Estimation of the tail-index in a conditional location-scale family of heavy-tailed distributions

Aboubacrène Ag Ahmad, El Hadji Deme, Aliou Diop, Stéphane Girard

HAL Id: hal-02132976
https://hal.inria.fr/hal-02132976v2

Submitted on 15 Nov 2019

Estimation of the tail-index in a conditional location-scale family of heavy-tailed distributions

Aboubacrène Ag Ahmad[1], El Hadji Deme[1], Aliou Diop[1] & Stéphane Girard[2,⋆]

[1] Université Gaston Berger, LERSTAD, UFR SAT, BP 234, Saint-Louis, Sénégal.

[2] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

[⋆] Corresponding author, Stephane.Girard@inria.fr

Abstract

We introduce a location-scale model for conditional heavy-tailed distributions when the covariate is deterministic. First, nonparametric estimators of the location and scale functions are introduced. Second, an estimator of the conditional extreme-value index is derived. The asymptotic properties of the estimators are established under mild assumptions and their finite sample properties are illustrated both on simulated and real data.

Keywords: Nonparametric estimation, location-scale function, tail-index, extreme-values, conditional quantile.

1 Introduction

The literature on extreme-value analysis of independent and identically distributed observations is very elaborate, see for instance [4, 13, 27]. However, the regression point of view has been less extensively studied. The goal is to describe how tail characteristics such as extreme quantiles or small exceedance probabilities of the quantity of interest Y may depend on some explanatory variable x. Furthermore, as noted in [4, Chapter 7], such covariate information allows to combine datasets from different sources which may lead to better point estimates and thus improved inference.

A parametric approach is considered in [35] where a linear trend is fitted to the expectation of the extreme-value distribution. We also refer to [12] for other examples of parametric models. Turning to semi-parametric models, [29] proposed to mix

1

a non-parametric estimation of the trend with a parametric assumption on Y given x. Similarly, a semi-parametric estimator of γ is introduced in [3] as $\gamma(\psi(\beta'x))$ where ψ is a known link function and β is interpreted as a vector of regression coefficients. Fully non-parametric estimators have been first introduced in [7, 11] through respectively local polynomial and spline models. We also refer to [13, Theorem 3.5.2] for the approximation of the nearest neighbors distribution using the Hellinger distance and to [14] for the study of their asymptotic distribution. Focusing on the estimation of the tail-index of the conditional distribution of Y given x, moving windows and nearest neighbors approaches are developed respectively by [16, 17] in a fixed design setting. Kernels methods are proposed in [10, 9, 20, 21, 25] to tackle the random design case. Finally, these methods have been adapted to the situation where the covariate is a random field or infinite dimensional, see respectively [1] and [18, 19].

The aim of our work is to estimate in a semi-parametric way the tail-index γ in a location-scale model for conditional heavy-tailed distributions. The so-called conditional tail-index is assumed to be constant while the location and scale parameters depend on the covariate, in a fixed design setting. The underlying idea of this model is to achieve a balance between the flexibility of non-parametric approaches (for the location and scale functions) and the stability of parametric estimators (for the conditional tail-index) compared to purely non-parametric ones. This intuition has also been implemented in [31]: An extreme-value distribution with constant extreme-value index is fitted to standardized rainfall maxima. Here, we introduce a statistical framework to assess the benefits of such approaches in terms of convergence rates of the estimators.

This paper is organized as follows. The location-scale model for heavy-tailed distribution is introduced in Section 2. The associated inference procedures are described in Section 3. 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 insurance data. Proofs are postponed to the Appendix.

2 Conditional location-scale family of heavy-tailed distributions

Let Y be a real random variable. We assume that the conditional survival function of Y given $x \in [0,1]$ can be written as

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

for $y \geq y_0(x) > a(x)$. The functions $a : [0, 1] \rightarrow \mathbb{R}$ and $b : [0, 1] \rightarrow \mathbb{R}^+$ are referred to as the location and scale functions respectively while $\bar{F}_Z$ is the survival function of a real random variable Z which is assumed to be heavy-tailed:

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

Here, $\gamma > 0$ is called the conditional tail-index and ℓ 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 [6] for a detailed account on regular variations. Combining (1) and (2) yields

$$\bar{F}_Y(y \mid x) = \left(\frac{y - a(x)}{b(x)} \right)^{-1/\gamma} \ell \left(\frac{y - a(x)}{b(x)} \right), \tag{3}$$

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-scale 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 modeling offers more flexibility than purely parametric approaches. On the other hand, assuming a constant conditional tail-index γ should yield more reliable estimates in small sample contexts than purely nonparametric approaches. A similar idea is developed in [31]: An extreme-value distribution with constant extreme-value index is fitted to standardized rainfall maxima.

In the following, a fixed design setting is adopted, and thus the covariate x is supposed to be nonrandom. Model (1) can be rewritten as

$$Y = a(x) + b(x)Z, \tag{4}$$

where $x \in [0, 1]$ and Z is a random variable distributed according to (2). Starting with a n-sample $\{(Y_1, x_1), \ldots, (Y_n, x_n)\}$ from (4), 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 $\bar{F}_Z$. To this end, for all $\alpha \in (0, 1)$ consider the αth quantile of Z:

$$q_Z(\alpha) = \inf\{z \in \mathbb{R}; \bar{F}_Z(z) \leq \alpha\},$$

and assume there exist $0 < \mu_3 < \mu_2 < \mu_1 < 1$ such that

$$q_Z(\mu_2) = 0 \text{ and } q_Z(\mu_3) - q_Z(\mu_1) = 1. \tag{5}$$

3

From (4), it straightforwardly follows that, for all $\alpha \in (0,1)$, the conditional quantile of Y given $x \in [0,1]$ is

$$q_Y(\alpha \mid x) = a(x) + b(x)q_Z(\alpha), \tag{6}$$

and therefore the location and scale functions are defined in an unique way by

$$a(x) = q_Y(\mu_2 \mid x) \text{ and } b(x) = q_Y(\mu_3 \mid x) - q_Y(\mu_1 \mid x), \tag{7}$$

for all $x \in [0,1]$. This remark is the starting point of the inference procedure.

3 Inference

Let $\{(Y_1, x_1), \ldots, (Y_n, x_n)\}$ be a n-sample from (4): $Y_i = a(x_i) + b(x_i)Z_i$, $i = 1, \ldots, n$ where $Z_1, \ldots, Z_n$ are independent and identically distributed (iid) from (2). For the sake of simplicity, it is assumed that the design points are equidistant: $x_i = i/n$ for all $i = 1, \ldots, n$ and $x_0 := 0$. This assumption could be weakened to $\max_i |x_i - x_{i-1}| = O(1/n)$ used for instance in [2, 33]. A three-stage inference procedure is adopted.

(i) First, let $\hat{q}_{n,Y}(\alpha \mid x)$ be a nonparametric estimator of the conditional quantile $q_Y(\alpha \mid x)$ where $\alpha \in (0,1)$ and $x \in [0,1]$. In view of (7), the location and scale functions are estimated for all $x \in [0,1]$ by

$$\hat{a}_n(x) = \hat{q}_{n,Y}(\mu_2 \mid x) \text{ and } \hat{b}_n(x) = \hat{q}_{n,Y}(\mu_3 \mid x) - \hat{q}_{n,Y}(\mu_1 \mid x). \tag{8}$$

(ii) Second, the non-observed $Z_1, \ldots, Z_n$ can be estimated by the residuals

$$\hat{Z}_i = \frac{Y_i - \hat{a}_n(x_i)}{\hat{b}_n(x_i)}, \tag{9}$$

for all $i = 1, \ldots, n$. In practice, nonparametric estimators can suffer from boundary effects [8, 32] and therefore only design points sufficiently far from 0 and 1 are considered. Let us denote by I_n the set of indices associated with such design points and set $m_n = \text{card}(I_n)$.

(iii) 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 an Hill-type statistics:

$$\hat{\gamma}_n = \frac{1}{k_n} \sum_{i=0}^{k_n-1} \log \hat{Z}_{m_n-i,m_n} - \log \hat{Z}_{m_n-k_n,m_n}. \tag{10}$$

4

This estimator is similar to Hill estimator [30], but in our context, it is built on non iid pseudo-observations.

The proposed procedure relies on the choice of an estimator for the conditional quantiles. Here, a kernel estimator for $\bar{F}_Y(y \mid x)$ is considered (see for instance [33]). For all $(x, y) \in [0, 1] \times \mathbb{R}$ let

$$\hat{\bar{F}}_{n,Y}(y \mid x) = \sum_{i=1}^{n} \mathbb{1}_{\{Y_i > y\}} \int_{x_{i-1}}^{x_i} K_h(x - t) dt, \tag{11}$$

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

$$\hat{q}_{n,Y}(\alpha \mid x) = \hat{\bar{F}}_{n,Y}^{\leftarrow}(\alpha \mid x) := \inf\{y; \ \hat{\bar{F}}_{n,Y}(y \mid x) \le \alpha\}. \tag{12}$$

In this context, $I_n = \{\lfloor nh \rfloor, n - \lfloor nh \rfloor\}$ and $m_n = n - 2\lfloor nh \rfloor + 1$. Remark that I_n is properly defined for all large n since $h < 1/2$ eventually. Nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution function have been extensively investigated, see, for example [5, 34, 36], among others.

4 Main results

The following general assumptions are required to establish our results. The first one gathers all the conditions to define a conditional location-scale families of heavy-tailed distributions.

(A.1) $(Y_1, x_1), \ldots, (Y_n, x_n)$ are independent observations from the conditional location-scale family of heavy-tailed distributions defined by (1), (2) and (5). The functions $a(\cdot)$ and $b(\cdot)$ are continuous on $[0, 1]$ and the survival function $\bar{F}_Z(\cdot)$ is continuously differentiable on $\mathbb{R}$ with associated density $f_Z(\cdot) = -\bar{F}'_Z(\cdot)$.

