

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

Aboubacrène Ag Ahmad<sup>(1)</sup>, El Hadji Deme<sup>(1)</sup>, Aliou Diop<sup>(1)</sup>,  
Stéphane Girard<sup>(2,\*)</sup> & Antoine Usseglio-Carleve<sup>(2)</sup>

<sup>(1)</sup> Université Gaston Berger, LERSTAD, UFR SAT, BP 234, Saint-Louis, Sénégal.

<sup>(2)</sup> Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

<sup>(\*)</sup> Corresponding author, [Stephane.Girard@inria.fr](mailto:Stephane.Girard@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 benefits from the same rate of convergence as in the unconditional situation. Its finite sample properties are illustrated both on simulated and real data.

**Keywords:** Semi-parametric estimation, regression and dispersion functions, tail-index, extreme conditional quantile.

## 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, 12, 22]. In many practical 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 [15],

the assessment of the optimal cost of the delivery activity of postal services [7], the analysis of longevity [27], the description of the upper tail of claim size distributions [1], the modeling of extremes in environmental time series [34], etc.

Here, we focus on the challenging situation where  $Y$  given  $x$  is heavy-tailed. Without additional assumptions on the pair  $(x, Y)$ , the estimation of conditional extreme quantiles may be addressed using nonparametric methods, see for instance the recent works of [9, 17, 19]. 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, 24]: a nonparametric regression model of the trend is combined with a parametric model for extreme values.

Our approach belongs to this 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 [36], 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. Moreover, this model imposes 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 feature to decouple the estimation of the nonparametric and extreme structures. As a consequence, we shall show that the semi-parametric estimators of conditional extreme quantiles are asymptotically Gaussian and benefit from the same rate of convergence as in the unconditional situation.

This paper is organized as follows. The location-dispersion regression model for heavy-tailed distribution 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 conditional extreme quantiles. Asymptotic results are provided in Section 4 while the finite sample behaviour of the estimators is illustrated in Section 5 on simulated data and in Section 6 on real data. Proofs are postponed to the Appendix.

## 2 Location-dispersion regression models 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

$$Y = a(x) + b(x)Z. \tag{1}$$

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

$$\bar{F}_Z(z) = z^{-1/\gamma} \ell(z), \quad z > 0. \quad (2)$$

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

$$\lim_{z \rightarrow \infty} \frac{\ell(\lambda z)}{\ell(z)} = 1.$$

$\bar{F}_Z$  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 (1) has been introduced by [36] in the random design setting where the functions  $a : \Pi \rightarrow \mathbb{R}$  and  $b : \Pi \rightarrow \mathbb{R}^+$  are referred to as the regression and dispersion functions respectively. Combining (1) and (2) yields

$$\bar{F}_Y(y | x) := \mathbb{P}(Y > y | x) = \bar{F}_Z\left(\frac{y - a(x)}{b(x)}\right) = \left(\frac{y - a(x)}{b(x)}\right)^{-1/\gamma} \ell\left(\frac{y - a(x)}{b(x)}\right), \quad (3)$$

for  $y \geq y_0(x) > a(x)$  where the functions  $a(\cdot)$ ,  $b(\cdot)$  and the conditional tail-index  $\gamma$  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  $\gamma$  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  $\gamma$  should yield more reliable estimates in small sample contexts than purely nonparametric approaches. A similar technique is implemented in [26]: An extreme-value distribution with constant extreme-value index is fitted to standardized rainfall maxima.

Starting with a  $n$ -sample  $\{(Y_1, x_1), \dots, (Y_n, x_n)\}$  from (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  $\alpha$ th quantile of  $Z$  and let  $0 < \mu_3 < \mu_2 < \mu_1 < 1$  such that

$$q_Z(\mu_2) = 0 \text{ and } q_Z(\mu_3) - q_Z(\mu_1) = 1. \quad (4)$$

From (1), it straightforwardly follows that, for all  $\alpha \in (0, 1)$ , the conditional quantile of  $Y$  given  $x \in \Pi$  is

$$q_Y(\alpha | x) = a(x) + b(x)q_Z(\alpha), \quad (5)$$

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

$$a(x) = q_Y(\mu_2 | x) \text{ and } b(x) = q_Y(\mu_3 | x) - q_Y(\mu_1 | x), \quad (6)$$

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

### 3 Inference

Let  $\{(Y_1, x_1), \dots, (Y_n, x_n)\}$  be a  $n$ -sample from (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). We assume that the design points  $x_i$ ,  $i = 1, \dots, n$  are included in  $\Pi$ , an open set of  $\mathbb{R}^d$  whose boundary has measure zero. Let  $\{\Pi_i, i = 1, \dots, n\}$  be a partition of  $\Pi$  such that  $x_i \in \Pi_i$  and let us denote by  $\lambda$  the Lebesgue measure and  $\|\cdot\|$  a norm on  $\mathbb{R}^d$ . 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 conditional extreme quantiles are derived by combining a plug-in method with Weissman's extrapolation device [38] 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  $\bar{F}_Y(y | x)$  is considered (see for instance [30, 31]). For all  $(x, y) \in \Pi \times \mathbb{R}$  let

$$\hat{F}_{n,Y}(y | x) = \sum_{i=1}^n \mathbb{1}_{\{Y_i > y\}} \int_{\Pi_i} K_h(x - t) dt, \quad (7)$$

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  $h = h_n \rightarrow 0$  as  $n \rightarrow \infty$  is a nonrandom sequence called the bandwidth. The corresponding estimator of  $q_Y(\alpha | x)$  is defined for all  $(x, \alpha) \in \Pi \times (0, 1)$  by

$$\hat{q}_{n,Y}(\alpha | x) = \hat{F}_{n,Y}^{\leftarrow}(\alpha | x) := \inf\{y; \hat{F}_{n,Y}(y | x) \leq \alpha\}. \quad (8)$$

Nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution function have been extensively investigated, see, for example [2, 32, 35], among others. In view of (6), the regression and dispersion functions are estimated by

$$\hat{a}_n(x) = \hat{q}_{n,Y}(\mu_2 | x) \text{ and } \hat{b}_n(x) = \hat{q}_{n,Y}(\mu_3 | x) - \hat{q}_{n,Y}(\mu_1 | x), \quad (9)$$

for all  $x \in \Pi$ .

### 3.2 Estimation of the conditional tail-index

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

$$\hat{Z}_i = \frac{Y_i - \hat{a}_n(x_i)}{\hat{b}_n(x_i)}, \quad (10)$$

for all  $i = 1, \dots, n$  where  $\hat{a}_n(\cdot)$  and  $\hat{b}_n(\cdot)$  are given in (9). In practice, nonparametric estimators can suffer from boundary effects [6, 28] and therefore only design points sufficiently far from the boundary of  $\Pi$  are considered. More specifically, consider  $\tilde{\Pi}^{(n)} = \{x \in \mathbb{R}^d, \text{ such that } B(x, h) \subset \Pi\}$  the erosion of the set  $\Pi$  by the ball  $B(0, h)$  centered at 0 and with radius  $h$ , see [33] for further details on mathematical morphology. Denote by  $I_n$  the set of indices associated with such design points  $I_n = \{i \in \{1, \dots, n\} \text{ such that } x_i \in \tilde{\Pi}^{(n)}\}$  and let  $m_n = \text{card}(I_n)$ . It can be shown that  $m_n = n(1 + O(h))$ , see Lemma 3 in the Appendix.

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

$$\hat{\gamma}_n = \frac{1}{k_n} \sum_{i=0}^{k_n-1} \log \hat{Z}_{m_n - i, m_n} - \log \hat{Z}_{m_n - k_n, m_n}. \quad (11)$$

This estimator is similar to Hill estimator [25], but in our context, it is built on non iid pseudo-observations.

### 3.3 Estimation of conditional extreme quantiles

Clearly, the purely nonparametric estimator (8) cannot estimate consistently extreme quantiles of levels  $\alpha_n$  arbitrarily small. For instance, when  $n\alpha_n \rightarrow 0$ , 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 conditional extreme quantile  $q_Y(\alpha_n | x)$ . To this end, we propose to take profit of the structure of the location-dispersion regression model (5) to define the plugin estimator

$$\tilde{q}_{n,Y}(\alpha_n | x) = \hat{a}_n(x) + \hat{b}_n(x) \hat{q}_{n,Z}(\alpha_n), \quad (12)$$

where  $\hat{a}_n(x)$  and  $\hat{b}_n(x)$  are given in (9) and  $\hat{q}_{n,Z}(\alpha_n)$  is the Weissman type estimator [38] defined by

$$\hat{q}_{n,Z}(\alpha_n) = \hat{Z}_{m_n - k_n, m_n} \left( \frac{\alpha_n m_n}{k_n} \right)^{-\hat{\gamma}_n}. \quad (13)$$

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 (12) cannot suffer from quantile crossing, a phenomenon which can occur with some quantile regression techniques [37].

