



**HAL**  
open science

## On kernel smoothing for extremal quantile regression

Abdelaati Daouia, Laurent Gardes, Stéphane Girard

► **To cite this version:**

Abdelaati Daouia, Laurent Gardes, Stéphane Girard. On kernel smoothing for extremal quantile regression. *Bernoulli*, 2013, 19 (5B), pp.2557-2589. 10.3150/12-BEJ466 . hal-00630726v3

**HAL Id: hal-00630726**

**<https://hal.science/hal-00630726v3>**

Submitted on 9 Jul 2012

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

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

# On kernel smoothing for extremal quantile regression

Abdelaati Daouia<sup>(1)</sup>, Laurent Gardes<sup>(2)</sup> & Stéphane Girard<sup>(3,\*)</sup>

<sup>(1)</sup> Institute of Statistics, Catholic University of Louvain, Belgium.

Toulouse School of Economics, University of Toulouse, France.

<sup>(2)</sup> Université de Strasbourg & CNRS, IRMA, UMR 7501, France.

<sup>(3)</sup> INRIA Rhône-Alpes & LJK, team Mistis, Inovallée, 655, av. de l'Europe, Montbonnot, 38334 Saint-Ismier cedex, France.

(\*) `Stephane.Girard@inria.fr` (corresponding author)

## Abstract

Nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution of the response are long established in statistics. Attention has been, however, restricted to ordinary quantiles staying away from the tails of the conditional distribution. The purpose of this paper is to extend their asymptotic theory far enough into the tails. We focus on extremal quantile regression estimators of a response variable given a vector of covariates in the general setting, whether the conditional extreme-value index is positive, negative, or zero. Specifically, we elucidate their limit distributions when they are located in the range of the data or near and even beyond the sample boundary, under technical conditions that link the speed of convergence of their (intermediate or extreme) order with the oscillations of the quantile function and a von-Mises property of the conditional distribution. A simulation experiment and an illustration on real data were presented. The real data are the American electric data where the estimation of conditional extremes is found to be of genuine interest.

*Key words* : Regression, extreme quantile, extreme-value index, kernel smoothing, von-Mises condition, asymptotic normality

## 1 Introduction

Quantile regression plays a fundamental role in various statistical applications. It complements the classical regression on the conditional mean by offering a more useful tool for examining how a vector of regressors  $X \in \mathbb{R}^p$  influences the entire distribution of a response variable  $Y \in \mathbb{R}$ . The nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution function are used widely in applied work and investigated extensively in theoretical statistics. See, for example [3, 28, 30, 31], among others. Attention has been, however, restricted to conditional quantiles having a fixed order  $\alpha \in (0, 1)$ . In the following, the order  $\alpha$  has to be understood as the conditional probability to be larger than the conditional quantile. In result, the available large sample theory does not apply sufficiently far in the tails.

There are many important applications in ecology, climatology, demography, biostatistics, econometrics, finance, insurance, to name a few, where extending that conventional asymptotic theory further into the tails of the conditional distribution is an especially welcome development. This translates into considering the order  $\alpha = \alpha_n \rightarrow 0$  or  $\alpha_n \rightarrow 1$  as the sample size  $n$  goes to infinity. Motivating examples include the study of extreme rainfall as a function of the geographical location [14], the estimation of factors of high risk in finance [32], the assessment of the optimal cost of the delivery activity of postal services [6], the analysis of survival at extreme durations [24], the edge estimation in image reconstruction [25], the accurate description of the upper tail of the

claim size distribution for reinsurers [2], the analysis of environmental time series with application to trend detection in ground-level ozone [29], the estimation of autoregressive models with asymmetric innovations [11], etc.

There have been several efforts to treat the asymptotics of extreme conditional quantile estimators in semi/parametric and other nonparametric regression models. For example, Chernozhukov [5] and Jurecková [23] considered the extreme quantiles in the linear regression model and derived their asymptotic distributions under various distributions of errors. Other parametric models are proposed in [9, 29], where some extreme-value based techniques are extended to the point-process view of high-level exceedances. A semi-parametric approach to modeling trends in sample extremes, based on local polynomial fitting of the Generalized extreme-value distribution, has been introduced in [8]. Hall and Tajvidi [20] suggested a nonparametric estimation of the temporal trend when fitting parametric models to extreme values. Another semi-parametric method has been developed in [1], where the regression is based on a Pareto-type conditional distribution of the response. Fully nonparametric estimators of extreme conditional quantiles have been discussed in [1, 4], where the former approach is based on the technique of local polynomial maximum likelihood estimation, while spline estimators are fitted in the latter by a maximum penalized likelihood method. Recently, [13, 15] proposed, respectively, a moving-window based estimator for the tail index and extreme quantiles of heavy-tailed conditional distributions, and they established their asymptotic properties.

In the context of kernel-smoothing, the asymptotic theory for quantile regression in the tails is relatively unexplored and still in full development. Daouia *et al.* [7] have extended the asymptotics further into the tails in the particular setting of a heavy-tailed conditional distribution, while [16, 17] have analyzed the case  $\alpha_n = 1/n$  in the particular situation where the response  $Y$  given  $X = x$  is uniformly distributed. The purpose of this paper is to develop a unified asymptotic theory for the kernel-smoothed conditional extremes in the general setting where the conditional distribution can be short, light or heavy-tailed. We will focus on the  $\alpha_n \rightarrow 0$  case, which corresponds to the class of large quantiles of the upper conditional tail. Similar considerations evidently apply to the case  $\alpha_n \rightarrow 1$ . Specifically, we first obtain the asymptotic normality of the extremal quantile regression under the ‘intermediate’ order condition  $nh^p\alpha_n \rightarrow \infty$  where  $h = h_n \rightarrow 0$  stands for the bandwidth involved in the kernel smoothing estimation. Next we extend the asymptotic normality far enough into the ‘most extreme’ order- $\beta_n$  regression quantiles with  $\beta_n/\alpha_n \rightarrow 0$ , thus providing a conditional analog of modern extreme-value results [19]. We also analyze kernel-smoothed Pickands type estimators of the conditional extreme-value index as in the familiar nonregression case [10].

The paper is organized as follows. Section 2 contains the basic notations and assumptions. Section 3 states the main results. Section 4 presents some simulation evidence and practical guidelines. Section 5 provides a motivating example in production theory, and Section 6 collects the proofs.