Under **(A.1)**, the quantile function $q_Z(\cdot)$ exists 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. The second assumption is a Lipschitz condition on the conditional survival function of Y. Lemma 1 in Appendix provides sufficient conditions on $a(\cdot)$, $b(\cdot)$ and $\bar{F}_Z(\cdot)$ such that it is verified.

(A.2) For any compact set $C \subset \mathbb{R}$, there exists $c_1 > 0$ such that for all $(s, t) \in [0, 1]^2$

$$\sup_{y \in C} \left| \frac{\bar{F}_Y(y \mid s)}{\bar{F}_Y(y \mid t)} - 1 \right| \le c_1 |s - t|.$$

5

The next assumption is standard in the nonparametric kernel estimation framework.

(A.3) K is a bounded density with support $S \subset [-1,1]$ and verifying the Lipschitz property: There exists $c_2 > 0$ such that

$$|K(u) - K(v)| \le c_2 |u - v|$$

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 [27, eq (3.2.5)]):

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

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

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

The rationale behind **(A.4)** is the following. From [6, Theorem 1.5.12], it is clear that (2) is equivalent to $U_Z \in \mathcal{RV}_\gamma$, that is $U_Z(\lambda z)/U_Z(z) \to \lambda^\gamma$ as $z \to \infty$ for all $\lambda > 0$. The role of the second-order condition 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 [27, Lemma 2.2.3]. It is then clear that ρ, referred to as the second-order parameter, is a crucial quantity, tuning the rate of convergence of most extreme-value estimators, see [27, Chapter 3] for examples.

Our first result states the joint asymptotic normality of the estimators (8) of the location and scale parameters at a point $t_n \in (0,1)$ not too close from the boundaries of the unit interval.

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* $nh \to \infty$ *and* $nh^3 \to 0$ *as* $n \to \infty$ *then, for all sequence* $(t_n) \subset [h, 1-h]$,

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

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. Let* $I_n = \{\lfloor nh \rfloor, \ldots, n - \lfloor nh \rfloor\}$ *and suppose* $nh/\log n \to \infty$ *and* $nh^3/\log n \to 0$ *as* $n \to \infty$. *Then,*

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

Theorem 2 will reveal useful to prove that the residuals $\hat{Z}_i$ are close to the unobserved Z_i, $i = 1, \ldots, n$. This justifies the computation of the Hill estimator (10) on the residuals. Our final main result provides the asymptotic normality of this conditional tail-index estimator.

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

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

It appears that our methodology is able to estimate the tail-index in the conditional location-scale family at the same rate $1/\sqrt{k_n}$ as in iid case, see [26] for a review. As expected, the conditional location-scale family is a more favorable situation than the purely nonparametric framework for the estimation of the conditional tail index where the rate of convergence $1/\sqrt{k_n h}$ is impacted by the covariate, see [10, Corollary 1 & 2], [9, Theorem 3] and [25, Theorem 2]. To be more specific, remark first that conditions $nh/(k_n \log n) \to \infty$ and $nh^3/\log n \to 0$ imply that $k_n = o((n/\log n)^{2/3})$. Second, following [27, Eq. (3.2.10)], if A is exactly a power function, then condition $\sqrt{k_n} A(n/k_n) \to 0$ as $n \to \infty$ yields $k_n = o(n^{-2\rho/(1-2\rho)})$. Up to logarithmic factors, the constraint is then $k_n = o(n^{(-2\rho/(1-2\rho))\wedge(2/3)})$. If $\rho \geq -1$, 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 [28, Remark 3].

Let us also remark that, since $nh/(k_n \log n) \to \infty$ and since $b(\cdot)$ is lower bounded under **(A.1)**, Theorem 1 and Theorem 3 entail that

$$\sqrt{k_n} \begin{pmatrix} \hat{\gamma}_n - \gamma \\ \hat{a}_n(t_n) - a(t_n) \\ \hat{b}_n(t_n) - b(t_n) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^3}, \gamma^2 E\right),$$

where the coefficients of the matrix E are given by $E_{1,1} = 1$ and $E_{i,j} = 0$ if $i \in \{2, 3\}$ or $j \in \{2, 3\}$. The joint limiting distribution is degenerated since $\hat{\gamma}_n$ converges at a slower rate than $\hat{a}_n(t_n)$ and $\hat{b}_n(t_n)$.

5 Illustration on simulated data

The finite-sample performance of the estimators of the location and scale functions as well as of the conditional tail-index are illustrated on simulated data from model (4).

The location and scale functions are defined respectively by $a(x) = \cos(2\pi x)$ and $b(x) = 1 + x^2$ for $x \in [0, 1]$. Let Z_0 be a standard Student-t_ν random variable where $\nu \in \{1, 2, 4\}$ denotes the degrees of freedom (df). Let $\mu_1 = 3/4$, $\mu_2 = 1/2$ and $\mu_3 = 1/4$ and introduce $Z = Z_0/(2q_{Z_0}(\mu_1))$ the rescaled Student random variable. By symmetry, $q_Z(\mu_2) = 0$ and $q_Z(\mu_3) = -q_Z(\mu_1)$. Besides, $q_Z(\mu_1) = q_{Z_0}(\mu_1)/(2q_{Z_0}(\mu_1)) = 1/2$ by construction and thus (5) holds. This choice also ensures that Z is heavy-tailed and that the second-order condition **(A.4)** holds with conditional tail-index $\gamma = 1/\nu$ and conditional second-order parameter $\rho = -2/\nu$.

In all the experiments, $N = 100$ replications of a dataset of size $n = 1000$ are considered. The kernel function K is chosen to be the quartic (or biweight) kernel

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

and the bandwidth is fixed to $h = 0.1$.

We denote respectively by $\hat{a}_{n,i}(\cdot)$, $\hat{b}_{n,i}(\cdot)$ and $\hat{\gamma}_{n,i}$ the estimates of $a(\cdot)$, $b(\cdot)$ and γ obtained on the i−th replication, $i \in \{1, \ldots, N\}$. The associated mean values are also computed as

$$\bar{\hat{a}}_n(\cdot) := \frac{1}{N} \sum_{i=1}^{N} \hat{a}_{n,i}(\cdot), \ \bar{\hat{b}}_n(\cdot) := \frac{1}{N} \sum_{i=1}^{N} \hat{b}_{n,i}(\cdot) \text{ and } \bar{\hat{\gamma}}_n := \frac{1}{N} \sum_{i=1}^{N} \hat{\gamma}_{n,i}.$$

The results are depicted on Figure 1 ($\nu = 1$), Figure 2 ($\nu = 2$) and Figure 3 ($\nu = 4$). On the top-left panels (a), the true conditional quantiles $q(\mu_j|\cdot)$, $j \in \{1, 2, 3\}$ are superimposed to one replication of the simulated datasets. The estimated location and scale functions $a(\cdot)$ and $b(\cdot)$ are compared with the mean estimates $\bar{\hat{a}}_n(\cdot)$ and $\bar{\hat{b}}_n(\cdot)$ on the top-right (b) and bottom-left panels (c) respectively. Finally, the estimated conditional tail-indices $\hat{\gamma}_{n,i}$, $i = 1, \ldots, N$, the mean estimated value $\bar{\hat{\gamma}}_n$ and the true conditional tail-index are displayed as functions of $k_n \in \{1, \ldots, 300\}$ on the bottom-right panels (d). As expected, it appears on Figure 1(a)–3(a) that the tail heaviness of $Y|x$ decreases as ν increases. The estimation accuracy of the location and scale function does not seem to be sensitive to ν, see Figure 1(b,c)–3(b,c). On the contrary, it appears on Figure 1(d)–3(d) that large values of ν yield a large bias in the estimation of the conditional tail-index. This trend was expected, since the conditional second-order parameter is the main driver of the bias, as explained in Section 4, and since

8

$|\rho| = 1/(2\nu)$ for a Student distribution. Small values of $|\rho|$ in **(A.4)** entail high bias in extreme-value estimators such as Hill's statistics. A way to mitigate this bias could be to replace the conditional tail-index estimator (10) by a bias-reduced Hill-type estimators, see for instance [26].

6   Real data example

We consider here a dataset on motorcycle insurance policies and claims over the period 1994-1998 collected from the former Swedish insurance provider Wasa. The dataset is available from `www.math.su.se/GLMbook` and the R package `insuranceData`. We focus on two variables: the claim severity Y (defined as the ratio of claim cost by number of claims for each given policyholder) in SEK, and the age x of the policyholder in years. Removing missing data and an affine transformation of a covariate result in $n = 670$ pairs (x_i, Y_i) with $x_i \in [0, 1]$. Some graphical diagnostics have been performed in [23] to check that the heavy-tailed assumption makes sense for Y. Our goal is to estimate the conditional extreme quantile $q_Y(\alpha_n \mid x)$ where $n\alpha_n \to 0$ and $x \in (0, 1)$. Two estimators are considered. The first one relies on the semi-parametric model via (6):

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

and on Weissman estimator [37] applied to the pseudo-observations $\hat{Z}_i$, $i \in I_n$:

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

The second one is the nonparametric conditional Weissman estimator introduced in [10]:

$$\check{q}_{n,Y}(\alpha_n \mid x) = \hat{q}_{n,Y}(k_n/m_n \mid x) \left(\frac{\alpha_n m_n}{k_n} \right)^{-\check{\gamma}_n(x)},$$

where $\hat{q}_{n,Y}(k_n/m_n \mid x)$ is defined in (12) and $\check{\gamma}_n(x)$ is an estimator of the conditional tail index. Here, we selected a recent estimator introduced in [23] and denoted by $\hat{\gamma}_{k_n}^{(3)}(x)$ in the previously mentioned paper.