## 4 Main results

The following general assumptions are required to establish the asymptotic behaviour 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)**  $(Y_1, x_1), \dots, (Y_n, x_n)$  are independent observations from the location-dispersion regression model for heavy-tailed distributions defined by (1), (2) and (4) and such that

$$\max_{i=1, \dots, n} \left| \lambda(\Pi_i) - \frac{\lambda(\Pi)}{n} \right| = o(1/n), \quad (14)$$

$$\max_{i=1, \dots, n} \sup_{(s,t) \in \Pi_i^2} \|s - t\| = O(n^{-1/d}). \quad (15)$$

We refer to [30, 31] 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  $\Pi$ ,  $b(\cdot)$  is lower bounded on  $\Pi$ ,  $b(t) \geq b_m > 0$  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  $c_2 > 0$  such that

$$|K(u) - K(v)| \leq c_2 \|u - v\|,$$

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, the so-called second-order condition is introduced (see for instance [22, eq (3.2.5)]):

(A.4) For all  $\lambda > 0$ , as  $z \rightarrow \infty$ ,

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

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

From [3, Theorem 1.5.12], it is clear that (2) is equivalent to  $U_Z \in \mathcal{RV}_\gamma$ , that is  $U_Z(\lambda z)/U_Z(z) \rightarrow \lambda^\gamma$  as  $z \rightarrow \infty$  for all  $\lambda > 0$ . 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  $\rho$ , see [22, Lemma 2.2.3]. It is then clear that  $\rho$ , referred to as the (conditional) second-order parameter, is a crucial quantity, tuning the rate of convergence of most extreme-value estimators, see [22, Chapter 3] for examples. Similarly to [31], the dimension  $d = 4$  plays a special role and we thus introduce for all  $d \geq 1$ :

$$\kappa(d) = 2 \max\{2, d/(d-2)\} = \begin{cases} 4 & \text{if } d \leq 4 \\ d/(d-2) & \text{if } d \geq 4. \end{cases}$$

Our first result states the joint asymptotic normality of the estimators (9) of the location and dispersion parameters.

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

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

where the coefficients of the matrix  $D$  are given by

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

A uniform consistency result can also be established:

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

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

As a consequence of Theorem 2, one can prove that the residuals  $\hat{Z}_i = (Y_i - \hat{a}_n(x_i))/\hat{b}_n(x_i)$ , see (10), are close to the unobserved  $Z_i$ ,  $i = 1, \dots, n$ .

**Corollary 1.** *Under the assumptions of Theorem 2, for all  $i \in I_n$ ,*

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

Our next main result provides the asymptotic normality of the conditional tail-index estimator (11) and the Weissman estimator computed on the residuals (13).

**Theorem 3.** *Assume (A.1)-(A.4) hold. Let  $(k_n)$  be an intermediate sequence of integers such that  $nh^d/(k_n \log n) \rightarrow \infty$ ,  $nh^{d+\kappa(d)}/\log n \rightarrow 0$  and  $\sqrt{k_n}A(n/k_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Then,*

$$(i) \quad \sqrt{k_n}(\hat{\gamma}_n - \gamma) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

(ii) *For all sequence  $(\alpha_n) \subset (0, 1)$  such that  $n\alpha_n/k_n \rightarrow 0$  as  $n \rightarrow \infty$ ,*

$$\frac{\sqrt{k_n}}{\log \left( \frac{k_n}{n\alpha_n} \right)} \left( \log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n) \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

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 [20] 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 [9, Corollary 1 & 2], [8, Theorem 3] and [19, Theorem 2]. To be more specific, remark first that conditions  $nh^d/(k_n \log n) \rightarrow \infty$  and  $nh^{d+\kappa(d)}/\log n \rightarrow 0$  imply that  $k_n = o((n/\log n)^{\kappa(d)/(d+\kappa(d))})$ . Second, following [22, Eq. (3.2.10)], if  $A$  is a power function, then condition  $\sqrt{k_n}A(n/k_n) \rightarrow 0$  as  $n \rightarrow \infty$  yields  $k_n = o(n^{-2\rho/(1-2\rho)})$  and the constraint is then  $k_n = o(n^{(-2\rho/(1-2\rho)) \wedge (\kappa(d)/(d+\kappa(d)))})$ , up to logarithmic factors. If  $\rho \geq -\kappa(d)/(2d)$ , the rate of convergence of  $\hat{\gamma}_n$  is thus  $n^{\rho/(1-2\rho)}$  which is the classical rate for estimators of the tail-index, see for instance [23, Remark 3]. For instance, in the situation where the dimension of the covariate is  $d \leq 4$ , then the  $n^{\rho/(1-2\rho)}$  rate is reached as soon as  $\rho \geq -2$  which encompasses most usual distributions.

Theorem 4 states the asymptotic normality of the estimator (12) of extreme conditional quantiles.

**Theorem 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  $nh^d/(k_n \log n) \rightarrow \infty$ ,  $nh^{d+\kappa(d)} \rightarrow 0$  and*



$\sqrt{k_n}A(n/k_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Then, for all sequences  $(t_n) \subset \tilde{\Pi}^{(n)}$  and  $(\alpha_n) \subset (0, 1)$  such that  $n\alpha_n/k_n \rightarrow 0$  as  $n \rightarrow \infty$ ,

$$\frac{\sqrt{k_n}}{q_Z(\alpha_n) \log\left(\frac{k_n}{n\alpha_n}\right)} \left( \frac{\tilde{q}_{n,Y}(\alpha_n | t_n) - q_Y(\alpha_n | t_n)}{b(t_n)} \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2). \quad (16)$$

Remark that  $b(t_n)q_Z(\alpha_n) \sim a(t_n) + b(t_n)q_Z(\alpha_n) = q_Y(\alpha_n | t_n)$  and therefore (16) can be rewritten as

$$\frac{\sqrt{k_n}}{\log\left(\frac{k_n}{n\alpha_n}\right)} \left( \frac{\tilde{q}_{n,Y}(\alpha_n | t_n)}{q_Y(\alpha_n | t_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

As a comparison, the rate of convergence of purely nonparametric methods involves an extra  $h^{d/2}$  factor, see for instance [16, Theorem 3] or [8, Theorem 3]. The location-dispersion regression model allows to dampen this vexing effect of the dimensionality.

A uniform consistency result can also be established:

**Theorem 5.** *Assume (A.1)-(A.4) hold. Let  $(k_n)$  be an intermediate sequence of integers. Suppose  $nh^d/(k_n \log n) \rightarrow \infty$ ,  $nh^{d+\kappa(d)}/\log n \rightarrow 0$  and  $\sqrt{k_n}A(n/k_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Then, for all sequence  $(\alpha_n) \subset (0, 1)$  such that  $n\alpha_n/k_n \rightarrow 0$  as  $n \rightarrow \infty$ ,*

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

## 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 conditional extreme 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- $t_\nu$  random variable where  $\nu \in \{1, 2, 4\}$  denotes the degrees of freedom and introduce  $Z = Z_0/(2q_{Z_0}(\mu_1))$  the associated rescaled Student random variable. Symmetry arguments yield  $q_Z(\mu_2) = 0$ ,  $q_Z(\mu_3) = -q_Z(\mu_1)$  and  $q_Z(\mu_1) = q_{Z_0}(\mu_1)/(2q_{Z_0}(\mu_1)) = 1/2$  by construction. Therefore (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  $\rho = -2/\nu$ .

- Let  $Z_0$  be a Burr random variable with parameters  $\gamma \in \{1/4, 1/2, 1\}$ ,  $\rho = -1$  and survival function  $\bar{F}_{Z_0}(z) = 1/(1 + z^{1/\gamma})$ ,  $z \geq 0$ . Introduce the translated and rescaled Burr( $\gamma, -1$ ) random variable

$$Z = \frac{Z_0 - (\mu_2^{-1} - 1)^\gamma}{(\mu_3^{-1} - 1)^\gamma - (\mu_1^{-1} - 1)^\gamma},$$

such as (4) holds. The second-order condition (A.4) is also fulfilled.

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

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

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], 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  $t_\nu$  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. The estimation of the conditional tail-index and conditional extreme quantiles require the selection of the sample fraction  $k_n$ . This parameter is selected using a mean-squared error criterion as in [22, Section 3.2]. Assuming that  $A(t) = ct^\rho$ , the optimal value of  $k_n$  is given by

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

Since  $\rho$  may be difficult to estimate in practice, in several works dealing with the bias reduction of tail-index estimators, a miss-specified value  $\rho = -1$  is considered, see [13] or [21] for instance. Letting moreover  $c = 1$ , we end up with  $k_n^* = \tilde{\gamma}^{2/3} n^{2/3}$  where  $\tilde{\gamma}$  is a prior naive estimation of  $\gamma$  computed with  $k_n = n^{1/2}$ .