## 2 The setting and assumptions

Let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , be independent copies of a random pair  $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$ . The conditional survival function (csf) of  $Y$  given  $X = x$  is denoted by  $\bar{F}(y|x) = \mathbb{P}(Y > y|X = x)$  and the probability density function (pdf) of  $X$  is denoted by  $g$ . We address the problem of estimating extreme conditional quantiles

$$q(\alpha_n|x) = \bar{F}^{\leftarrow}(\alpha_n|x) = \inf\{t, \bar{F}(t|x) \leq \alpha_n\},$$

where  $\alpha_n \rightarrow 0$  as  $n$  goes to infinity. In the following we denote by  $y_F(x) = q(0|x) \in (-\infty, \infty]$  the endpoint of the conditional distribution of  $Y$  given  $X = x$ . The kernel estimator of  $\bar{F}(y|x)$  is

defined for all  $(x, y) \in \mathbb{R}^p \times \mathbb{R}$  by

$$\hat{F}_n(y|x) = \sum_{i=1}^n K_h(x - X_i) \mathbb{I}\{Y_i > y\} \Big/ \sum_{i=1}^n K_h(x - X_i), \quad (1)$$

where  $\mathbb{I}\{\cdot\}$  is the indicator function and  $h = h_n$  is a nonrandom sequence such that  $h \rightarrow 0$  as  $n \rightarrow \infty$ . We have also introduced  $K_h(t) = K(t/h)/h^p$  where  $K$  is a pdf on  $\mathbb{R}^p$ . In this context,  $h$  is called the window-width. Similarly, the kernel estimators of conditional quantiles  $q(\alpha|x)$  are defined via the generalized inverse of  $\hat{F}_n(\cdot|x)$ :

$$\hat{q}_n(\alpha|x) = \hat{F}_n^{\leftarrow}(\alpha|x) = \inf\{t, \hat{F}_n(t|x) \leq \alpha\}, \quad (2)$$

for all  $\alpha \in (0, 1)$ . Many papers are dedicated to the asymptotic properties of this type of estimator for fixed  $\alpha \in (0, 1)$ : weak and strong consistency are proved respectively in [30] and [12], asymptotic normality being established in [31, 28, 3]. In Theorem 1 below, the asymptotic distribution of (2) is investigated when estimating extreme quantiles, *i.e.* when  $\alpha = \alpha_n$  goes to 0 as the sample size  $n$  goes to infinity. The asymptotic behavior of such estimators then depends on the nature of the conditional distribution tail. In this paper, we assume that the csf satisfies the following von-Mises condition, see for instance [19, equation (1.11.30)]:

**(A.1)** The function  $\bar{F}(\cdot|x)$  is twice differentiable and

$$\lim_{y \uparrow y_F(x)} \frac{\bar{F}(y|x) \bar{F}''(y|x)}{(\bar{F}')^2(y|x)} = \gamma(x) + 1,$$

where  $\bar{F}'(\cdot|x)$  and  $\bar{F}''(\cdot|x)$  are respectively the first and the second derivatives of  $\bar{F}(\cdot|x)$ .

Here,  $\gamma(\cdot)$  is an unknown function of the covariate  $x$  referred to as the conditional extreme-value index. Let us consider, for all  $z \in \mathbb{R}$ , the classical  $K_z$  function defined for all  $u \in \mathbb{R}$  by

$$K_z(u) = \int_1^u v^{z-1} dv.$$

The associated inverse function is denoted by  $K_z^{-1}$ . Then, **(A.1)** implies that there exists a positive auxiliary function  $a(\cdot|x)$  such that,

$$\lim_{y \uparrow y_F(x)} \frac{\bar{F}(y + t(x)a(y|x)|x)}{\bar{F}(y|x)} = \frac{1}{K_{\gamma(x)}^{-1}(t(x))}, \quad (3)$$

where  $t(x) \in \mathbb{R}$  is such that  $1 + t(x)\gamma(x) > 0$ . Besides, (3) implies in turn that the conditional distribution of  $Y$  given  $X = x$  is in the maximum domain of attraction (MDA) of the extreme-value distribution with shape parameter  $\gamma(x)$ , see [19, Theorem 1.1.8] for a proof. The case  $\gamma(x) > 0$  corresponds to the Fréchet MDA and  $\bar{F}(\cdot|x)$  is heavy-tailed while the case  $\gamma(x) = 0$  corresponds to the Gumbel MDA and  $\bar{F}(\cdot|x)$  is light-tailed. The case  $\gamma(x) < 0$  represents most of the situations where  $\bar{F}(\cdot|x)$  is short-tailed, *i.e.*  $\bar{F}(\cdot|x)$  has a finite endpoint  $y_F(x)$ , this is referred to as the Weibull MDA.

The convergence (3) is also equivalent to

$$b(t, \alpha|x) := \frac{q(t\alpha|x) - q(\alpha|x)}{a(q(\alpha|x)|x)} - K_{\gamma(x)}(1/t) \rightarrow 0 \quad (4)$$

for all  $t > 0$  as  $\alpha \rightarrow 0$ , see [19, Theorem 1.1.6]. For all  $(x, x') \in \mathbb{R}^p \times \mathbb{R}^p$ , the Euclidean distance between  $x$  and  $x'$  is denoted by  $d(x, x')$ . The following Lipschitz condition is introduced:

**(A.2)** There exists  $c_g > 0$  such that  $|g(x) - g(x')| \leq c_g d(x, x')$ .

The last assumption is standard in the kernel estimation framework.

**(A.3)**  $K$  is a bounded pdf on  $\mathbb{R}^p$ , with support  $S$  included in the unit ball of  $\mathbb{R}^p$ .

### 3 Main results

Let  $B(x, h)$  be the ball centered at  $x$  with radius  $h$ . The oscillations of the csf are controlled by

$$\Delta_\kappa(x, \alpha) := \sup_{(x', \beta) \in B(x, h) \times [\kappa\alpha, \alpha]} \left| \frac{\bar{F}(q(\beta|x)|x')}{\beta} - 1 \right|,$$

where  $(\kappa, \alpha) \in (0, 1)^2$ . Under assumption **(A.1)**,  $\bar{F}(\cdot|x)$  is differentiable and the associated conditional density will be denoted in the sequel by  $f(\cdot|x)$ . We first establish the asymptotic normality of  $\hat{q}_n(\alpha_n|x)$ .

**Theorem 1.** *Suppose **(A.1)**, **(A.2)** and **(A.3)** hold. Let  $0 < \tau_J < \dots < \tau_2 < \tau_1 \leq 1$  where  $J$  is a positive integer and  $x \in \mathbb{R}^p$  such that  $g(x) > 0$ . If  $\alpha_n \rightarrow 0$  and there exists  $\kappa \in (0, \tau_J)$  such that*

$$nh^p \alpha_n \rightarrow \infty, \quad nh^p \alpha_n (h \vee \Delta_\kappa(x, \alpha_n))^2 \rightarrow 0,$$

then, the random vector

$$\left\{ f(q(\alpha_n|x)|x) \sqrt{nh^p \alpha_n^{-1}} (\hat{q}_n(\tau_j \alpha_n|x) - q(\tau_j \alpha_n|x)) \right\}_{j=1, \dots, J}$$

is asymptotically Gaussian, centered, with covariance matrix  $\|K\|_2^2/g(x)\Sigma(x)$  where  $\Sigma_{j,j'}(x) = (\tau_j \tau_{j'})^{-\gamma(x)} \tau_{j \wedge j'}^{-1}$  for  $(j, j') \in \{1, \dots, J\}^2$ .