As in Section 5, we set the normalizing parameters to $\mu_1 = 3/4$, $\mu_2 = 1/2$ and $\mu_3 = 1/4$. The quartic kernel is used and the bandwidth $h = 0.065$ is chosen by the cross-validation procedure implemented in R as `h.cv`. The estimated location and scaled functions are superimposed to the dataset on Figure 4. The residuals are then computed according to (9).

To confirm that the location-scale model (3) is appropriate, Figure 5 displays a quantile-quantile plot of the weighted log-spacings within the top of the residuals

against the quantiles of the standard exponential distribution. Formally, let $W_{i,m_n} = i\log(\hat{Z}_{m_n-i+1,m_n}/\hat{Z}_{m_n-i,n})$, $1 \leq i \leq k_n - 1$, denote the weighted log-spacings computed from the consecutive top order statistics of the residuals. It is known that, if $\hat{Z}$ is heavy-tailed with tail-index γ then, the W_{i,m_n} are approximately independent copies of an exponential random variable with mean γ, see for instance [4]. Here, the number of upper statistics is fixed to $k_n = 130$ by a visual inspection of the Hill plot (not reproduced here). The relationship appearing on Figure 5 is approximately linear, which constitutes a graphical evidence that the heavy-tail assumption (2) on Z makes sense and that the choice of k_n is appropriate.

Finally, the two conditional quantile estimators $\tilde{q}_{n,Y}(\alpha_n \mid \cdot)$ and $\check{q}_{n,Y}(\alpha_n \mid \cdot)$ are graphically compared on Figure 6 for $\alpha_n = 8/n$. Both of them yield level curves with similar shapes and located above the sample. Unsurprisingly, the estimator $\tilde{q}_{n,Y}(\alpha_n \mid \cdot)$ based on the location-scale model has a smoother behavior than $\check{q}_{n,Y}(\alpha_n \mid \cdot)$ since it relies on the assumption that the tail-index does not depend on the covariate.

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

We begin by providing some sufficient conditions such that **(A.2)** holds.

Lemma 1. *If* **(A.1)** *holds and there exist* $(c_a, c_b, c_F) \in \mathbb{R}^3_+$ *and* $m_b > 0$ *such that for all* $(y, z, t, s) \in \mathbb{R}^2 \times [0,1]^2$,

$$
\begin{aligned}
m_b &\leq |b(t)|, \\
|a(t) - a(s)| &\leq c_a|t - s|, \\
|b(t) - b(s)| &\leq c_b|t - s|, \\
|\log \bar{F}_Z(y) - \log \bar{F}_Z(z)| &\leq c_F|y - z|,
\end{aligned}
$$

then **(A.2)** *holds.*

Proof. Let us first remark that, since $|a(\cdot)|$ and $|b(\cdot)|$ are continuous functions on the compact set $[0,1]$, they are necessarily upper bounded by some finite constants denoted

10

by M_a and M_b. Second, consider the quantity

$$\Delta(y, t, s) := \log \bar{F}_Y(y \mid t) - \log \bar{F}_Y(y \mid s) = \log \bar{F}_Z\left(\frac{y - a(t)}{b(t)}\right) - \log \bar{F}_Z\left(\frac{y - a(s)}{b(s)}\right).$$

The Lipschitz assumption on $\log \bar{F}_Z$ yields for all $(t, s) \in [0, 1]^2$ and $y \in \mathbb{R}$:

$$
\begin{aligned}
|\Delta(y, t, s)| &\leq c_F \left| \frac{y - a(t)}{b(t)} - \frac{y - a(s)}{b(s)} \right| \\
&= c_F \left| \frac{y(b(s) - b(t)) + a(s)(b(t) - b(s)) + b(s)(a(s) - a(t))}{b(t)b(s)} \right| \\
&\leq \frac{c_F}{m_b^2}(|y|c_b + M_a c_b + M_b c_a)|t - s|,
\end{aligned}
$$

in view of the assumptions on $a(\cdot)$ and $b(\cdot)$. Let $C \subset \mathbb{R}$ be a compact set. It follows that the supremum of $|\Delta(y, t, s)|$ on $(y, t, s) \in C \times [0, 1]^2$ is bounded and thus there exists $\tilde{c} > 0$ such that

$$\sup_{y \in C} |\exp(\Delta(y, t, s)) - 1| \leq \tilde{c} \sup_{y \in C} |\Delta(y, t, s)|.$$

Letting $M_y := \sup\{|y| \in C\}$, assumption **(A.2)** holds with $c_1 = \tilde{c}\, c_F((M_y + M_a)c_b + M_b c_a)/m_b^2$. $\qquad\square$

The next result is an adaptation of Bochner's lemma to our fixed design setting.

Lemma 2. *Let $\psi(.\mid.) : \mathbb{R}^p \times [0, 1] \to \mathbb{R}^+$, $p \geq 1$, be a positive function and C a compact subset of $\mathbb{R}^p$. For all sequences $(t_n) \subset [h, 1 - h]$ and $(y_n) \subset C$, define*

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

where $x_i = i/n$ for all $i = 0, \ldots, n$ and $Q_h(\cdot) = Q(\cdot/h)/h$, with Q is a measurable positive function with support $S \subset [-1, 1]$. If there exists $c > 0$ such that $\forall (x, s) \in [0, 1]^2$,

$$\sup_{y \in C} \left| \frac{\psi(y \mid x)}{\psi(y \mid s)} - 1 \right| \leq c|x - s|,$$

then, as $n \to \infty$,

$$\left| \frac{\psi_n(y_n \mid t_n)}{\psi(y_n \mid t_n)} - \int_S Q(u)du \right| = O\left(\frac{1}{n}\right) + O(h).$$

11

Proof. Consider the expansion

$$\frac{\psi_n(y_n \mid t_n)}{\psi(y_n \mid t_n)} - \int_S Q(u)du \;=\; \frac{\sum_{i=1}^n \psi(y_n \mid x_i) \int_{x_{i-1}}^{x_i} Q_h(t_n - s)ds}{\psi(y_n \mid t_n)} - \int_S Q(u)du$$

$$= \; \frac{\int_0^1 \psi(y_n \mid s)Q_h(t_n - s)ds - \psi(y_n \mid t_n) \int_S Q(u)du}{\psi(y_n \mid t_n)}$$

$$+ \; \frac{\sum_{i=1}^n \psi(y_n \mid x_i) \int_{x_{i-1}}^{x_i} Q_h(t_n - s)ds - \int_0^1 \psi(y_n \mid s)Q_h(t_n - s)ds}{\psi(y_n \mid t_n)}$$

$$=: \; \frac{T_{n,1}}{\psi(y_n \mid t_n)} + \frac{T_{n,2}}{\psi(y_n \mid t_n)},$$

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

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

Since $(t_n) \subset [h, 1-h]$, it follows that $S \subset [-1, 1] \subset \left[\dfrac{t_n - 1}{h}, \dfrac{t_n}{h}\right]$ and therefore

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

As a consequence, for all $y_n \in C$,

$$\left| \frac{T_{n,1}}{\psi(y_n \mid t_n)} \right| \leq \int_S \left| \frac{\psi(y_n \mid t_n - uh)}{\psi(y_n \mid t_n)} - 1 \right| Q(u)du \leq ch \int_S |u|Q(u)du = O(h). \qquad (13)$$

Let us now turn to the second term

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

We have, for all $y_n \in C$,

$$\left| \frac{T_{n,2}}{\psi(y_n \mid t_n)} \right| \; \leq \; \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \frac{\psi(y_n \mid s)}{\psi(y_n \mid t_n)} \left| \frac{\psi(y_n \mid x_i)}{\psi(y_n \mid s)} - 1 \right| Q_h(t_n - s)ds$$

$$\leq \; \frac{c}{\psi(y_n \mid t_n)} \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \psi(y_n \mid s)|x_i - s|Q_h(t_n - s)ds$$

$$\leq \; \frac{c}{n\psi(y_n \mid t_n)} \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \psi(y_n \mid s)Q_h(t_n - s)ds$$

$$= \; \frac{c}{n\psi(y_n \mid t_n)} \int_0^1 \psi(y_n \mid s)Q_h(t_n - s)ds$$

$$= \; \frac{c}{n} \left(\frac{T_{n,1}}{\psi(y_n \mid t_n)} + \int_S Q(u)du \right)$$

$$= \; O\left(\frac{1}{n}\right), \qquad (14)$$

in view of (13). Finally, collecting (13) and (14), the conclusion follows. $\qquad \square$

12

As a consequence of Lemma 2, the asymptotic bias and variance of the estimator (11) of the conditional survival function can be derived.

Lemma 3. *Suppose* **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset [h, 1-h]$ *and* $(y_n) \subset C$, *where* C *is a compact subset of* $\mathbb{R}$, *be two nonrandom sequences.*

(i) *Then,*

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

(ii) *If, moreover,* $nh \to \infty$ *as* $n \to \infty$ *and* $\liminf F_Y(y_n \mid t_n) > 0$, *then*

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

Proof. (i) Remarking that

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

the conclusion follows from Lemma 2.