The estimation results associated with the conditional tail-index and the conditional quantiles  $q_Y(1/n | \cdot)$  of extreme level  $\alpha_n = 1/n$  are displayed on Figure 4 and Figure 3

respectively. It appears that the tail-index estimator (11) has a stronger bias for larger values of  $\nu$ . These empirical results are in line with the properties of the Student distribution. Indeed, the second-order parameter  $\rho = -2/\nu$  being increasing with  $\nu$ , the bias of the Hill-type estimator increases as well. As expected, the estimated conditional extreme 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 conditional extreme quantile estimator (12) thanks to a comparison with a purely nonparametric one. The nonparametric estimator is based on the ideas of the moving window approach introduced in [14]. For each  $x \in \tilde{\Pi}^{(n)}$ , a subsample

$$\{(Y_i^*, x_i^*)\}_{i=1, \dots, n^*} = \{(Y_i, x_i), 1 \leq i \leq n, \text{ s.t. } \|x - x_i\| < h\}$$

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

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

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

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

The comparison between the true and estimated conditional extreme quantiles is based on a relative median-squared relative error (RMSE) computed on the  $N = 100$  replications and  $m_n$  design points in the square  $\tilde{\Pi}^{(n)}$ :

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

where  $\hat{q}_{n,Y}^{[r]}(\alpha_n | \cdot)$  denotes either  $\tilde{q}_{n,Y}(\alpha_n | \cdot)$  or  $\hat{q}_{n,Y}^*(\alpha_n | \cdot)$  computed on the  $r$ th replication. Here, both Student- $t_\nu$  and Burr( $\gamma, -1$ ) distributions are considered with  $n \in \{20^2, 40^2, 60^2, 80^2, 100^2\}$ ,  $\nu \in \{1, 2, 4\}$ ,  $\gamma \in \{1, 1/2, 1/4\}$  and  $\alpha_n = 1/n$ . The RMSE are reported in Table 1. For both estimators, it appears that the main driver of the relative error is the tail heaviness. The nonparametric estimator even seems not to converge on the Burr distribution with tail-index  $\gamma = 1$ . Unsurprisingly, the semi-parametric estimator  $\tilde{q}_{n,Y}$  provides much better results than the nonparametric one  $\hat{q}_{n,Y}^*$ : Its RMSE is smaller and converges towards 0 at a faster rate when the sample size  $n$  increases.

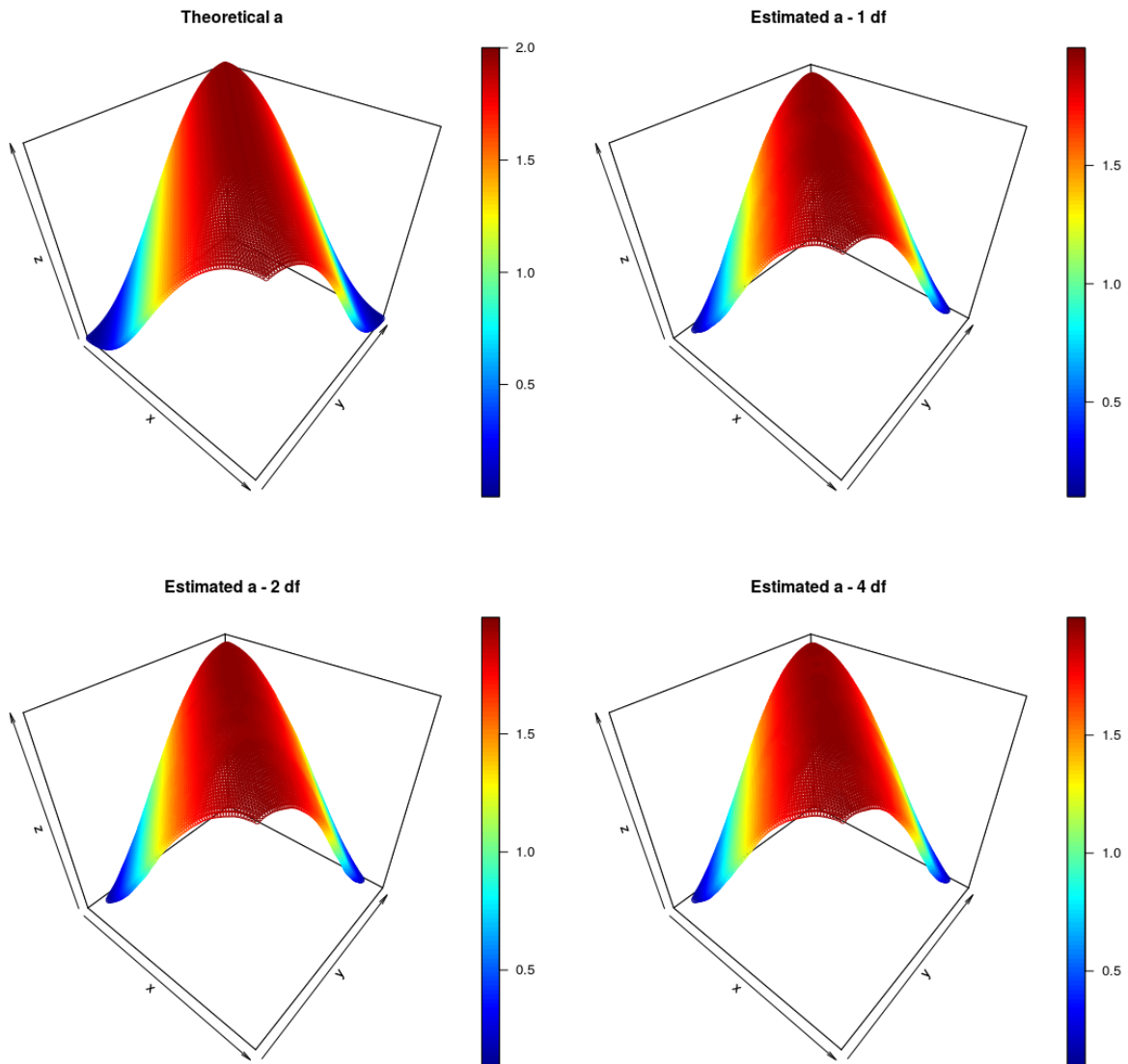


Figure 1: Simulation results obtained on a Student- $t_\nu$  distribution. From top to bottom, left to right : Theoretical function  $a(\cdot)$ , and means over  $N = 100$  replications of estimates  $\hat{a}_n(\cdot)$  computed on  $n = 10,000$  observations for  $\nu \in \{1, 2, 4\}$ . All coordinates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

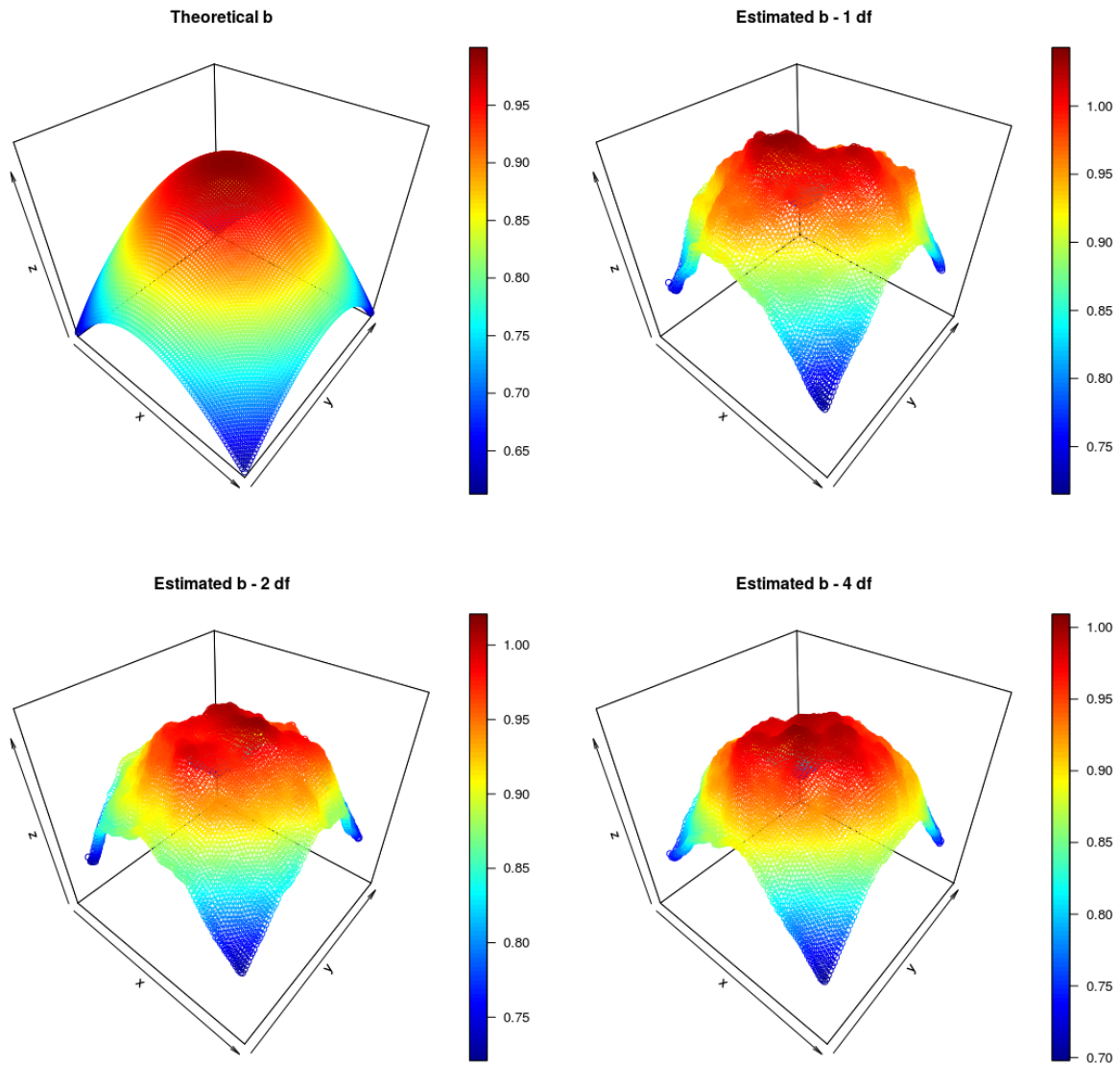


Figure 2: Simulation results obtained on a Student- $t_\nu$  distribution. From top to bottom, left to right : Theoretical function  $b(\cdot)$ , and means over  $N = 100$  replications of estimates  $\hat{b}_n(\cdot)$  computed on  $n = 10,000$  observations for  $\nu \in \{1, 2, 4\}$ . All coordinates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