Let us remark that, in the particular case where  $J = 1$ ,  $\tau_1 = 1$  and  $\alpha_n = \alpha$  is fixed in  $(0, 1)$ , we find back the result of [3, Theorem 6.4]. Theorem 1 can be equivalently rewritten as

**Corollary 1.** *Under the assumptions of Theorem 1, the random vector*

$$\left\{ \sqrt{nh^p \alpha_n} \frac{q(\alpha_n|x)}{a(q(\alpha_n|x)|x)} \left( \frac{\hat{q}_n(\tau_j \alpha_n|x)}{q(\tau_j \alpha_n|x)} - 1 \right) \right\}_{j=1, \dots, J}$$

is asymptotically Gaussian, centered, with covariance matrix  $\|K\|_2^2/g(x)\tilde{\Sigma}(x)$  where  $\tilde{\Sigma}_{j,j'}(x) = (\tau_j \tau_{j'})^{-\gamma(x) \wedge 0} \tau_{j \wedge j'}^{-1}$  for  $(j, j') \in \{1, \dots, J\}^2$ .

Moreover [19, Theorem 1.2.5] and [19, page 33] show that

$$\lim_{y \uparrow y_F(x)} \frac{a(y|x)}{y} = \gamma(x) \vee 0. \quad (5)$$

Under the assumptions of Theorem 1, and from (5), it follows that  $\hat{q}_n(\tau_j \alpha_n|x)/q(\tau_j \alpha_n|x) \xrightarrow{P} 1$  when  $n \rightarrow \infty$  which can be read as a weak consistency result for the considered estimator. Besides, if  $\gamma(x) > 0$ , then collecting (5) and Corollary 1 shows that the random vector

$$\left\{ \sqrt{nh^p \alpha_n} \left( \frac{\hat{q}_n(\tau_j \alpha_n|x)}{q(\tau_j \alpha_n|x)} - 1 \right) \right\}_{j=1, \dots, J}$$

is asymptotically Gaussian, centered, with covariance matrix  $\|K\|_2^2 \gamma^2(x)/g(x)\tilde{\Sigma}(x)$  where the coefficients of the covariance matrix can be simplified  $\tilde{\Sigma}_{j,j'}(x) = \tau_{j \wedge j'}^{-1}$  for  $(j, j') \in \{1, \dots, J\}^2$ . Our results thus build on and complement the analysis given by [7, Theorem 2] in the case  $\gamma(x) > 0$ .

As pointed out in [7], the condition  $nh^p \alpha_n \rightarrow \infty$  implies  $\alpha_n > \log^p(n)/n$  eventually. This condition provides a lower bound on the order of the extreme conditional quantiles for the asymptotic normality of kernel estimators to hold. We now propose a scheme to estimate extreme conditional quantiles without this restriction. Let  $\alpha_n \rightarrow 0$  and  $\beta_n/\alpha_n \rightarrow 0$  as  $n \rightarrow \infty$ . Suppose one has  $\hat{\gamma}_n(x)$

and  $\hat{a}_n(x)$  two estimators of  $\gamma(x)$  and  $a(q(\alpha_n|x)|x)$  respectively. Then, starting from the estimator  $\hat{q}_n(\alpha_n|x)$  of  $q(\alpha_n|x)$  defined in (2) and making use of (4), it is possible to build an estimator  $\tilde{q}_n(\beta_n|x)$  of  $q(\beta_n|x)$  which is an extreme conditional quantile of higher order than  $q(\alpha_n|x)$ :

$$\tilde{q}_n(\beta_n|x) = \hat{q}_n(\alpha_n|x) + K_{\hat{\gamma}_n(x)}(\alpha_n/\beta_n)\hat{a}_n(x). \quad (6)$$

Let us consider, for all  $z \in \mathbb{R}$ , the function defined for all  $u > 1$  by

$$K'_z(u) = \frac{\partial K_z(u)}{\partial z} = \int_1^u v^{z-1} \log(v) dv.$$

The following result provides a quantile regression analog of [19, Theorem 4.3.1].

**Theorem 2.** *Suppose (A.1) holds and let  $\alpha_n \rightarrow 0$ ,  $\beta_n/\alpha_n \rightarrow 0$ . Let  $\hat{q}_n(\alpha_n|x)$  be the kernel estimator of  $q(\alpha_n|x)$  defined in (2). Let  $\hat{\gamma}_n(x)$  and  $\hat{a}_n(x)$  be two estimators of  $\gamma(x)$  and  $a(q(\alpha_n|x)|x)$  respectively such that*

$$\Lambda_n^{-1} \left( \hat{\gamma}_n(x) - \gamma(x), \frac{\hat{a}_n(x)}{a(q(\alpha_n|x)|x)} - 1, \frac{\hat{q}_n(\alpha_n|x) - q(\alpha_n|x)}{a(q(\alpha_n|x)|x)} \right)^t \xrightarrow{d} \zeta(x), \quad (7)$$

where  $\zeta(x)$  is a non-degenerate  $\mathbb{R}^3$  random vector,

$$\Lambda_n \log(\alpha_n/\beta_n) \rightarrow 0 \text{ and } \Lambda_n^{-1} \frac{b(\beta_n/\alpha_n, \alpha_n|x)}{K'_{\gamma(x)}(\alpha_n/\beta_n)} \rightarrow 0$$

as  $n \rightarrow \infty$ . Then,

$$\Lambda_n^{-1} \left( \frac{\tilde{q}_n(\beta_n|x) - q(\beta_n|x)}{a(q(\alpha_n|x)|x)K'_{\gamma(x)}(\alpha_n/\beta_n)} \right) \xrightarrow{d} c(x)^t \zeta(x),$$

where  $c(x)^t = (1, -(\gamma(x) \wedge 0), (\gamma(x) \wedge 0)^2)$ .