(ii) Let us consider the expansion:

$$\begin{aligned}
var\left(\hat{\bar{F}}_{n,Y}(y_n \mid t_n) \right) &= \sum_{i=1}^n var\left(\mathbb{1}_{\{Y_i > y_n\}} \int_{x_{i-1}}^{x_i} K_h(t_n - s)ds \right) \\
&= \sum_{i=1}^n \bar{F}_Y(y_n \mid x_i) S_{n,i} - \sum_{i=1}^n \bar{F}_Y^2(y_n \mid x_i) S_{n,i} \\
&=: T_{n,1} - T_{n,2},
\end{aligned}$$

where

$$S_{n,i} := \left(\int_{x_{i-1}}^{x_i} K_h(t_n - s)ds \right)^2 = \frac{1}{h^2} \int_{x_{i-1}}^{x_i} \int_{x_{i-1}}^{x_i} K\left(\frac{t_n - s_1}{h} \right) K\left(\frac{t_n - s_2}{h} \right) ds_1 ds_2. \tag{15}$$

Let us write

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

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

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

13

uniformly on $(s_1, s_2) \in [x_{i-1}, x_i]^2$ and $i = 1, \ldots, n$. It thus follows that

$$
\begin{aligned}
S_{n,i} &= \frac{1}{h^2} \int_{x_{i-1}}^{x_i} \int_{x_{i-1}}^{x_i} \left[K^2 \left(\frac{t_n - s_1}{h} \right) + K \left(\frac{t_n - s_1}{h} \right) O \left(\frac{1}{nh} \right) \right] ds_1 ds_2 \\
&= \frac{1}{nh^2} \int_{x_{i-1}}^{x_i} K^2 \left(\frac{t_n - s}{h} \right) ds + O \left(\frac{1}{n^2 h^3} \right) \int_{x_{i-1}}^{x_i} K \left(\frac{t_n - s}{h} \right) ds.
\end{aligned}
$$

Defining $M(v) = K^2(v) / \|K\|_2^2$ yields

$$
S_{n,i} = \frac{\|K\|_2^2}{nh} \int_{x_{i-1}}^{x_i} M_h(t_n - s) ds + O \left(\frac{1}{n^2 h^2} \right) \int_{x_{i-1}}^{x_i} K_h(t_n - s) ds. \tag{16}
$$

Replacing in $T_{n,1}$, we obtain:

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

Applying Lemma 2 twice and recalling that $nh \to \infty$ as $n \to \infty$ entail

$$
T_{n,1} = \frac{\|K\|_2^2}{nh} \bar{F}_Y(y_n \mid t_n) \left(1 + O(h) + O \left(\frac{1}{nh} \right) \right).
$$

Similarly,

$$
T_{n,2} = \frac{\|K\|_2^2}{nh} \bar{F}_Y^2(y_n \mid t_n) \left(1 + O(h) + O \left(\frac{1}{nh} \right) \right),
$$

and the conclusion follows:

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

under the assumption $\liminf F_Y(y_n \mid t_n) > 0$. $\qquad \square$

The next lemma controls the error between each unobserved random variable Z_i and its estimation $\hat{Z}_i$, for all $i = 1, \ldots, n$.

Lemma 4. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $I_n = \{\lfloor nh \rfloor, \ldots, n - \lfloor nh \rfloor\}$ *and suppose* $nh / \log n \to \infty$ *and* $nh^3 / \log n \to 0$ *as* $n \to \infty$. *Then, for all* $i \in I_n$,

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

14

Proof. 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 $i \in I_n$,

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

On the one hand, Theorem 2 entails

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

On the other hand,

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

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

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

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

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

15

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

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

which completes the proof of the lemma. $\qquad\square$

Finally, Lemma 5 is an adaptation of [22, Proposition 1]. It permits to derive the error made on the estimation of the order statistics Z_{m_n-i,m_n}, $i = 0, \ldots, m_n - 1$ from the error made on the unsorted Z_i, $i \in I_n$.

Lemma 5. *Let $I_n = \{\lfloor nh \rfloor, \ldots, n - \lfloor nh \rfloor\}$ and $m_n = \mathrm{card}(I_n)$. 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. Remarking that $m_n = n - 2\lfloor nh \rfloor + 1 \sim n$ as $n \to \infty$ and (2) entails that the distribution of Z has an infinite upper endpoint, the conclusion follows by applying [22, Proposition 1]. $\qquad\square$

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 (11) of the conditional survival function evaluated at points depending on n.

Proposition 1. *Assume (A.1), (A.2) and (A.3) hold. Let $(t_n) \subset [h, 1-h]$ and $(\alpha_j)_{j=1,\ldots,J}$ a strictly decreasing sequence in $(0,1)$. For all $j \in \{1, \ldots, J\}$, define $y_{j,n} = q_Y(\alpha_j \mid t_n) + b(t_n)\epsilon_{j,n}$, where $\epsilon_{j,n} \to 0$ as $n \to \infty$. If in addition $nh \to \infty$ and $nh^3 \to 0$ as $n \to \infty$, then*

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

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

Proof. Let us first remark that, for all $j \in \{1, \ldots, J\}$, in view of (6), the sequence $y_{j,n} = a(t_n) + b(t_n)(q_Z(\alpha_j) + \epsilon_{j,n})$ is bounded since $\epsilon_{j,n} \to 0$ as $n \to \infty$ and since $a(\cdot)$ and $b(\cdot)$ are continuous functions defined on compact sets. Besides, from (1),

16

$F_Y(y_{j,n} \mid t_n) = F_Z(q_Z(\alpha_j) + \epsilon_{j,n}) \to 1 - \alpha_j > 0$ as $n \to \infty$ and thus the assumptions of Lemma 3(i,ii) are satisfied. Let $\beta \neq 0$ in $\mathbb{R}^J$, $J \geq 1$ and consider the random variable

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

Let us first consider the random term:

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

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

$$
\begin{aligned}
var(\Gamma_{n,1}) &= \sum_{i=1}^{n} var(T_{i,n}) \\
&= \sum_{i=1}^{n} \left(\int_{x_{i-1}}^{x_i} K_h(t_n - s)ds \right)^2 var\left(\sum_{j=1}^{J} \beta_j \mathbb{1}_{\{Y_i > y_{j,n}\}} \right) \\
&=: \sum_{i=1}^{n} S_{n,i} \, \beta^t \Sigma^{(i,n)} \beta, \tag{17}
\end{aligned}
$$

where $S_{n,i}$ is defined by (15) in the proof of Lemma 3, and where $\Sigma^{(i,n)}$ is the matrix whose coefficients are defined for $(k,l) \in \{1, \ldots, J\}^2$ by $\Sigma_{k,l}^{(i,n)} = cov\left(\mathbb{1}_{\{Y_i > y_{k,n}\}}, \mathbb{1}_{\{Y_i > y_{l,n}\}} \right)$. In view of (16),

$$
S_{n,i} = \frac{\|K\|_2^2}{nh} \int_{x_{i-1}}^{x_i} M_h(t_n - s)ds + O\left(\frac{1}{n^2 h^2} \right) \int_{x_{i-1}}^{x_i} K_h(t_n - s)ds
$$

and

$$
\begin{aligned}
\Sigma_{k,l}^{(i,n)} &= \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{k,n} \vee y_{l,n}\}} \right) - \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{k,n}\}} \right) \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{l,n}\}} \right) \\
&= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) - \bar{F}_Y(y_{k,n} \mid x_i)\bar{F}_Y(y_{l,n} \mid x_i) \\
&= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i) - \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i)\bar{F}_Y(y_{k,n} \wedge y_{l,n} \mid x_i) \\
&= \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid x_i)F_Y(y_{k,n} \wedge y_{l,n} \mid x_i) \\
&=: \varphi(y_{k,n}, y_{l,n} \mid x_i), \tag{18}
\end{aligned}
$$

17

where φ is the function $\mathbb{R}^2 \times [0,1] \to [0,1]$ defined by $\varphi(.,. \mid .) = \bar{F}_Y(. \vee . \mid .) F_Y(. \wedge . \mid .)$. Replacing in (17) yields $var(\Gamma_{n,1}) = \beta^t C^{(n)} \beta$, where $C^{(n)}$ is the covariance matrix whose coefficients are defined by

$$
\begin{aligned}
C_{k,l}^{(n)} &= \frac{\|K\|_2^2}{nh} \sum_{i=1}^{n} \varphi(y_{k,n}, y_{l,n} \mid x_i) \int_{x_{i-1}}^{x_i} M_h(t_n - s) ds \\
&+ O\left(\frac{1}{n^2 h^2}\right) \sum_{i=1}^{n} \varphi(y_{k,n}, y_{l,n} \mid x_i) \int_{x_{i-1}}^{x_i} K_h(t_n - s) ds.
\end{aligned}
$$

Applying Lemma 2 twice and recalling that $nh \to \infty$ entail

$$
\begin{aligned}
C_{k,l}^{(n)} &= \frac{\|K\|_2^2}{nh} \varphi(y_{k,n}, y_{l,n} \mid t_n)\, (1 + O(h)) + O\left(\frac{1}{n^2 h^2}\right) \varphi(y_{k,n}, y_{l,n} \mid t_n)\, (1 + O(h)) \\
&= \frac{\|K\|_2^2}{nh} \varphi(y_{k,n}, y_{l,n} \mid t_n)(1 + o(1)).
\end{aligned}
$$

As a result,

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

where

$$
B_{k,l}^{(n)} = \varphi(y_{k,n}, y_{l,n} \mid t_n) = \bar{F}_Y(y_{k,n} \vee y_{l,n} \mid t_n) F_Y(y_{k,n} \wedge y_{l,n} \mid t_n).
$$

Let us remark that, in view of (6),

$$
\begin{aligned}
y_{k,n} - y_{l,n} &= q_Y(\alpha_k \mid t_n) - q_Y(\alpha_l \mid t_n) + b(t_n)(\epsilon_{k,n} - \epsilon_{l,n}) \\
&= b(t_n)(q_Z(\alpha_k) - q_Z(\alpha_l) + \epsilon_{k,n} - \epsilon_{l,n}) \\
&\sim b(t_n)(q_Z(\alpha_k) - q_Z(\alpha_l)),
\end{aligned}
$$

as $n \to \infty$. Thus, 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

$$
B_{k,l}^{(n)} = \bar{F}_Y(y_{k \vee l, n} \mid t_n) F_Y(y_{k \wedge l, n} \mid t_n).
$$

From (1) and (6), we have

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

in view of the continuity of $\bar{F}_Z$. As a result, $B_{k,l}^{(n)} \to B_{k,l} = \alpha_{k \vee l}(1 - \alpha_{k \wedge l})$ as $n \to \infty$ and therefore

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

18

The proof of the asymptotic normality of $\Gamma_{n,1}$ is based on Lyapounov criteria for triangular arrays of independent random variables:

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \bigg/ var(\Gamma_{n,1}) \to 0 \tag{19}$$

as $n \to \infty$. Let us first remark that, for all $i = 1, \ldots, n$, the random variable $T_{i,n}$ is bounded:

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

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

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

and it is thus clear that (19) holds under the assumption $nh \to \infty$. A a result,

$$\sqrt{nh}\Gamma_{n,1} \xrightarrow{d} \mathcal{N}\left(0, \|K\|_2^2 \, \beta^t B\beta\right). \tag{20}$$

Let us now turn to the nonrandom term

$$\Gamma_{n,2} = \sum_{j=1}^{J} \beta_j \bar{F}_Y(y_{j,n} \mid t_n) \left[\frac{\mathbb{E}\left[\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) \right]}{\bar{F}_Y(y_{j,n} \mid t_n)} - 1 \right].$$

Lemma 3(i) together with the assumptions $nh^3 \to 0$ and $nh \to \infty$ as $n \to \infty$ entail

$$\sqrt{nh}|\Gamma_{n,2}| \leq \sqrt{nh} \sum_{j=1}^{J} |\beta_j| \left| \frac{\mathbb{E}\left[\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) \right]}{\bar{F}_Y(y_{j,n} \mid t_n)} - 1 \right| = O(\sqrt{nh^3}) = o(1). \tag{21}$$

Finally, collecting (20) and (21), $\sqrt{nh}\Gamma_n$ converges to a centered Gaussian random variable with variance $\|K\|_2^2 \, \beta^t B\beta$, and the result follows. $\qquad \square$

The following proposition provides the joint asymptotic normality of the estimator (12) of conditional quantiles. It can be read as an adaptation of classical results [5, 34, 36] to the location-scale setting.

Proposition 2. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset [h, 1-h]$ *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* $nh \to \infty$ *and* $nh^3 \to 0$ *as* $n \to \infty$, *then*

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

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 $(s_1, \dots, s_J) \in \mathbb{R}^J$, $\nu_{j,n} := s_j b(t_n)/\sqrt{nh}$ for all $j = 1, \dots, J$ and consider:

$$
\begin{aligned}
W_n(s_1, \dots, s_J) &= \mathbb{P}\left(\bigcap_{j=1}^{J} \left\{ \frac{\sqrt{nh}}{b(t_n)} \left(\hat{q}_{n,Y}(\alpha_j \mid t_n) - q_Y(\alpha_j \mid t_n) \right) \le s_j \right\} \right) \\
&= \mathbb{P}\left(\bigcap_{j=1}^{J} \left\{ \hat{q}_{n,Y}(\alpha_j \mid t_n) \le q_Y(\alpha_j \mid t_n) + \nu_{j,n} \right\} \right) \\
&= \mathbb{P}\left(\bigcap_{j=1}^{J} \left\{ V_{j,n} \le v_{j,n} \right\} \right),
\end{aligned}
$$

where, for $j = 1, \dots, J$,

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

Let us first examine the nonrandom term $v_{j,n}$. In view of (1) and (6), it follows that

$$
\begin{aligned}
\bar{F}_Y(q_Y(\alpha_j \mid t_n) + \nu_{j,n} \mid t_n) &= \bar{F}_Z\left(\frac{q_Y(\alpha_j \mid t_n) + \nu_{j,n} - a(t_n)}{b(t_n)} \right) \\
&= \bar{F}_Z\left(q_Z(\alpha_j) + \frac{s_j}{\sqrt{nh}} \right).
\end{aligned}
$$

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

$$\bar{F}_Z\left(q_Z(\alpha_j) + \frac{s_j}{\sqrt{nh}} \right) = \alpha_j - \frac{s_j}{\sqrt{nh}} f_Z\left(q_Z(\alpha_j) + \frac{s_j \theta_{j,n}}{\sqrt{nh}} \right).$$

In view of the continuity of $f_Z(\cdot)$ and since $s_j/\sqrt{nh} \to 0$ as $n \to \infty$, it follows that

$$f_Z\left(q_Z(\alpha_j) + \frac{s_j \theta_{j,n}}{\sqrt{nh}} \right) = \frac{1 + o(1)}{H_Z(\alpha_j)},$$

20

leading to

$$v_{j,n} = \frac{s_j}{H_Z(\alpha_j)}(1 + o(1)). \tag{22}$$

Let us now turn to the random variable $V_{j,n}$. For all $j = 1, \ldots, J$, let

$$y_{j,n} = q_Y(\alpha_j \mid t_n) + \nu_{j,n} = q_Y(\alpha_j \mid t_n) + b(t_n)\frac{s_j}{\sqrt{nh}} =: q_Y(\alpha_j \mid t_n) + b(t_n)\epsilon_{j,n},$$

where $\epsilon_{j,n} \to 0$ as $n \to \infty$. Then, Proposition 1 entails that

$$\left\{ \sqrt{nh} \left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \bar{F}_Y(y_{j,n} \mid t_n) \right) \right\}_{j=1,\ldots,J} = \{V_{j,n}\}_{j=1,\ldots,J}$$

converges to a centered Gaussian random variable with covariance matrix $\|K\|_2^2 \, B$. Taking account of (22) yields that W_n converges to the cumulative distribution function of a centered Gaussian distribution with covariance matrix $\|K\|_2^2 \, C$, evaluated at $(s_1, \ldots, s_J)$, which is the desired result. $\qquad\square$

The following proposition provides a uniform consistency result for the estimator (12) of conditional quantiles of Y given a sequence of design points (not too close from the boundaries 0 and 1).

Proposition 3. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $I_n = \{\lfloor nh \rfloor, \ldots, n - \lfloor nh \rfloor\}$ *and suppose* $nh/\log n \to \infty$ *and* $nh^3/\log n \to 0$ *as* $n \to \infty$. *Then, for all* $\alpha \in (0,1)$,

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

Proof. Let $\epsilon \in (0,1)$ and $\alpha \in (0,1)$. Define $v_n = (nh/\log n)^{1/2}$,

$$M(\epsilon, \alpha) = 2\|K\|_2 H_Z(\alpha) \left(\alpha(1-\alpha) \left(1 - \log(\epsilon/2) \right) \right)^{1/2},$$

21

and for all $i \in I_n$ let $q_{i,n}^{\pm} = q_Y(\alpha \mid x_i) \pm M(\epsilon, \alpha) b(x_i)/v_n$. Let us consider the expansion:

$$\delta_n := \mathbb{P}\left(v_n \max_{i \in I_n} \left| \frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)} \right| \geq M(\epsilon, \alpha)\right)$$

$$= \mathbb{P}\left(\bigcup_{i \in I_n} \left| \frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)} \right| \geq M(\epsilon, \alpha)/v_n\right)$$

$$= \mathbb{P}\left(\bigcup_{i \in I_n} \left\{\hat{q}_{n,Y}(\alpha \mid x_i) \geq q_{i,n}^{+}\right\} \cup \left\{\hat{q}_{n,Y}(\alpha \mid x_i) \leq q_{i,n}^{-}\right\}\right)$$

$$= \mathbb{P}\left(\bigcup_{i \in I_n} \left\{\alpha \leq \hat{\bar{F}}_{n,Y}\left(q_{i,n}^{+} \mid x_i\right)\right\} \cup \left\{\alpha \geq \hat{\bar{F}}_{n,Y}\left(q_{i,n}^{-} \mid x_i\right)\right\}\right)$$

$$\leq \mathbb{P}\left(\bigcup_{i \in I_n} \left\{\alpha - \mathbb{E}\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{+} \mid x_i\right) \leq (\hat{\bar{F}}_{n,Y} - \mathbb{E}\hat{\bar{F}}_{n,Y})(q_{i,n}^{+} \mid x_i)\right\}\right)$$

$$+ \mathbb{P}\left(\bigcup_{i \in I_n} \left\{\alpha - \mathbb{E}\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{-} \mid x_i\right) \geq (\hat{\bar{F}}_{n,Y} - \mathbb{E}\hat{\bar{F}}_{n,Y})(q_{i,n}^{-} \mid x_i)\right\}\right)$$

$$=: \mathbb{P}\left(\bigcup_{i \in I_n} \alpha_{i,n}^{+} \leq \xi_{i,n}^{+}\right) + \mathbb{P}\left(\bigcup_{i \in I_n} \alpha_{i,n}^{-} \geq \xi_{i,n}^{-}\right)$$

$$=: \delta_n^{+} + \delta_n^{-}.$$

Let us focus on the term δ_n^{+}. Assumption $nh/\log n \to \infty$ entails that $v_n \to \infty$ as $n \to \infty$ and thus $q_{i,n}^{+}$ is bounded. Therefore Lemma 3(i) yields

$$\alpha_{i,n}^{+} := \alpha - \mathbb{E}\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{+} \mid x_i\right) = \alpha - \bar{F}_Y\left(q_{i,n}^{+} \mid x_i\right)(1 + O(h))$$

$$= \bar{F}_Z(q_Z(\alpha)) - \bar{F}_Z\left(q_Z(\alpha) + \frac{M(\epsilon, \alpha)}{v_n}\right)(1 + O(h))$$

$$= \frac{M(\epsilon, \alpha)}{v_n} f_Z\left(q_Z(\alpha) + \frac{M(\epsilon, \alpha)}{v_n}\theta\right) + O(h),$$

for some $\theta \in (0, 1)$. Since $f_Z(.)$ is continuous, it follows that

$$\alpha_{i,n}^{+} = \frac{M(\epsilon, \alpha)}{v_n H_Z(\alpha)}(1 + o(1)) + O(h) =: \frac{\kappa_1(\epsilon, \alpha)}{v_n}(1 + o(1)),$$

in view of the assumption $nh^3/\log n \to 0$ as $n \to \infty$. As a preliminary result,

$$\delta_n^{+} = \mathbb{P}\left(\bigcup_{i \in I_n} \xi_{i,n}^{+} \geq \frac{\kappa_1(\epsilon, \alpha)}{v_n}(1 + o(1))\right) \leq \sum_{i \in I_n} \mathbb{P}\left(\xi_{i,n}^{+} \geq \frac{\kappa_1(\epsilon, \alpha)}{v_n}(1 + o(1))\right). \quad (23)$$