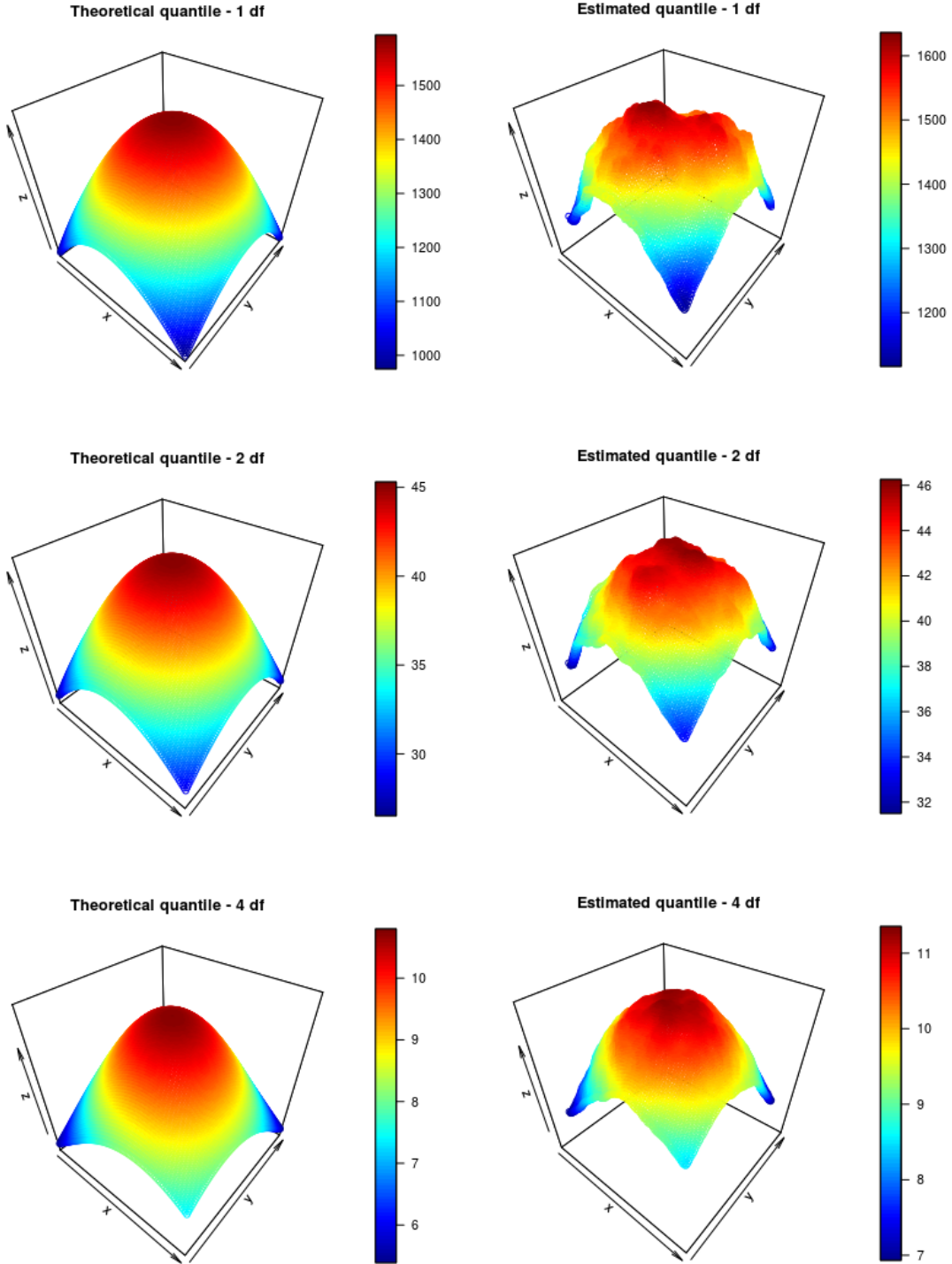


Figure 3: Simulation results obtained on a Student- $t_\nu$  distribution for  $\nu = 1$  (top),  $\nu = 2$  (middle) and  $\nu = 4$  (bottom). Left panels: Theoretical quantiles  $q_Y(1/n | \cdot)$ . Right panels: Means over  $N = 100$  replications of estimates  $\tilde{q}_{n,Y}(1/n | \cdot)$  computed 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.

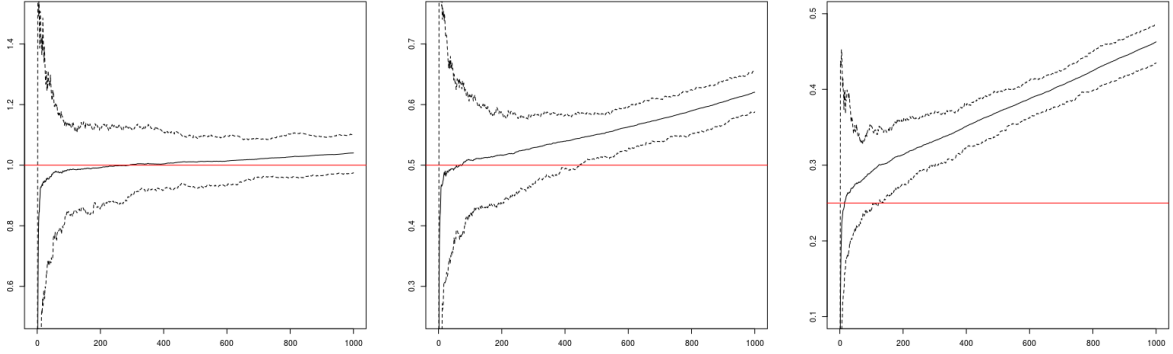


Figure 4: Simulation results obtained on a Student- $t_\nu$  distribution for  $\nu = 1$  (left),  $\nu = 2$  (middle) and  $\nu = 4$  (right). Estimator of the tail-index (11) (continuous line) and 95% empirical confidence intervals (dotted lines) as functions of the sample fraction  $k_n$ . The true value  $\gamma = 1/\nu$  is depicted by a red horizontal line.

$n$	Student, $\nu = 1$	Student, $\nu = 2$	Student, $\nu = 4$
400	0.547 (0.890)	0.129 (0.643)	0.062 (0.442)
1,600	0.138 (0.867)	0.065 (0.533)	0.020 (0.284)
3,600	0.145 (0.855)	0.048 (0.477)	0.012 (0.226)
6,400	0.061 (0.845)	0.032 (0.456)	0.011 (0.206)
10,000	0.045 (0.820)	0.026 (0.425)	0.013 (0.184)
$n$	Burr, $\gamma = 1, \rho = -1$	Burr, $\gamma = 1/2, \rho = -1$	Burr, $\gamma = 1/4, \rho = -1$
400	0.525 (0.746)	0.197 (0.329)	0.104 (0.129)
1,600	0.182 (0.796)	0.068 (0.348)	0.038 (0.124)
3,600	0.157 (0.825)	0.056 (0.333)	0.023 (0.118)
6,400	0.096 (0.827)	0.054 (0.311)	0.020 (0.107)
10,000	0.070 (0.845)	0.030 (0.301)	0.023 (0.102)

Table 1: Relative median squared errors associated with the estimation of the conditional extreme quantile  $q_Y(1/n|\cdot)$ . Results obtained with the semi-parametric estimator  $\tilde{q}_{n,Y}$  and comparison with the purely nonparametric one  $\hat{q}_{n,Y}^*$  (between brackets).

## 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 occurred. We focus on the 2011 Tohoku tsunami, in Japan. This earthquake was the cause of the Fukushima Daiichi nuclear disaster. Indeed, a wave of height more 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 associated with small probability wave heights 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 in  $n = 5,364$  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). In each station, we propose to estimate an extreme quantile of the wave height, following the methodology introduced in this paper.