As an illustration, for all  $r \in (0, 1)$ , let us consider  $\tau_j = r^{j-1}$ ,  $j = 1, \dots, J$ . The following estimators of  $\gamma(x)$  and  $a(q(\alpha_n|x)|x)$  are introduced

$$\begin{aligned} \hat{\gamma}_n^{\text{RP}}(x) &= \frac{1}{\log r} \sum_{j=1}^{J-2} \pi_j \log \left( \frac{\hat{q}_n(\tau_j \alpha_n|x) - \hat{q}_n(\tau_{j+1} \alpha_n|x)}{\hat{q}_n(\tau_{j+1} \alpha_n|x) - \hat{q}_n(\tau_{j+2} \alpha_n|x)} \right) \\ \hat{a}_n^{\text{RP}}(x) &= \frac{1}{K_{\hat{\gamma}_n^{\text{RP}}(x)}(r)} \sum_{j=1}^{J-2} \pi_j r^{\hat{\gamma}_n^{\text{RP}}(x)j} (\hat{q}_n(\tau_j \alpha_n|x) - \hat{q}_n(\tau_{j+1} \alpha_n|x)), \end{aligned}$$

where  $(\pi_j)$  is a sequence of weights summing to one. Let us highlight that  $\hat{\gamma}_n^{\text{RP}}(x)$  is an adaptation to the conditional case of the Refined Pickands estimator introduced in [10]. The joint asymptotic normality of  $(\hat{\gamma}_n^{\text{RP}}(x), \hat{a}_n^{\text{RP}}(x), \hat{q}_n(\alpha_n|x))$  is established in the next theorem.

**Theorem 3.** *Suppose (A.1), (A.2) and (A.3) hold. Let  $x \in \mathbb{R}^p$  such that  $g(x) > 0$ . If  $\alpha_n \rightarrow 0$  and there exists  $\kappa \in (0, \tau_J)$  such that*

$$nh^p \alpha_n \rightarrow \infty, \quad nh^p \alpha_n \left( h \vee \Delta_\kappa(x, \alpha_n) \vee \bigvee_{j=1}^J b(\tau_j, \alpha_n|x) \right)^2 \rightarrow 0,$$

as  $n \rightarrow \infty$ , then the random vector

$$\sqrt{nh^p \alpha_n} \left( \hat{\gamma}_n^{\text{RP}}(x) - \gamma(x), \frac{\hat{a}_n^{\text{RP}}(x)}{a(q(\alpha_n|x)|x)} - 1, \frac{\hat{q}_n(\alpha_n|x) - q(\alpha_n|x)}{a(q(\alpha_n|x)|x)} \right)^t$$

is asymptotically centered and Gaussian.

The asymptotic covariance matrix is denoted by  $S(x)$ . It can be explicitly calculated from (27) in the proof of Theorem 3, but the result would be too complicated to be reported here. As a consequence of the two above theorems, one obtains the asymptotic normality of the extreme conditional quantile estimator built on  $\hat{\gamma}_n^{\text{RP}}(x)$  and  $\hat{a}_n^{\text{RP}}(x)$ :

$$\tilde{q}_n^{\text{RP}}(\beta_n|x) := \hat{q}_n(\alpha_n|x) + K_{\hat{\gamma}_n^{\text{RP}}(x)}(\alpha_n/\beta_n)\hat{a}_n^{\text{RP}}(x).$$

**Corollary 2.** *Suppose (A.1), (A.2) and (A.3) hold. Let  $x \in \mathbb{R}^p$  such that  $g(x) > 0$ . If  $\alpha_n \rightarrow 0$ ,  $\beta_n/\alpha_n \rightarrow 0$  and there exists  $\kappa \in (0, \tau_J)$  such that*

$$\frac{nh^p\alpha_n}{(\log(\alpha_n/\beta_n))^2} \rightarrow \infty, \quad nh^p\alpha_n \left( h \vee \Delta_\kappa(x, \alpha_n) \vee \bigvee_{j=1}^J b(\tau_j, \alpha_n|x) \vee \frac{b(\beta_n/\alpha_n, \alpha_n|x)}{K'_{\gamma(x)}(\alpha_n/\beta_n)} \right)^2 \rightarrow 0,$$

as  $n \rightarrow \infty$ , then

$$\sqrt{nh^p\alpha_n} \left( \frac{\tilde{q}_n^{\text{RP}}(\beta_n|x) - q(\beta_n|x)}{a(q(\alpha_n|x)|x)K'_{\gamma(x)}(\alpha_n/\beta_n)} \right)$$

is asymptotically Gaussian, centered with variance  $c(x)^t S(x) c(x)$ .

Finally, two particular cases of  $\hat{\gamma}_n^{\text{RP}}(x)$  may be considered. First, constant weights  $\pi_1 = \dots = \pi_{J-2} = 1/(J-2)$  yield

$$\hat{\gamma}_n^{\text{RP},1}(x) = \frac{1}{(J-2)\log r} \log \left( \frac{\hat{q}_n(\tau_1\alpha_n|x) - \hat{q}_n(\tau_2\alpha_n|x)}{\hat{q}_n(\tau_{J-1}\alpha_n|x) - \hat{q}_n(\tau_J\alpha_n|x)} \right).$$

Clearly, when  $J = 3$ , this estimator reduces the kernel Pickands estimator introduced and studied in [7] in the situation where  $\gamma(x) > 0$ . Second, linear weights  $\pi_j = 2j/((J-1)(J-2))$  for  $j = 1, \dots, J-2$  give rise to a new estimator

$$\hat{\gamma}_n^{\text{RP},2}(x) = \frac{2}{(J-1)(J-2)\log r} \sum_{j=1}^{J-2} \log \left( \frac{\hat{q}_n(\tau_j\alpha_n|x) - \hat{q}_n(\tau_{j+1}\alpha_n|x)}{\hat{q}_n(\tau_{J-1}\alpha_n|x) - \hat{q}_n(\tau_J\alpha_n|x)} \right),$$

which can be read as the average of  $J-1$  estimators  $\hat{\gamma}_n^{\text{RP},1}(x)$ . These estimators are now compared on finite sample situations.

## 4 Some simulation evidence

Subsection 4.1 provides Monte Carlo evidence that the extreme quantile function estimator  $\tilde{q}_n^{\text{RP},1}(\beta_n|x)$  is efficient relative to the version  $\tilde{q}_n^{\text{RP},2}(\beta_n|x)$ , whether  $\gamma(x)$  is positive, negative or zero, and outperforms the estimator  $\hat{q}_n(\beta_n|x)$  for heavy-tailed conditional distributions. Subsection 4.2 provides a comparison with the promising local smoothing approach introduced in [1] and [2, Section 7.5.2]. Practical guidelines for selecting the bandwidth  $h$  and the order  $\alpha_n$  are suggested in Subsection 4.3.