In addition,

$$\mathbb{P}\left(\xi_{i,n}^{+} \geq \frac{\kappa_1(\epsilon, \alpha)}{v_n}(1 + o(1))\right) = \mathbb{P}\left((\hat{\bar{F}}_{n,Y} - \mathbb{E}\hat{\bar{F}}_{n,Y})(q_{i,n}^{+} \mid x_i) \geq \frac{\kappa_1(\epsilon, \alpha)}{v_n}(1 + o(1))\right)$$

$$:= \mathbb{P}\left(\sum_{j=1}^{n} \tilde{X}_j \geq \frac{\kappa_1(\epsilon, \alpha)}{v_n}(1 + o(1))\right), \quad (24)$$

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

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

are independent, centered and bounded:

$$|\tilde{X}_j| \leq \int_{x_{j-1}}^{x_j} K_h(x_i - s)ds \leq \frac{\|K\|_\infty}{nh}.$$

Lemma 3(ii) entails

$$\sum_{j=1}^n \mathbb{E}(\tilde{X}_j^2) = var\left(\sum_{j=1}^n \tilde{X}_j\right) = var\left[\hat{\bar{F}}_{n,Y}\left(q_{i,n}^+ \mid x_i\right)\right]$$

$$= \frac{\bar{F}_Y\left(q_{i,n}^+ \mid x_i\right) F_Y\left(q_{i,n}^+ \mid x_i\right)}{nh} \|K\|_2^2 (1 + o(1)).$$

Besides, $q_{i,n}^+ \to q_Y(\alpha \mid x_i)$ as $n \to \infty$ 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 $\bar{F}_Y(\cdot \mid x_i)$. It follows that,

$$\sum_{j=1}^n \mathbb{E}(\tilde{X}_j^2) = \frac{\alpha(1-\alpha)}{nh} \|K\|_2^2 (1 + o(1)) =: \frac{\kappa_2(\alpha)}{nh}(1 + o(1)).$$

Applying Bernstein's inequality for bounded random variables yields

$$
\begin{aligned}
(24) \quad &\leq \exp\left(-\frac{\frac{\kappa_1^2(\epsilon, \alpha) \log n (1 + o(1))}{2nh}}{\frac{\kappa_2(\alpha)(1 + o(1))}{nh} + \frac{\kappa_1(\epsilon, \alpha)(1 + o(1))}{3nhv_n}}\right) \\
&= \exp\left(-\frac{\kappa_1^2(\epsilon, \alpha) \log n}{2\kappa_2(\alpha) + \frac{2\kappa_1(\epsilon, \alpha)(1 + o(1))}{3v_n}}(1 + o(1))\right) \\
&= \exp\left(-\frac{\kappa_1^2(\epsilon, \alpha) \log n}{2\kappa_2(\alpha)}(1 + o(1))\right) \\
&= \exp\left[-2\left(1 - \log(\epsilon/2)\right) \log n \left(1 + o(1)\right)\right] \\
&\leq \exp\left[-\left(1 - \log(\epsilon/2)\right) \log n\right], \quad (25)
\end{aligned}
$$

for n large enough. Collecting (23)-(25) yields

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

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

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

which is the desired result. $\qquad \square$

7.3 Proofs of main results

The proof of Theorem 1 directly relies on Proposition 2:

Proof of Theorem 1. Let us remark that

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

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

Applying Proposition 2 with $J = 3$, $\alpha_1 = \mu_1$, $\alpha_2 = \mu_2$ and $\alpha_3 = \mu_3$ yields

$$\xi_n \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^3}, \|K\|_2^2 \, C\right),$$

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,

$$\tilde{A}\xi_n \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^2}, \|K\|_2^2 \, \tilde{A}C\tilde{A}^t\right),$$

and the conclusion follows from standard calculations. □

Theorem 2 is a straightforward consequence of Proposition 3:

Proof of Theorem 2. Remarking that

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

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

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

and the conclusion follows from Proposition 3 successively applied with $\alpha = \mu_3$ and $\alpha = \mu_1$. □

24

Proof of Theorem 3. 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 $Z_1,\ldots,Z_n$. The first term is controlled by remarking that

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

Combining Lemma 4 and Lemma 5 yields

$$|\Upsilon_{1,n}| = O_{\mathbb{P}}\left(\sqrt{\frac{k_n \log n}{nh}}\right) = o_{\mathbb{P}}(1), \tag{26}$$

in view of the assumption $nh/(k_n \log n) \to \infty$ as $n \to \infty$. Let us now focus on $\Upsilon_{2,n}$. Remarking that $m_n \sim n$ as $n \to \infty$ 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, $\sqrt{k_n}A(m_n/k_n) \to 0$ as $n \to \infty$ and, since $Z_1,\ldots,Z_n$ are iid from (2), classical results on Hill estimator apply, see for instance [27, Theorem 3.2.5], leading to

$$\Upsilon_{2,n} \xrightarrow{d} \mathcal{N}(0,\gamma^2). \tag{27}$$

The conclusion follows by combining (26) and (27). $\square$

Acknowledgments

References

[1] Abdi, S., Abdi, A., Dabo-Niang, S. and Diop, A. (2010). Consistency of a nonparametric conditional quantile estimator for random fields. *Mathematical Methods of Statistics*, 19(1), 1–21.

[2] Antoch, J. and Janssen, P. (1989). Nonparametric regression M-quantiles. *Statistics and Probability Letters*, 8(4), 355–362.

[3] Beirlant, J. and Goegebeur, Y. (2003). Regression with response distributions of Pareto-type. *Computational Statistics and Data Analysis*, 42, 595–619, 2003.

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

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

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

[7] 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.

[8] 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.

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

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

[11] 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.

[12] 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.

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

[14] Gangopadhyay, A. K. (1995). A note on the asymptotic behavior of conditional extremes, *Statistics and Probability Letters*, 25, 163–170.

[15] Gardes, L. (2015). A general estimator for the extreme value index: applications to conditional and heteroscedastic extremes. *Extremes*, 18(3), 479–510.

[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.

[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.

[18] Gardes, L. and Girard, S. (2012). Functional kernel estimators of large conditional quantiles, *Electronic Journal of Statistics*, 6, 1715–1744.

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

[20] Gardes, L., Guillou, A. and Schorgen, A. (2012). Estimating the conditional tail index by integrating a kernel conditional quantile estimator, *Journal of Statistical Planning and Inference*, 142(6), 1586–1598.

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

[22] Girard, S. and Stupfler, G. (2019). Estimation of high-dimensional extreme conditional expectiles. *Working paper.*

[23] Girard, S., Stupfler, G. and A. Usseglio-Carleve (2019). Nonparametric extreme conditional expectile estimation. *Submitted*, http://hal.inria.fr/hal-02114255.

[24] Goegebeur, Y., Guillou, A. and Osmann, M. (2014). A local moment type estimator for the extreme value index in regression with random covariates. *The Canadian Journal of Statistics*, 42, 487–507.

[25] 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.

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

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

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

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

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

[31] 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.

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

[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.

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

[35] 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.