To this end, we first have to select a bandwidth. Noticing that the standard deviations of  $x^{(1)}$  and  $x^{(2)}$  are respectively 1.63 and 1.16, we fixed  $h_n^* = 1.63 \times n^{-1/6} \simeq 0.4$ . The regression and dispersion functions are then estimated via (9) 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 (10). In a first time, it is necessary to check whether the residuals have a heavy-tailed behavior. The common practice is to use a graphical diagnosis. 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 Section 5. 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 Hill estimator (11) is depicted as a function of  $k_n$  on the bottom-left panel of Figure 6. It features a nice stable behaviour, confirming the heavy-tail assumption, and pointing towards a tail-index close to 0.25. Finally, the conditional extreme quantile estimator (12) 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  $\tilde{q}_{n,Y}(\alpha_n | x_1), \dots, \tilde{q}_{n,Y}(\alpha_n | x_n)$  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.

## 7 Appendix: Proofs

Technical lemmas are collected in Paragraph 7.1 while preliminary results of general interest are provided in Paragraph 7.2. Finally, the proofs of the main results are given in Paragraph 7.3.

### 7.1 Auxiliary lemmas

The first result is an adaptation of Bochner's lemma to the multidimensional fixed design setting.

**Lemma 1.** *Let  $\psi(\cdot | \cdot) : \mathbb{R}^p \times \Pi \rightarrow \mathbb{R}^+$  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  $\psi(\cdot | \cdot)$  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*

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

where  $x_i \in \Pi_i$  such that (14) and (15) 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, as  $n \rightarrow \infty$ ,

$$\psi_n(y_n | t_n) = \psi(y_n | t_n) \int_S Q(u) du + O(n^{-1/d}) + O(h^2).$$

*Proof.* Consider the expansion

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

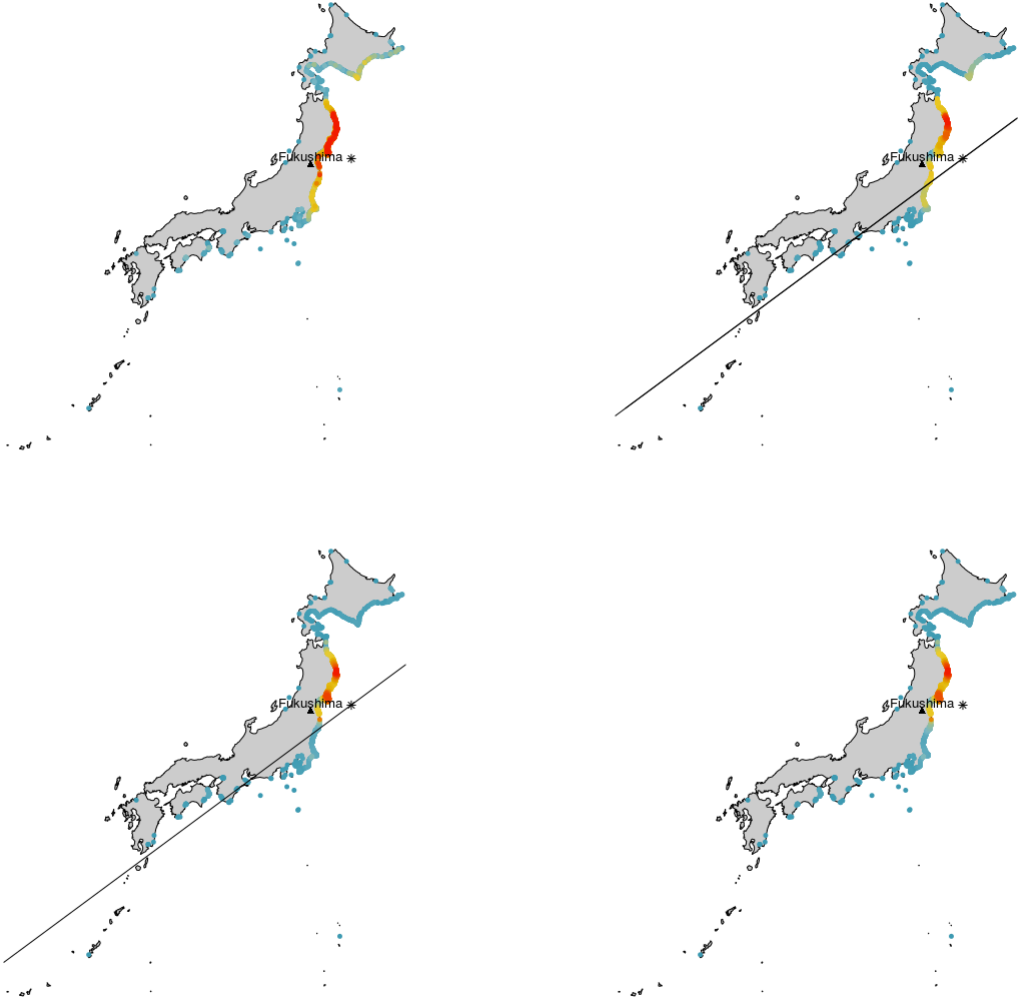


Figure 5: Results on tsunami data. Top-left: Maximum wave height recorded at each station. Top-right: Regression function estimate  $\hat{a}_n(\cdot)$  at each station. Bottom-left: Dispersion function estimate  $\hat{b}_n(\cdot)$  at each station. Bottom-right: Quantile estimate  $\tilde{q}_{n,Y}(10/n | \cdot)$  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.

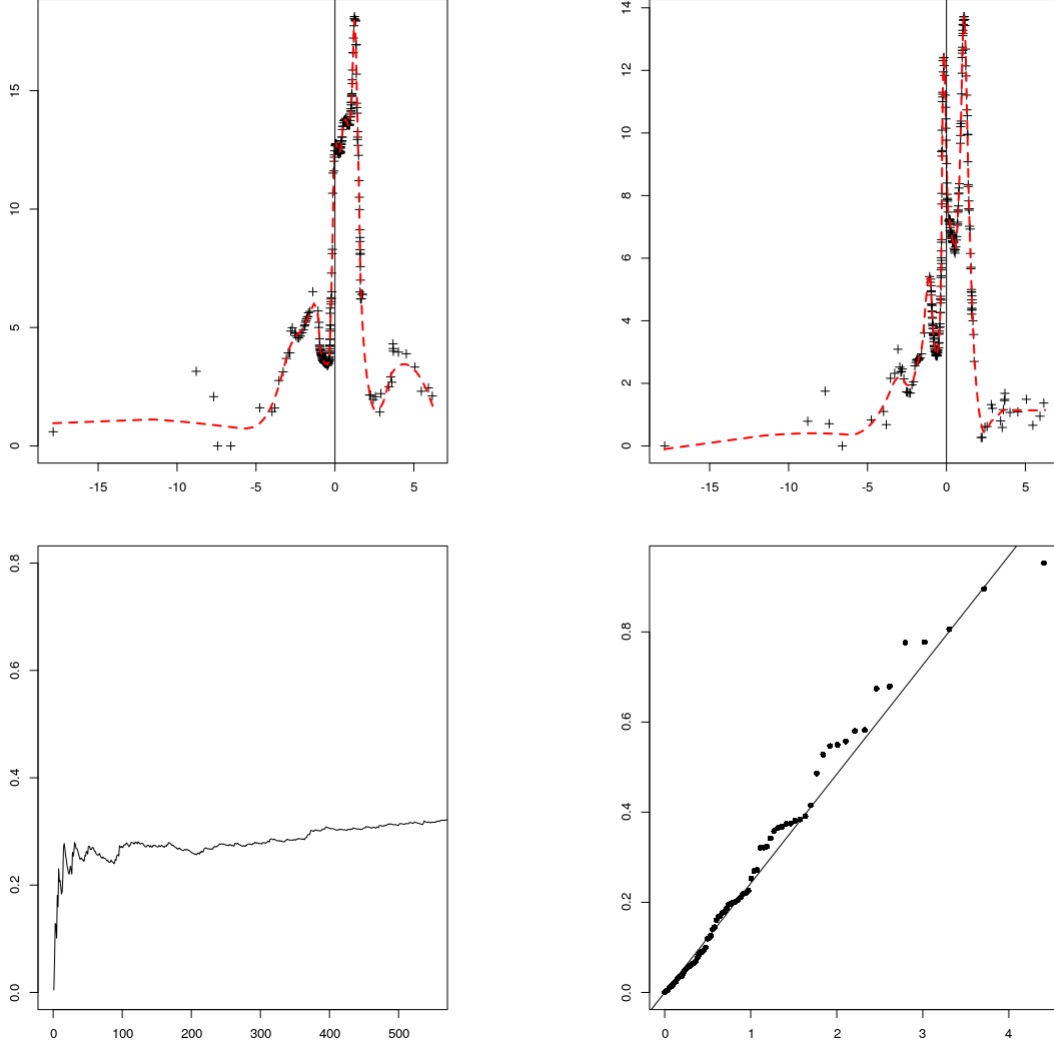


Figure 6: Results on tsunami data. Top: Regression (left) and dispersion (right) function estimates  $\hat{a}_n(\cdot)$  and  $\hat{b}_n(\cdot)$  along the principal axis  $x^{(2)} = 1.64x^{(1)} + 80.35$ . The estimates at each station (black +) are smoothed (red dashed line) for the visualization sake. The vertical black line displays the projection of the epicenter on the principal axis. Bottom left: Hill estimator (11) as a function of  $k_n$ . Bottom right: Log-excesses of the residuals  $\log(\hat{Z}_{n-i+1,n}/\hat{Z}_{n-k_n^*+1,n})$  versus  $\log(k_n^*/i)$ ,  $1 \leq i \leq k_n^* = 82$ . The straight line has slope  $\hat{\gamma}_n \simeq 0.25$ .