### 4.1 Monte Carlo experiments

To evaluate finite-sample performance of the conditional extreme-value index and extreme quantile estimators described above we have undertaken some simulation experiments following the model

$$Y_i = \mathcal{G}(X_i) + \sigma(X_i)U_i, \quad i = 1, \dots, n.$$

The local scale factor,  $\sigma(x) = (1+x)/10$ , is linearly increasing in  $x$ , while the local location parameter

$$\mathcal{G}(x) = \sqrt{x(1-x)} \sin \left( \frac{2\pi(1+2^{-7/5})}{x+2^{-7/5}} \right)$$

has been introduced in [27, Section 17.5.1]. The design points  $X_i$  are generated following a standard uniform distribution. The  $U_i$ 's are independent and their conditional distribution given  $X_i = x$  is chosen to be standard Gaussian, Student  $t_{k(x)}$ , or Beta( $\nu(x), \nu(x)$ ), with

$$k(x) = [\nu(x)] + 1, \quad \nu(x) = \left\{ \left( \frac{1}{10} + \sin(\pi x) \right) \left( \frac{11}{10} - \frac{1}{2} \exp\{-64(x - 1/2)^2\} \right) \right\}^{-1},$$

and  $[\nu(x)]$  being the integer part of  $\nu(x)$ . Let us recall that the Gaussian distribution belongs to the Gumbel MDA, *i.e.*  $\gamma(x) = 0$ , the Student distribution  $t_{k(x)}$  belongs to the Fréchet MDA with  $\gamma(x) = 1/k(x) > 0$  and the Beta distribution belongs to the Weibull MDA with  $\gamma(x) = -1/\nu(x) < 0$ .

In all cases we have  $q(\beta|x) = \mathcal{G}(x) + \sigma(x)\bar{F}_{U|X}^{\leftarrow}(\beta|x)$ , for  $\beta \in (0, 1)$ . All the experiments were performed over 400 simulations for  $n = 200$ , and the kernel function  $K$  was chosen to be the Triweight kernel

$$K(t) = \frac{35}{32}(1 - t^2)^3 \mathbb{I}\{-1 \leq t \leq 1\}.$$

Monte Carlo experiments were first devoted to accuracy of the two conditional extreme-value index estimators  $\hat{\gamma}_n^{\text{RP},1}(x)$  and  $\hat{\gamma}_n^{\text{RP},2}(x)$ . The measures of efficiency for each simulation used were the mean squared error and the bias

$$\text{MSE}\{\hat{\gamma}_n(\cdot)\} = \frac{1}{L} \sum_{\ell=1}^L \{\hat{\gamma}_n(x_\ell) - \gamma(x_\ell)\}^2, \quad \text{Bias}\{\hat{\gamma}_n(\cdot)\} = \frac{1}{L} \sum_{\ell=1}^L \{\hat{\gamma}_n(x_\ell) - \gamma(x_\ell)\}$$

for  $\hat{\gamma}_n(x) = \hat{\gamma}_n^{\text{RP},1}(x)$ ,  $\hat{\gamma}_n^{\text{RP},2}(x)$ , with the  $x_\ell$ 's being  $L = 100$  points regularly distributed in  $[0, 1]$ . To guarantee a fair comparison among the two estimation methods, we used for each estimator the parameters  $(\alpha_n, h)$  minimizing its mean squared error, with  $\alpha_n$  ranging over  $\mathcal{A} = \{0.1, 0.15, 0.2, \dots, 0.95\}$  and the bandwidth  $h$  ranging over a grid  $\mathcal{H}$  of 50 points regularly distributed between  $h_{\min} = \max_{1 \leq i < n} |X_{(i+1)} - X_{(i)}|$  and  $h_{\max} = |X_{(n)} - X_{(1)}|/2$ , where  $X_{(1)} \leq \dots \leq X_{(n)}$  are the ordered observations. The resulting values of MSE and bias are averaged on the 400 Monte Carlo replications and reported in Table 1 for  $J \in \{3, 4, 5\}$  and  $r \in \{1/J, (J-1)/J\}$ .

|                 | $r = 1/J$                         |                                   |                                   |                                   | $r = (J-1)/J$                     |                                   |                                   |                                   |
|-----------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
|                 | MSE                               |                                   | Bias                              |                                   | MSE                               |                                   | Bias                              |                                   |
|                 | $\hat{\gamma}_n^{\text{RP},1}(x)$ | $\hat{\gamma}_n^{\text{RP},2}(x)$ | $\hat{\gamma}_n^{\text{RP},1}(x)$ | $\hat{\gamma}_n^{\text{RP},2}(x)$ | $\hat{\gamma}_n^{\text{RP},1}(x)$ | $\hat{\gamma}_n^{\text{RP},2}(x)$ | $\hat{\gamma}_n^{\text{RP},1}(x)$ | $\hat{\gamma}_n^{\text{RP},2}(x)$ |
| <b>Gaussian</b> |                                   |                                   |                                   |                                   |                                   |                                   |                                   |                                   |
| J=3             | 0.2026                            | 0.2026                            | -0.2415                           | -0.2415                           | 0.7656                            | 0.7656                            | -0.3213                           | -0.3213                           |
| J=4             | 0.1915                            | 0.2018                            | -0.3270                           | -0.3501                           | 0.6730                            | 0.7960                            | -0.3455                           | -0.3747                           |
| J=5             | NaN                               | NaN                               | NaN                               | NaN                               | 0.7305                            | 0.9128                            | -0.4104                           | -0.4107                           |
| <b>Student</b>  |                                   |                                   |                                   |                                   |                                   |                                   |                                   |                                   |
| J=3             | 0.2882                            | 0.2882                            | -0.2964                           | -0.2964                           | 1.1109                            | 1.1109                            | -0.4497                           | -0.4497                           |
| J=4             | 0.3350                            | 0.2837                            | -0.4167                           | -0.3480                           | 0.9991                            | 1.1997                            | -0.4384                           | -0.4597                           |
| J=5             | NaN                               | NaN                               | NaN                               | NaN                               | 1.1245                            | 1.3331                            | -0.5715                           | -0.5872                           |
| <b>Beta</b>     |                                   |                                   |                                   |                                   |                                   |                                   |                                   |                                   |
| J=3             | 0.1157                            | 0.1157                            | -0.0730                           | -0.0730                           | 0.6737                            | 0.6737                            | -0.2591                           | -0.2591                           |
| J=4             | 0.0510                            | 0.0597                            | -0.0811                           | -0.0750                           | 0.5861                            | 0.6891                            | -0.2338                           | -0.2432                           |
| J=5             | NaN                               | NaN                               | NaN                               | NaN                               | 0.6431                            | 0.8167                            | -0.2185                           | -0.2757                           |

Table 1: Performance of  $\hat{\gamma}_n^{\text{RP},1}(x)$  and  $\hat{\gamma}_n^{\text{RP},2}(x)$  – Results averaged on 400 simulations with  $n = 200$ . The results may not be available for  $r = 1/J$  and  $J = 5$  since the numerator  $\{\hat{q}_n(\tau_j \alpha_n | x) - \hat{q}_n(\tau_{j+1} \alpha_n | x)\}$  and the denominator  $\{\hat{q}_n(\tau_{J-1} \alpha_n | x) - \hat{q}_n(\tau_J \alpha_n | x)\}$  in the definitions of both estimators might be null when  $n$  is not large enough.