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

[37] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, 73, 812–815.
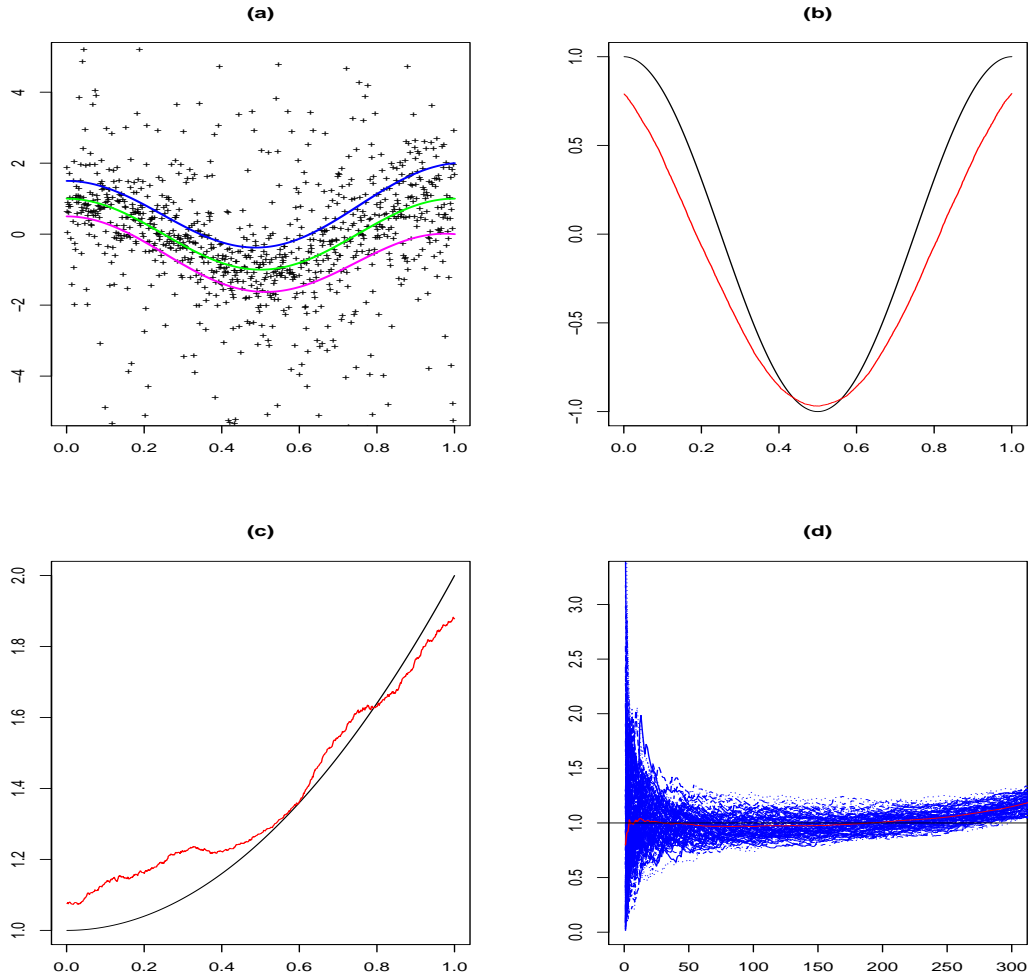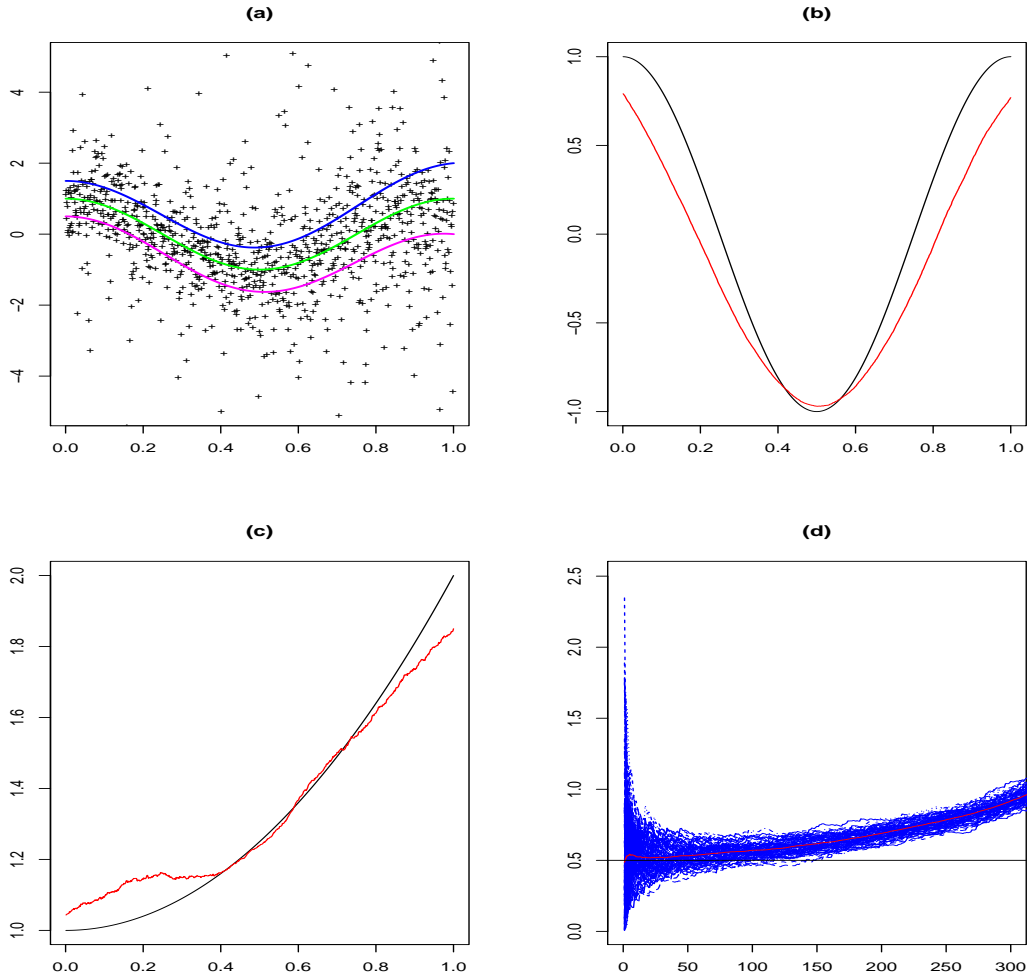
Figure 1: Student distribution with $\nu = 1$ df. (a): Simulated data (+) and conditional quantiles $q(3/4|\cdot)$ (magenta), $q(1/2|\cdot)$ (green) and $q(1/4|\cdot)$ (blue). (b): Location function $a(\cdot)$ (black) and mean estimate $\bar{\hat{a}}_n(\cdot)$ (red). (c): Scale function $b(\cdot)$ (black) and mean estimate $\bar{\hat{b}}_n(\cdot)$ (red). (d): Conditional tail-index γ (black), estimates $\hat{\gamma}_{n,i}$, $i = 1, \ldots, N$ (blue) and mean estimate $\bar{\hat{\gamma}}_n$ (red) as functions of $k_n \in \{1, \ldots, 300\}$.
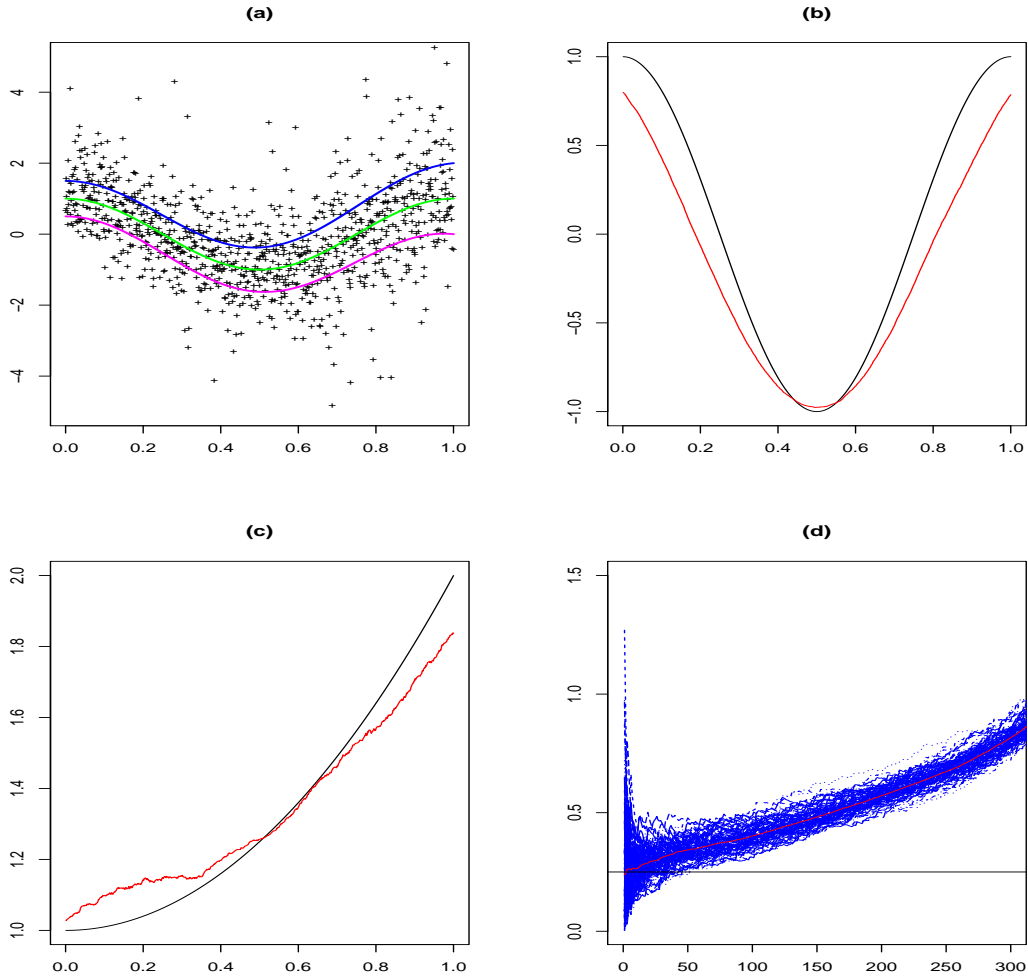
Figure 2: Student distribution with $\nu = 2$ df. (a): Simulated data (+) and conditional quantiles $q(3/4|\cdot)$ (magenta), $q(1/2|\cdot)$ (green) and $q(1/4|\cdot)$ (blue). (b): Location function $a(\cdot)$ (black) and mean estimate $\bar{\hat{a}}_n(\cdot)$ (red). (c): Scale function $b(\cdot)$ (black) and mean estimate $\bar{\hat{b}}_n(\cdot)$ (red). (d): Conditional tail-index γ (black), estimates $\hat{\gamma}_{n,i}$, $i = 1, \ldots, N$ (blue) and mean estimate $\bar{\hat{\gamma}}_n$ (red) as functions of $k_n \in \{1, \ldots, 300\}$.