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

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

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

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

Let  $\nabla_2[\psi](\cdot, \cdot)$  denote the gradient of  $\psi(\cdot | \cdot)$  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  $y_n \in C$ ,

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

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

$$T_{n,1} = O(h^2). \tag{17}$$

Let us now turn to the second term

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

Since  $\psi(\cdot | \cdot)$  is continuously differentiable with respect to its second argument, there exists  $c_1 > 0$  such that

$$|T_{n,2}| \leq \sum_{i=1}^n \int_{\Pi_i} |\psi(y_n | x_i) - \psi(y_n | s)| Q_h(t_n - s)ds \leq c_1 \sum_{i=1}^n \int_{\Pi_i} \|x_i - s\| Q_h(t_n - s)ds$$

Moreover, under assumption (15),

$$\begin{aligned} |T_{n,2}| &= \sum_{i=1}^n \int_{\Pi_i} Q_h(t_n - s)ds O(n^{-1/d}) = \int_{\Pi} Q_h(t_n - s)ds O(n^{-1/d}) \\ &= O(n^{-1/d}). \end{aligned} \tag{18}$$

Finally, collecting (17) and (18), the conclusion follows.  $\square$

As a consequence of Lemma 1, the asymptotic bias and variance of the estimator (7) of the conditional survival function can be derived.

**Lemma 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,

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

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

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

*Proof.* (i) Clearly,

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

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

(ii) As a consequence of the independence assumption,

$$\begin{aligned} \text{var} \left( \hat{F}_{n,Y}(y_n | t_n) \right) &= \sum_{i=1}^n \bar{F}_Y(y_n | x_i) S_{n,i} - \sum_{i=1}^n \bar{F}_Y^2(y_n | x_i) S_{n,i} \\ &=: T_{n,1} - T_{n,2}, \end{aligned}$$

where

$$S_{n,i} := \left( \int_{\Pi_i} K_h(t_n - s) ds \right)^2 = \frac{1}{h^{2d}} \int_{\Pi_i} \int_{\Pi_i} K \left( \frac{t_n - s_1}{h} \right) K \left( \frac{t_n - s_2}{h} \right) ds_1 ds_2. \quad (19)$$

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 (15),

$$\left| K \left( \frac{t_n - s_2}{h} \right) - K \left( \frac{t_n - s_1}{h} \right) \right| \leq \frac{c_2 \|s_2 - s_1\|}{h} = O \left( \frac{1}{n^{1/d} h} \right),$$

uniformly on  $(s_1, s_2) \in \Pi_i^2$  and  $i = 1, \dots, n$ . It thus follows from (14) that

$$\begin{aligned} S_{n,i} &= \frac{1}{h^{2d}} \int_{\Pi_i} \int_{\Pi_i} \left[ K^2 \left( \frac{t_n - s_1}{h} \right) + K \left( \frac{t_n - s_1}{h} \right) O \left( \frac{1}{n^{1/d} h} \right) \right] ds_1 ds_2 \\ &= \frac{\lambda(\Pi)}{nh^{2d}} \int_{\Pi_i} K^2 \left( \frac{t_n - s}{h} \right) ds (1 + o(1)) + O \left( \frac{1}{n^{1+1/d} h^{2d+1}} \right) \int_{\Pi_i} K \left( \frac{t_n - s}{h} \right) ds \\ &= \frac{\lambda(\Pi) \|K\|_2^2}{nh^d} \int_{\Pi_i} M_h(t_n - s) ds (1 + o(1)) + O \left( \frac{1}{n^{1+1/d} h^{d+1}} \right) \int_{\Pi_i} K_h(t_n - s) ds \end{aligned} \quad (20)$$

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

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

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

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

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

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

and the conclusion follows:

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

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

Finally, Lemma 3 is an adaptation of [18, Lemma 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 3.** *Recall that  $I_n = \{i \in \{1, \dots, n\} \text{ such that } x_i \in \tilde{\Pi}^{(n)}\}$  and  $m_n = \text{card}(I_n)$ . Assume  $nh^d \rightarrow \infty$  as  $n \rightarrow \infty$ .*

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

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

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

*Proof.* (i) Let  $C_n = \Pi \setminus \tilde{\Pi}^{(n)}$ ,  $J_n = \{i \in \{1, \dots, n\} \text{ such that } x_i \in C_n\}$  and  $N_n := \text{card}(J_n)$ . For all  $i \in J_n$ ,  $x_i \in C_n$  and for  $n$  large enough,  $nh^d \rightarrow \infty$  together with (15) entail that  $\Pi_i \subset C_n$ . Therefore, as the sets  $\Pi_i$  are disjoint:

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

in view of the absolute continuity of the erosion with respect to Lebesgue measure, see [29]. From (14),  $\lambda(\Pi_i) \sim \lambda(\Pi)/n$  uniformly on  $i$  and thus  $N_n = O(nh)$ . Therefore,  $m_n = n - N_n = n(1 + O(h)) \sim n$  as  $n \rightarrow +\infty$ .

(ii) The conclusion follows by remarking that in view of (2) the distribution of  $Z$  has an infinite upper endpoint and by applying [18, Lemma 3].  $\square$

## 7.2 Preliminary results

Let  $\vee$  (resp.  $\wedge$ ) denote the maximum (resp. the minimum). The next proposition provides a joint asymptotic normality result for the estimator (7) of the conditional survival function evaluated at points depending on  $n$ .

**Proposition 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  $\varepsilon_{j,n} \rightarrow 0$  as  $n \rightarrow \infty$ . If  $nh^d \rightarrow \infty$  and  $nh^{d+\kappa(d)} \rightarrow 0$  as  $n \rightarrow \infty$ , then*

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

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

*Proof.* Let us first remark that, for all  $j \in \{1, \dots, J\}$ , in view of (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  $\varepsilon_{j,n} \rightarrow 0$  as  $n \rightarrow \infty$ . Besides, from (3),  $F_Y(y_{j,n} \mid t_n) = F_Z(q_Z(\alpha_j) + \varepsilon_{j,n}) \rightarrow 1 - \alpha_j > 0$  as  $n \rightarrow \infty$  and thus the assumptions of Lemma 2(i,ii) are satisfied. Now, let  $\beta \neq 0$  in  $\mathbb{R}^J$ ,  $J \geq 1$  and consider the random variable

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

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  $Y_1, \dots, Y_n$ ,

$$\text{var}(\Gamma_{n,1}) = \sum_{i=1}^n \left( \int_{\Pi_i} K_h(t_n - s) ds \right)^2 \text{var} \left( \sum_{j=1}^J \beta_j \mathbb{1}_{\{Y_i > y_{j,n}\}} \right) =: \sum_{i=1}^n S_{n,i} \beta^t \Sigma^{(i,n)} \beta, \quad (21)$$

where  $S_{n,i}$  is defined by (19) in the proof of Lemma 2, and where  $\Sigma^{(i,n)}$  is the matrix whose coefficients are defined for  $(k, l) \in \{1, \dots, J\}^2$  by  $\Sigma_{k,l}^{(i,n)} = \text{cov}(\mathbb{1}_{\{Y_i > y_{k,n}\}}, \mathbb{1}_{\{Y_i > y_{l,n}\}})$ . In view

of (20),

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

and straightforward calculations yield

$$\begin{aligned} \Sigma_{k,l}^{(i,n)} &= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) - \bar{F}_Y(y_{k,n} \mid x_i) \bar{F}_Y(y_{l,n} \mid x_i) \\ &= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) - \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) \bar{F}_Y(y_{k,n} \wedge y_{l,n} \mid x_i) \\ &= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) F_Y(y_{k,n} \wedge y_{l,n} \mid x_i) \\ &=: \varphi(y_{k,n}, y_{l,n} \mid x_i), \end{aligned} \tag{22}$$

where  $\varphi$  is the function  $\mathbb{R}^2 \times \Pi \rightarrow [0, 1]$  defined by  $\varphi(\cdot, \cdot \mid \cdot) = \bar{F}_Y(\cdot \vee \cdot \mid \cdot) F_Y(\cdot \wedge \cdot \mid \cdot)$ . Replacing in (21) yields  $\text{var}(\Gamma_{n,1}) = \beta^t C^{(n)} \beta$ , where  $C^{(n)}$  is the covariance matrix whose coefficients are defined by