It does appear that the results for  $r = 1/J$  are superior to those for  $r = (J-1)/J$ , uniformly in  $J$ . For these desirable results, it may be seen that the estimator  $\hat{\gamma}_n^{\text{RP},1}(x)$  performs better than  $\hat{\gamma}_n^{\text{RP},2}(x)$  in the Gaussian error model, whereas the latter is superior to the former in the Student



error model. It may be also seen that there is no winner in the Beta error model in terms of both MSE and bias.

Turning to the performance of the extreme conditional quantile estimators, we consider as above the two measures of performance

$$\text{MSE}\{q_n(\beta_n|\cdot)\} = \frac{1}{L} \sum_{\ell=1}^L \{q_n(\beta_n|x_\ell) - q(\beta_n|x_\ell)\}^2, \quad \text{Bias}\{q_n(\beta_n|\cdot)\} = \frac{1}{L} \sum_{\ell=1}^L \{q_n(\beta_n|x_\ell) - q(\beta_n|x_\ell)\},$$

for  $q_n(\beta_n|x) = \hat{q}_n(\beta_n|x)$ ,  $\tilde{q}_n^{\text{RP},1}(\beta_n|x)$ ,  $\tilde{q}_n^{\text{RP},2}(\beta_n|x)$ . The averaged MSE and bias of these three estimators of  $q(\beta_n|x)$ , computed for  $\beta_n \in \{0.05, 0.01, 0.005\}$ ,  $J \in \{3, 4\}$  and  $r = 1/J$ , over 400 Monte Carlo simulations are displayed in Table 2. Here also, we used for each estimator the smoothing parameters  $(\alpha_n, h)$  minimizing its MSE over the grid of values  $\mathcal{A} \times \mathcal{H}$  described above. When comparing the estimators  $\tilde{q}_n^{\text{RP},1}(\beta_n|x)$  and  $\tilde{q}_n^{\text{RP},2}(\beta_n|x)$  themselves with  $\hat{q}_n(\beta_n|x)$ , the results (both in terms of MSE and bias) indicate that  $\tilde{q}_n^{\text{RP},2}(\beta_n|x)$  is slightly less efficient than  $\tilde{q}_n^{\text{RP},1}(\beta_n|x)$  in all cases, and that the latter is appreciably more efficient than  $\hat{q}_n(\beta_n|x)$  only in the Student error model. It may be also noticed that  $\hat{q}_n(\beta_n|x)$  is more efficient but not by much (especially when  $J = 3$ ) in the Gaussian and Beta error models.

| $r = 1/J, \beta_n = 0.05$ |  |  |                        |  |  |                        |
|---------------------------|--|--|------------------------|--|--|------------------------|
|                           | MSE                                    |  |                        | Bias                                   |  |                        |
|                           | $\tilde{q}_n^{\text{RP},1}(\beta_n x)$ | $\tilde{q}_n^{\text{RP},2}(\beta_n x)$ | $\hat{q}_n(\beta_n x)$ | $\tilde{q}_n^{\text{RP},1}(\beta_n x)$ | $\tilde{q}_n^{\text{RP},2}(\beta_n x)$ | $\hat{q}_n(\beta_n x)$ |
| <b>Gaussian</b>           |  |  |                        |  |  |                        |
| J=3                       | 0.0110                                 | 0.0110                                 | 0.0108                 | 0.0001                                 | 0.0001                                 | 0.0063                 |
| J=4                       | 0.0591                                 | 0.0796                                 | 0.0108                 | 0.1136                                 | 0.1131                                 | 0.0063                 |
| <b>Student</b>            |  |  |                        |  |  |                        |
| J=3                       | 0.0307                                 | 0.0307                                 | 0.0771                 | -0.0134                                | -0.0134                                | 0.0871                 |
| J=4                       | 0.0532                                 | 0.0743                                 | 0.0771                 | 0.0792                                 | 0.0792                                 | 0.0871                 |
| <b>Beta</b>               |  |  |                        |  |  |                        |
| J=3                       | 0.0091                                 | 0.0091                                 | 0.0022                 | 0.0505                                 | 0.0505                                 | 0.0135                 |
| J=4                       | 0.0745                                 | 0.1002                                 | 0.0022                 | 0.1746                                 | 0.1752                                 | 0.0135                 |

| $r = 1/J, \beta_n = 0.01$ |  |  |                        |  |  |                        |
|---------------------------|--|--|------------------------|--|--|------------------------|
|                           | MSE                                    |  |                        | Bias                                   |  |                        |
|                           | $\tilde{q}_n^{\text{RP},1}(\beta_n x)$ | $\tilde{q}_n^{\text{RP},2}(\beta_n x)$ | $\hat{q}_n(\beta_n x)$ | $\tilde{q}_n^{\text{RP},1}(\beta_n x)$ | $\tilde{q}_n^{\text{RP},2}(\beta_n x)$ | $\hat{q}_n(\beta_n x)$ |
| <b>Gaussian</b>           |  |  |                        |  |  |                        |
| J=3                       | 0.0265                                 | 0.0265                                 | 0.0161                 | -0.0776                                | -0.0776                                | -0.0360                |
| J=4                       | 0.0693                                 | 0.0926                                 | 0.0161                 | 0.1092                                 | 0.1225                                 | -0.0360                |
| <b>Student</b>            |  |  |                        |  |  |                        |
| J=3                       | 0.1115                                 | 0.1115                                 | 0.6825                 | -0.0895                                | -0.0895                                | -0.0959                |
| J=4                       | 0.1304                                 | 0.3992                                 | 0.6825                 | 0.0018                                 | 0.1089                                 | -0.0959                |
| <b>Beta</b>               |  |  |                        |  |  |                        |
| J=3                       | 0.0143                                 | 0.0143                                 | 0.0034                 | 0.0523                                 | 0.0523                                 | 0.0212                 |
| J=4                       | 0.1038                                 | 0.1265                                 | 0.0034                 | 0.1964                                 | 0.2064                                 | 0.0212                 |