Figure 3: Student distribution with $\nu = 4$ df. (a): Simulated data ($+$) and conditional quantiles $q(3/4|\cdot)$ (magenta), $q(1/2|\cdot)$ (green) and $q(1/4|\cdot)$ (blue). (b): Location function $a(\cdot)$ (black) and mean estimate $\bar{\hat{a}}_n(\cdot)$ (red). (c): Scale function $b(\cdot)$ (black) and mean estimate $\bar{\hat{b}}_n(\cdot)$ (red). (d): Conditional tail-index γ (black), estimates $\hat{\gamma}_{n,i}$, $i = 1, \ldots, N$ (blue) and mean estimate $\bar{\hat{\gamma}}_n$ (red) as functions of $k_n \in \{1, \ldots, 300\}$.
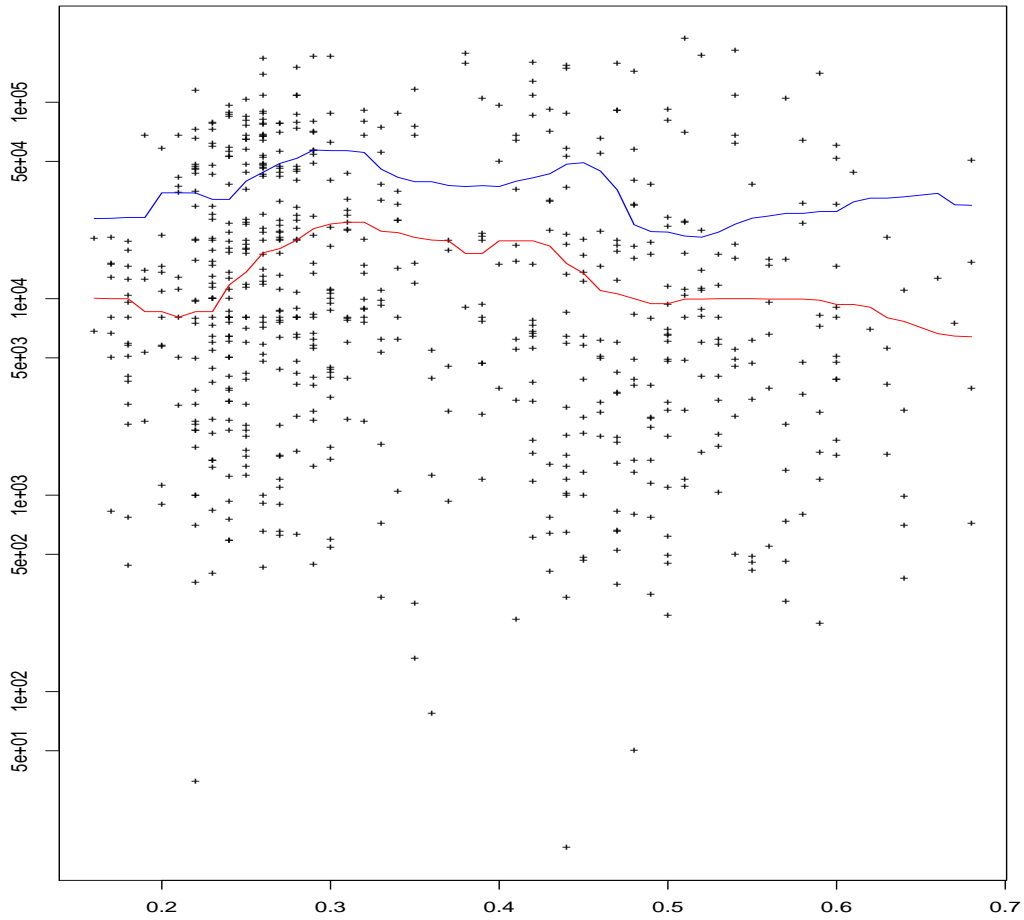
Figure 4: Illustration on motorcycle insurance data. Horizontally: Age of the policyholder (hundred of years), vertically: Claim severity (SEK, log scale). Data (+), estimated location function $\hat{a}_n(\cdot)$ (red) and scale function $\hat{b}_n(\cdot)$ (blue).
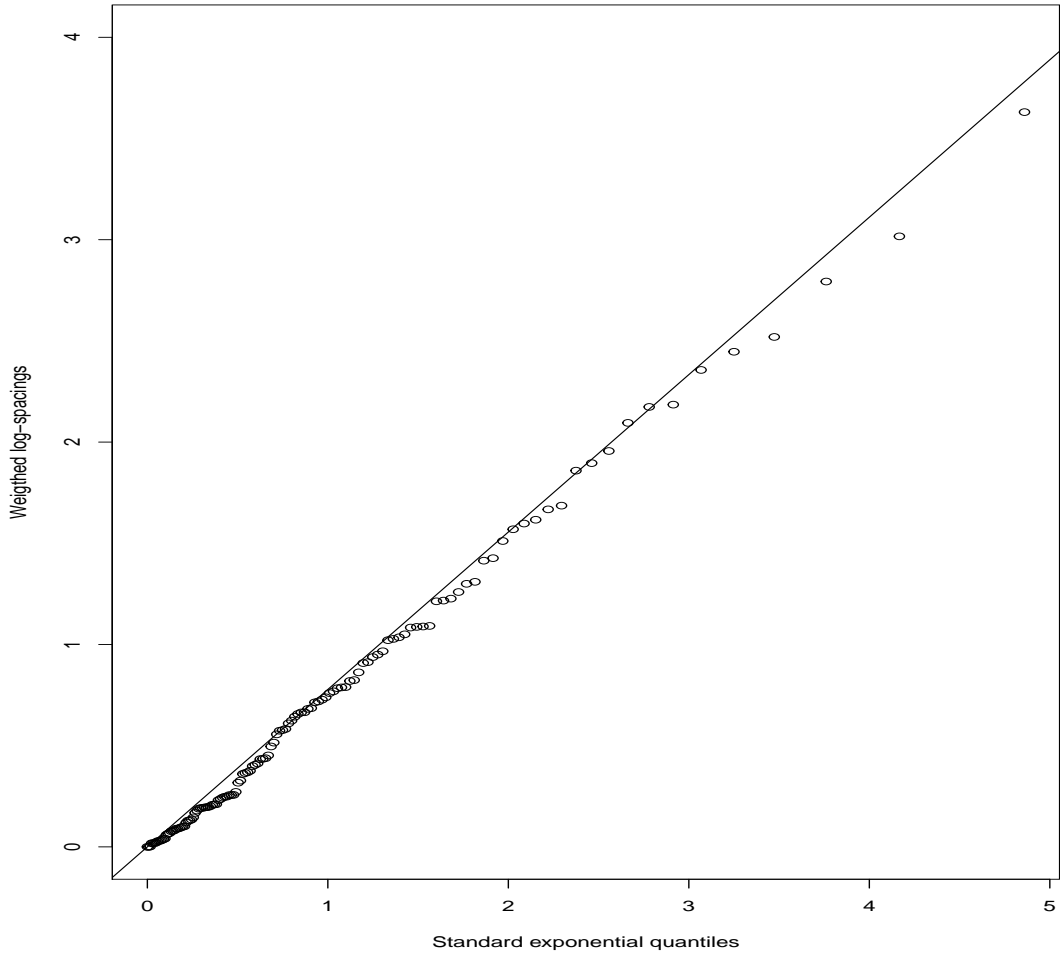
Figure 5: Illustration on motorcycle insurance data, quantile-quantile plot. Horizontally: Standard exponential quantiles, vertically: Weighted log-spacings computed on the residuals. The continuous line has slope $\hat{\gamma}_n$.
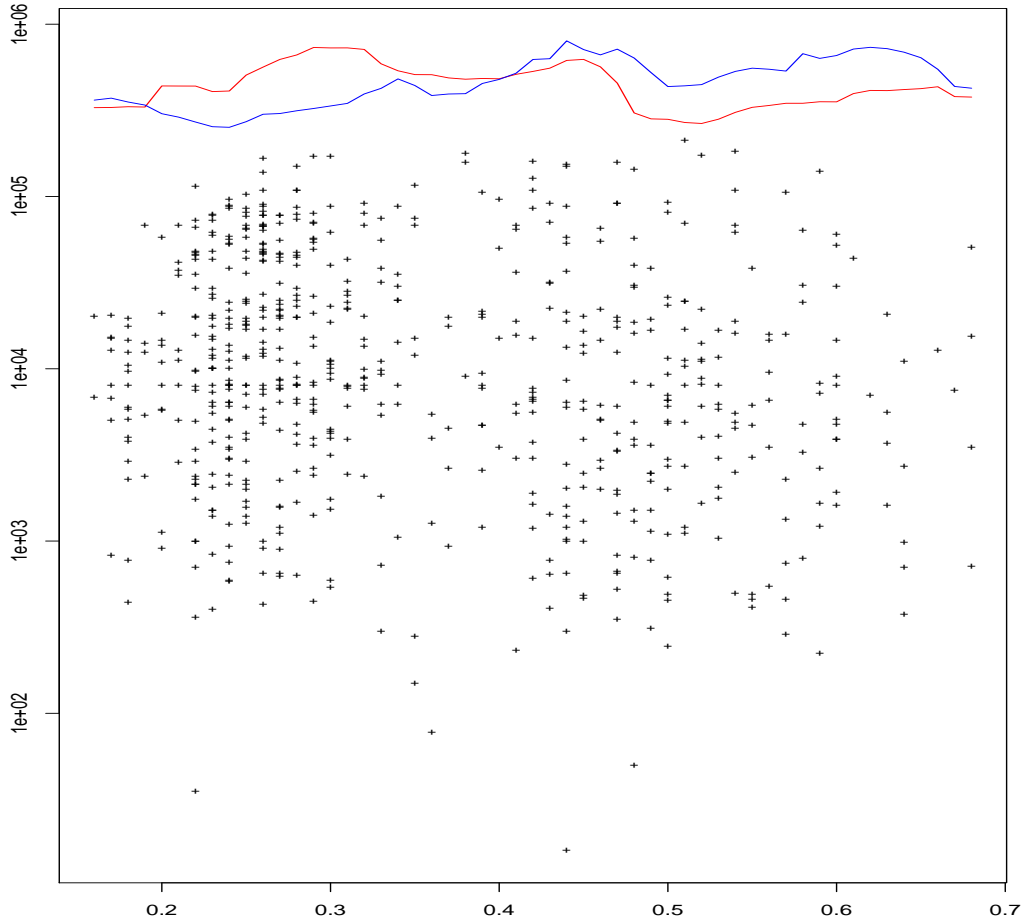
Figure 6: Illustration on motorcycle insurance data. Horizontally: Age of the policyholder (hundred of years), vertically: Claim severity (SEK, log scale). Data (+), nonparametric conditional Weissman estimator $\check{q}_{n,Y}(\alpha_n \mid \cdot)$ (blue) and semi-parametric extreme quantile estimator $\tilde{q}_{n,Y}(\alpha_n \mid \cdot)$ (red).