$$\begin{aligned} C_{k,l}^{(n)} &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \sum_{i=1}^n \varphi(y_{k,n}, y_{l,n} \mid x_i) \int_{\Pi_i} M_h(t_n - s) ds (1 + o(1)) \\ &+ O\left(\frac{1}{n^{1+1/d}h^{d+1}}\right) \sum_{i=1}^n \varphi(y_{k,n}, y_{l,n} \mid x_i) \int_{\Pi_i} K_h(t_n - s) ds \\ &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} [\varphi(y_{k,n}, y_{l,n} \mid t_n) + O(h^2) + O(n^{-1/d})] (1 + o(1)) \\ &+ O\left(\frac{1}{n^{1+1/d}h^{d+1}}\right) [\varphi(y_{k,n}, y_{l,n} \mid t_n) + O(h^2) + O(n^{-1/d})] \\ &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} [\varphi(y_{k,n}, y_{l,n} \mid t_n)(1 + o(1)) + O(h^2) + O(n^{-1/d})], \end{aligned} \tag{23}$$

from Lemma 1 applied twice with  $p = 2$  and recalling that  $nh^d \rightarrow \infty$ . Besides, let us remark that, in view of (5),

$$y_{k,n} - y_{l,n} = b(t_n)(q_Z(\alpha_k) - q_Z(\alpha_l) + \varepsilon_{k,n} - \varepsilon_{l,n}) = b(t_n)(q_Z(\alpha_k) - q_Z(\alpha_l))(1 + o(1)),$$

as  $n \rightarrow \infty$ . Therefore, assuming for instance  $k < l$  implies  $\alpha_k > \alpha_l$  and thus  $q_Z(\alpha_k) < q_Z(\alpha_l)$  leading to  $y_{k,n} < y_{l,n}$  for  $n$  large enough. More generally,  $y_{k,n} \vee y_{l,n} = y_{k \vee l, n}$  and  $y_{k,n} \wedge y_{l,n} = y_{k \wedge l, n}$  for  $n$  large enough and thus  $\varphi(y_{k,n}, y_{l,n} \mid t_n) = \bar{F}_Y(y_{k \vee l, n} \mid t_n) F_Y(y_{k \wedge l, n} \mid t_n)$ . From (3) and (5), we have

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

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

$$\varphi(y_{k,n}, y_{l,n} \mid t_n) \rightarrow B_{k,l} = \alpha_{k \vee l}(1 - \alpha_{k \wedge l}) \text{ as } n \rightarrow \infty. \tag{24}$$



Collecting (23) and (24), one has

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

and therefore

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

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

$$\sum_{i=1}^n \mathbb{E}|T_{i,n}|^3 \Big/ \text{var}(\Gamma_{n,1}) \rightarrow 0 \quad (25)$$

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

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

in view of (A.3) and (14). As a consequence,

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

and it is thus clear that (25) holds under the assumption  $nh^d \rightarrow \infty$ . As a result,

$$\sqrt{nh^d} \Gamma_{n,1} \xrightarrow{d} \mathcal{N}(0, \lambda(\Pi)\|K\|_2^2 \beta^t B \beta). \quad (26)$$

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

$$\sqrt{nh^d} |\Gamma_{n,2}| \leq \sqrt{nh^d} \sum_{j=1}^J |\beta_j| \left| \mathbb{E} \left( \hat{F}_{n,Y}(y_{j,n} | t_n) \right) - \bar{F}_Y(y_{j,n} | t_n) \right| = O(\sqrt{nh^{d+\kappa(d)}}) = o(1). \quad (27)$$

Finally, collecting (26) and (27),  $\sqrt{nh^d} \Gamma_n$  converges to a centered Gaussian random variable with variance  $\lambda(\Pi)\|K\|_2^2 \beta^t B \beta$ , and the result follows.  $\square$

The following proposition provides the joint asymptotic normality of the estimator (8) of conditional quantiles. It can be read as an adaptation of classical results [2, 32, 35] to the location-dispersion regression model in the multivariate fixed design setting.

**Proposition 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  $nh^d \rightarrow \infty$  and  $nh^{d+\kappa(d)} \rightarrow 0$  as  $n \rightarrow \infty$ , then

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

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

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

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

These notations yield

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

From (3) and (5), the nonrandom term can be rewritten as

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

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

$$v_{j,n} = s_j f_Z(q_Z(\alpha_j) + \theta_{j,n} \varepsilon_{j,n}).$$

In view of the continuity of  $f_Z(\cdot)$  and since  $\varepsilon_{j,n} \rightarrow 0$  as  $n \rightarrow \infty$ , it follows that

$$v_{j,n} = \frac{s_j}{H_Z(\alpha_j)} (1 + o(1)). \tag{28}$$

Let us now turn to the random term. Recalling that, for all  $j = 1, \dots, J$ ,

$$y_{j,n} = q_Y(\alpha_j | t_n) + b(t_n) \varepsilon_{j,n},$$

with  $\varepsilon_{j,n} \rightarrow 0$  as  $n \rightarrow \infty$ , Proposition 1 entails  $\{V_{j,n}\}_{j=1,\dots,J}$  converges to a centered Gaussian random variable with covariance matrix  $\lambda(\Pi)\|K\|_2^2 B$ . Taking account of (28) 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.  $\square$

The following proposition provides a uniform consistency result for the estimator (8) 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  $\Pi$ .

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

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

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

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

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

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

so that the following expansion holds:

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

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

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

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

$$\alpha_{i,n}^+ = \frac{M(\varepsilon, \alpha)}{v_n H_Z(\alpha)} (1 + o(1)) + O(h^2) + O(n^{-1/d}) = \frac{\kappa_1(\varepsilon, \alpha)}{v_n} (1 + o(1)), \quad (29)$$

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

$$\delta_n^+ = \mathbb{P}\left(\bigcup_{i \in I_n} \left\{ \xi_{i,n}^+ \geq \frac{\kappa_1(\varepsilon, \alpha)}{v_n} (1 + o(1)) \right\}\right) \leq \sum_{i \in I_n} \mathbb{P}\left(\xi_{i,n}^+ \geq \frac{\kappa_1(\varepsilon, \alpha)}{v_n} (1 + o(1))\right). \quad (30)$$

Moreover,

$$\mathbb{P}\left(\xi_{i,n}^+ \geq \frac{\kappa_1(\varepsilon, \alpha)}{v_n} (1 + o(1))\right) =: \mathbb{P}\left(\sum_{j=1}^n \tilde{X}_j \geq \frac{\kappa_1(\varepsilon, \alpha)}{v_n} (1 + o(1))\right), \quad (31)$$

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

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

are independent, centered and bounded from (14):

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

Lemma 2(ii) entails

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

since  $\alpha_{i,n}^+ \rightarrow 0$  as  $n \rightarrow \infty$  from (29) and thus  $\bar{F}_Y(q_{i,n}^+ | x_i) \rightarrow \alpha$  as  $n \rightarrow \infty$  in view of the continuity of  $\bar{F}_Y(\cdot | x_i)$ . Bernstein's inequality for bounded random variables yields

$$\begin{aligned}
(31) &\leq \exp\left(-\frac{\kappa_1^2(\varepsilon, \alpha) \log n}{2\kappa_2(\alpha) + \frac{2\kappa_1(\varepsilon, \alpha)(1+o(1))}{3v_n}}(1+o(1))\right) \\
&= \exp\left(-\frac{\kappa_1^2(\varepsilon, \alpha) \log n}{2\kappa_2(\alpha)}(1+o(1))\right) \\
&= \exp[-2(1 - \log(\varepsilon/2)) \log n (1+o(1))] \\
&\leq \exp[-(1 - \log(\varepsilon/2)) \log n], \tag{32}
\end{aligned}$$

for  $n$  large enough. Collecting (30)-(32) yields

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

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

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

which is the desired result.  $\square$

### 7.3 Proofs of main results

The proof of Theorem 1 directly relies on Proposition 2:

*Proof of Theorem 1.* Let us remark that

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

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

Proposition 2 with  $J = 3$ ,  $\alpha_1 = \mu_1$ ,  $\alpha_2 = \mu_2$  and  $\alpha_3 = \mu_3$  yields that  $\xi_n$  converges in distribution to the centred  $\mathcal{N}(0_{\mathbb{R}^3}, \lambda(\Pi) \|K\|_2^2 C)$  distribution where

$$C = \begin{pmatrix} \mu_1(1 - \mu_1)H_Z^2(\mu_1) & \mu_2(1 - \mu_1)H_Z(\mu_2)(H_Z(\mu_1)) & \mu_3(1 - \mu_1)H_Z(\mu_3)H_Z(\mu_1) \\ \mu_2(1 - \mu_1)H_Z(\mu_2)H_Z(\mu_1) & \mu_2(1 - \mu_2)H_Z^2(\mu_2) & \mu_3(1 - \mu_2)H_Z(\mu_2)H_Z(\mu_3) \\ \mu_3(1 - \mu_1)H_Z(\mu_3)H_Z(\mu_1) & \mu_3(1 - \mu_2)H_Z(\mu_2)H_Z(\mu_3) & \mu_3(1 - \mu_3)H_Z^2(\mu_3) \end{pmatrix}.$$