| $r = 1/J, \beta_n = 0.005$ |  |  |                        |  |  |                        |
|----------------------------|--|--|------------------------|--|--|------------------------|
|                            | MSE                                    |  |                        | Bias                                   |  |                        |
|                            | $\tilde{q}_n^{\text{RP},1}(\beta_n x)$ | $\tilde{q}_n^{\text{RP},2}(\beta_n x)$ | $\hat{q}_n(\beta_n x)$ | $\tilde{q}_n^{\text{RP},1}(\beta_n x)$ | $\tilde{q}_n^{\text{RP},2}(\beta_n x)$ | $\hat{q}_n(\beta_n x)$ |
| <b>Gaussian</b>            |  |  |                        |  |  |                        |
| J=3                        | 0.0354                                 | 0.0354                                 | 0.0203                 | -0.0981                                | -0.0981                                | -0.0524                |
| J=4                        | 0.0719                                 | 0.0932                                 | 0.0203                 | 0.0982                                 | 0.1073                                 | -0.0524                |
| <b>Student</b>             |  |  |                        |  |  |                        |
| J=3                        | 0.2919                                 | 0.2919                                 | 0.9782                 | -0.1623                                | -0.1623                                | -0.2605                |
| J=4                        | 0.4569                                 | 0.9748                                 | 0.9782                 | -0.1920                                | 0.0280                                 | -0.2605                |
| <b>Beta</b>                |  |  |                        |  |  |                        |
| J=3                        | 0.0155                                 | 0.0155                                 | 0.0038                 | 0.0536                                 | 0.0536                                 | 0.0239                 |
| J=4                        | 0.1130                                 | 0.1337                                 | 0.0038                 | 0.1871                                 | 0.2111                                 | 0.0239                 |

Table 2: Performance of  $\tilde{q}_n^{\text{RP},1}(\beta_n|x)$ ,  $\tilde{q}_n^{\text{RP},2}(\beta_n|x)$  and  $\hat{q}_n(\beta_n|x)$  with  $\beta_n = 0.05$  (top),  $\beta_n = 0.01$  (middle) and  $\beta_n = 0.005$  (bottom) – Results averaged on 400 simulations with  $n = 200$ .

## 4.2 Benchmark nonparametric estimators of $\gamma(x)$ and $q(\beta_n|x)$

Alternative modern smoothing techniques were discussed in *e.g.* [2, Section 7.5]. For comparison, we focus on the prominent local polynomial maximum likelihood estimation. This contribution fits a generalized Pareto (GP) model to the exceedances  $Z_i^x = Y_j - u_x$  given  $Y_j > u_x$ , for a high threshold  $u_x$ , where  $j$  denotes the original index of the  $i$ th exceedance. Let  $N_x$  be the number of all exceedances over  $u_x$  and rearrange the indices of the explanatory variable such that  $X_i$  denotes the covariate observation associated with exceedance  $Z_i^x$ . If  $\tilde{g}(z; \sigma, \gamma)$  stands for the GP density, then the local polynomial maximum likelihood approach maximizes the kernel weighted log-likelihood function

$$L_{N_x}(\beta_1, \beta_2) = \frac{1}{N_x} \sum_{i=1}^{N_x} \log \tilde{g} \left( Z_i^x; \sum_{j=0}^{p_1} \beta_{1j} (X_i - x)^j, \sum_{j=0}^{p_2} \beta_{2j} (X_i - x)^j \right) K_h(X_i - x)$$

with respect to  $(\beta'_1, \beta'_2) = (\beta_{10}, \dots, \beta_{1p_1}, \beta_{20}, \dots, \beta_{2p_2})$  to get the estimates  $\hat{\sigma}_n^{GP}(x) = \hat{\beta}_{10}$  and  $\hat{\gamma}_n^{GP}(x) = \hat{\beta}_{20}$  of the parameter functions  $\sigma(x)$  and  $\gamma(x)$  of the GP distribution fitted to the exceedances over  $u_x$ . Note that local polynomial fitting also provides estimates of the derivatives of  $\sigma(x)$  and  $\gamma(x)$  up to order  $p_1$  and  $p_2$ , respectively. In order to not overload the estimation procedure, we confine ourselves to  $p_1 = p_2 = 0$ . The Monte Carlo results for  $\hat{\gamma}_n^{GP}(\cdot)$  are reported in Table 3 (l-h.s). For each simulation, we used the parameters  $(h, u)$  that minimize the  $\text{MSE}\{\hat{\gamma}_n^{GP}(\cdot)\}$ , with the bandwidth ranging over the grid  $\mathcal{H}$  described above and the threshold ranging over the  $\alpha$ th sample quantiles of  $Y$ , where  $\alpha \in \mathcal{A}$ . The estimator  $\hat{\gamma}_n^{GP}$  has clearly smaller MSEs than the  $\hat{\gamma}_n^{RP}$  estimators in the Gaussian and Student error models, but it seems to be less efficient in the Beta error model than both  $\hat{\gamma}_n^{RP,1}$  and  $\hat{\gamma}_n^{RP,2}$  for  $J = 4$  and  $r = 1/J$ . From a theoretical point of view, it should be clear that the pointwise asymptotic normality of  $\hat{\gamma}_n^{GP}(x)$  is proved in [1] only in case  $\gamma(x) > 0$ . Moreover, the proof is restricted to the setting where the design points  $X_i$  are deterministic.

|                 | MSE $\{\hat{\gamma}_n^{GP}\}$ | Bias $\{\hat{\gamma}_n^{GP}\}$ | $\beta_n = 0.05$ |        | $\beta_n = 0.01$ |        | $\beta_n = 0.005$ |         |
|-----------------|-------------------------------|--------------------------------|------------------|--------|------------------|--------|-------------------|---------|
|                 |                               |                                | MSE              | Bias   | MSE              | Bias   | MSE               | Bias    |
| <b>Gaussian</b> | 0.1324                        | -0.2671                        | 0.0184           | 0.0974 | 0.0278           | 0.0952 | 0.0315            | 0.0861  |
| <b>Student</b>  | 0.1310                        | -0.2238                        | 0.1346           | 0.1526 | 0.6924           | 0.0895 | 1.0232            | -0.0452 |
| <b>Beta</b>     | 0.0675                        | -0.0221                        | 0.0364           | 0.1578 | 0.0659           | 0.2067 | 0.0786            | 0.2242  |

Table 3: Performance of  $\hat{\gamma}_n^{GP}$  and  $\hat{q}_n^{GP}(\beta_n|\cdot)$  – Results averaged on 400 simulations with  $n = 200$ .