Therefore,  $\tilde{A} \xi_n \xrightarrow{d} \mathcal{N}(0_{\mathbb{R}^2}, \lambda(\Pi) \|K\|_2^2 \tilde{A} C \tilde{A}^t)$  and the conclusion follows from standard calculations.  $\square$

Theorem 2 is a straightforward consequence of Proposition 3:

*Proof of Theorem 2.* Remarking that

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

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

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

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

*Proof of Corollary 1.* Remark that for all  $i \in I_n$ , one has

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

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

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

On the one hand, Theorem 2 entails

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

On the other hand,

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

Again, Theorem 2 shows that the following uniform consistency holds: For all  $\varepsilon > 0$ , there exists  $M(\varepsilon) > 0$  such that

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

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

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

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.  $\square$

*Proof of Theorem 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  $m_n = \text{card}(I_n)$  where  $I_n = \{i \in \{1, \dots, n\} \text{ such that } x_i \in \tilde{\Pi}^{(n)}\}$ . The first term is controlled by remarking that

$$\begin{aligned} |\Upsilon_{1,n}| &= \sqrt{k_n} |\hat{\gamma}_n - \tilde{\gamma}_n| \\ &\leq \frac{1}{\sqrt{k_n}} \sum_{i=0}^{k_n-1} \left| \log \frac{\hat{Z}_{m_n-i, m_n}}{Z_{m_n-i, m_n}} \right| + \left| \log \frac{\hat{Z}_{m_n-k_n, m_n}}{Z_{m_n-k_n, m_n}} \right| \\ &\leq \sqrt{k_n} \max_{0 \leq i \leq k_n} \left| \log \frac{\hat{Z}_{m_n-i, m_n}}{Z_{m_n-i, m_n}} \right|. \end{aligned}$$

Combining Corollary 1 and Lemma 3(ii) yields

$$|\Upsilon_{1,n}| = O_{\mathbb{P}} \left( \sqrt{\frac{k_n \log n}{nh^d}} \right) = o_{\mathbb{P}}(1). \quad (33)$$

Let us now focus on  $\Upsilon_{2,n}$ . Remarking that  $m_n \sim n$  as  $n \rightarrow \infty$  in view of Lemma 3(i), it is clear that  $m_n/k_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Besides, since  $|A| \in \mathcal{RV}_\rho$ , we thus have  $A(m_n/k_n) \sim A(n/k_n)$

as  $n \rightarrow \infty$ . Therefore,  $\sqrt{k_n}A(m_n/k_n) \rightarrow 0$  as  $n \rightarrow \infty$  and, since  $Z_1, \dots, Z_n$  are iid from (2), classical results on Hill estimator apply, see for instance [22, Theorem 3.2.5], leading to

$$\Upsilon_{2,n} \xrightarrow{d} \mathcal{N}(0, \gamma^2). \quad (34)$$

The conclusion follows from (33) and (34).

(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 1 and Lemma 3(ii) together with the assumptions  $k_n \log n/(nh^d) \rightarrow 0$  and  $k_n/(n\alpha_n) \rightarrow \infty$  as  $n \rightarrow \infty$ ,

$$T_{1,1,n} = \frac{\sqrt{k_n}}{\log \left( \frac{k_n}{n\alpha_n} \right)} O_{\mathbb{P}} \left( \sqrt{\frac{\log n}{nh^d}} \right) = \sqrt{\frac{k_n \log n}{nh^d}} O_{\mathbb{P}} \left( \frac{1}{\log \left( \frac{k_n}{n\alpha_n} \right)} \right) = o_{\mathbb{P}}(1). \quad (35)$$

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

$$T_{1,2,n} = |\Upsilon_{1,n}|(1 + o_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1), \quad (36)$$

in view of (33). Collecting (35) and (36) yields

$$T_{1,n} = v_n(\log \hat{q}_{n,Z}(\alpha_n) - \log \tilde{q}_{n,Z}(\alpha_n)) = o_{\mathbb{P}}(1). \quad (37)$$

Let us now focus on  $T_{2,n}$ . As a consequence of [22, Theorem 4.3.8], Weissman estimator inherits its asymptotic distribution from Hill estimator:

$$v_n \left( \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2),$$

in view of (34). As a result,

$$T_{2,n} \xrightarrow{d} \mathcal{N}(0, \gamma^2). \quad (38)$$

The conclusion follows from (37) and (38).  $\square$



*Proof of Theorem 4.* Let  $v_n = \sqrt{k_n}/\log(k_n/n\alpha_n)$  and consider the following expansion:

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

From Theorem 1,  $\xi_n^{(a)} := \sqrt{nh^d} \left( \frac{\hat{a}_n(t_n) - a(t_n)}{b(t_n)} \right) = O_{\mathbb{P}}(1)$ ,  $\xi_n^{(b)} := \sqrt{nh^d} \left( \frac{\hat{b}_n(t_n) - b(t_n)}{b(t_n)} \right) = O_{\mathbb{P}}(1)$  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  $k_n/(nh^d) \rightarrow 0$ ,  $q_Z(\alpha_n) \rightarrow \infty$  and  $n\alpha_n/k_n \rightarrow 0$  as  $n \rightarrow \infty$ . In addition, since  $\xi_n^{(b)} = O_{\mathbb{P}}(1)$ , it follows that

$$\frac{\hat{b}_n(t_n)}{b(t_n)} = 1 + \frac{\xi_n^{(b)}}{\sqrt{nh^d}} \xrightarrow{\mathbb{P}} 1. \quad (39)$$

Besides, from Theorem 3(ii),

$$v_n \left( \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) = v_n (\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n)) (1 + o_{\mathbb{P}}(1)) \xrightarrow{d} \mathcal{N}(0, \gamma^2), \quad (40)$$

and collecting (39) and (40) yields

$$v_n \frac{\hat{b}_n(t_n)}{b(t_n)} \left( \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

The conclusion follows.  $\square$

*Proof of Theorem 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:

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

From Theorem 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  $q_Z(\alpha_n) \rightarrow \infty$  and under the assumptions  $nh^d/(k_n \log n) \rightarrow \infty$  and  $n\alpha_n/k_n \rightarrow 0$  as  $n \rightarrow \infty$ . In addition,

$$\max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} \right| \leq \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} - 1 \right| + 1 = O_{\mathbb{P}}(1), \quad (41)$$

from Theorem 2, and

$$v_n \left| \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right| = v_n |(\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n))(1 + o_{\mathbb{P}}(1))| = O_{\mathbb{P}}(1), \quad (42)$$

in view of Theorem 3(ii). Collecting (41) and (42) 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.  $\square$

## Acknowledgments

This work is supported by CEA-MITIC, Inria via the SIMERGE project, and 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 also 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.

## References

- [1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004). *Statistics of extremes: theory and applications*, John Wiley and Sons.
- [2] Berline, A., Gannoun, A. and Matzner-Løber, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles, *Statistics*, 35(2), 139–169.
- [3] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation*, Encyclopedia of Mathematics and its Applications, 27, Cambridge University Press.
- [4] Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations*, Vol. 18, OUP Oxford.
- [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.

- [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.
- [7] Daouia, A., Florens, J-P. and Simar, L. (2010). Frontier estimation and extreme value theory. *Bernoulli*, 16, 1039–1063.
- [8] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557–2589.
- [9] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20(2), 311–333.
- [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.
- [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.
- [12] Falk, M., Hüsler, J. and Reiss, R. D. (2004). *Laws of small numbers: Extremes and rare events*, 2nd edition, Birkhäuser.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101(2), 419–433.
- [17] Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17(1), 45–75.
- [18] Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2020). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *Working paper*.

- [19] 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.
- [20] Gomes, M. I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review*, 83(2), 263–292.
- [21] Gomes, M. I., and Martins, M. J. (2004). Bias reduction and explicit semi-parametric estimation of the tail index. *Journal of statistical planning and inference*, 124(2), 361–378.
- [22] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, New York, Springer.
- [23] de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1), 60–70.
- [24] Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 15, 153–167.
- [25] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3, 1163–1174.
- [26] 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.
- [27] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of American Statistical Association*, 96, 458–468.
- [28] Kyung-Joon, C. and Schucany, W. (1998). Nonparametric kernel regression estimation near endpoints. *Journal of Statistical Planning and Inference*, 66, 289–304.
- [29] Matheron, G. (1981). Sur la négligabilité du squelette et la continuité absolue des érosions. *Technical Report N-724*, Centre de Géostatistique et de Morphologie Mathématique, 1–26.
- [30] 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.

- [31] Müller, H. G. and Prewitt, K. (1993). Multiparameter bandwidth processes and adaptive surface smoothing. *Journal of Multivariate Analysis*, 47, 1–21.
- [32] Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics and Probability Letters*, 7(5), 407–412.
- [33] Serra, J. (1983). *Image analysis and mathematical morphology*. Academic Press, Inc.
- [34] 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.
- [35] Stone, C. J. (1977). Consistent nonparametric regression (with discussion). *The Annals of Statistics*, 5(4), 595–645.
- [36] 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.
- [37] Wang, H. J., Li, D. and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107, 1453–1464.
- [38] 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.