On the other hand, as suggested in [2, Section 7.5.2] and [1], the extreme conditional quantile  $q(\beta_n|x)$  can be estimated by

$$\hat{q}_n^{GP}(\beta_n|x) := u_x + \frac{\hat{\sigma}_n^{GP}(x)}{\hat{\gamma}_n^{GP}(x)} \left[ \left( \frac{n_{xh}^* \beta_n}{k_x} \right)^{-\hat{\gamma}_n^{GP}(x)} - 1 \right]$$

where  $n_{xh}^*$  is the number of observations in  $[x - h, x + h]$  and  $k_x$  is the number of exceedances receiving positive weight. Table 3 (r-h.s) reports the Monte Carlo estimates obtained by using in each simulation the parameters  $(h, u)$  that minimize the  $\text{MSE}\{\hat{q}_n^{GP}(\beta_n|\cdot)\}$ , where  $h \in \mathcal{H}$  and  $u$  ranges over the  $\alpha$ th sample quantiles of  $Y$  with  $\alpha \in \mathcal{A}$ . In all cases, the regression quantile (RQ) estimator  $\hat{q}_n(\beta_n|\cdot)$  do appear to be more efficient than  $\hat{q}_n^{GP}(\beta_n|\cdot)$ . Compared with the  $\tilde{q}_n^{RP}(\beta_n|\cdot)$  estimators (for  $J = 3$  and  $r = 1/J$ ),  $\hat{q}_n^{GP}(\beta_n|\cdot)$  seems to be more efficient only in the Gaussian error model for  $\beta_n = 0.005 = 1/n$ , but not by much. A typical realization of the experiment in each simulated scenario is shown in Figure 1, where the smoothing parameters of each estimator were chosen in such a way to minimize its MSE. From a theoretical viewpoint, unlike our estimators  $\hat{q}_n(\beta_n|x)$  and  $\tilde{q}_n^{RP}(\beta_n|x)$ , the asymptotic distribution of  $\hat{q}_n^{GP}(\beta_n|x)$  is not elucidated yet.

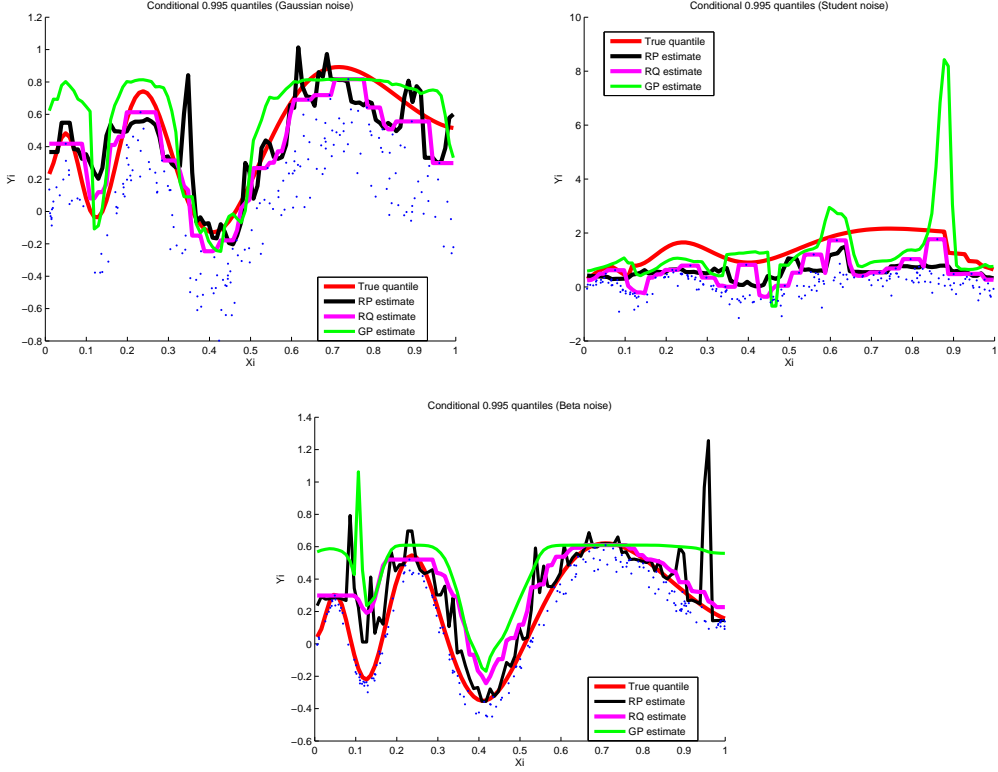


Figure 1: Typical realizations for simulated samples of size  $n = 200$ . From left to right and from top to bottom,  $Y|X$  is Gaussian, Student, Beta. The true quantile function  $q(\beta_n|\cdot)$  in red with  $\beta_n = 1/n$ . Its estimators  $\hat{q}_n(\beta_n|\cdot)$  in magenta,  $\tilde{q}_n^{RP,1}(\beta_n|\cdot) \equiv \tilde{q}_n^{RP,2}(\beta_n|\cdot)$  in black with  $r = 1/J$  and  $J = 3$ , and  $\hat{q}_n^{GP}(\beta_n|\cdot)$  in green. The observations  $(X_i, Y_i)$  are depicted as blue points.

### 4.3 Data-driven rules for selecting the parameters $h$ and $\alpha_n$

The use of the ‘RQ’ estimator  $\hat{q}_n(\beta_n|x) \equiv \hat{q}_n(\beta_n|x; h)$ , which relies on the inversion of  $\hat{F}_n(\cdot)$ , requires only the choice of the bandwidth  $h$  in an interval  $\mathcal{H}$  of lower and upper bounds given respectively by, say,  $h_{min} := \max_{1 \leq i < n} (X_{(i+1)} - X_{(i)})$  and  $h_{max} := (X_{(n)} - X_{(1)})/4$ . One way to select this parameter is by employing the cross-validation criterion as in [7] to obtain

$$h_{cv} = \arg \min_{h \in \mathcal{H}} \sum_{i=1}^n \sum_{j=1}^n \left\{ \mathbb{I}(Y_i \geq Y_j) - \hat{F}_{n,-i}(Y_j|X_i) \right\}^2,$$

where  $\hat{F}_{n,-i}(\cdot)$  is the estimator  $\hat{F}_n(\cdot)$  computed from the sample  $\{(X_j, Y_j), 1 \leq j \leq n, j \neq i\}$ . The empirical procedure of [33] could be used to get the alternative data-driven global bandwidth

$$h_{yj} = h_{cv} \left( \frac{\beta_n(1 - \beta_n)}{\phi(\Phi^{-1}(\beta_n))^2} \right)^{1/5},$$

where  $\phi$  and  $\Phi$  stand respectively for the standard normal density and distribution functions. However, the use of the ‘RP’ estimators  $\tilde{q}_n^{RP,i}(\beta_n|x)$  and  $\hat{\gamma}_n^{RP,i}(x)$ , for  $i = 1, 2$ , requires in addition the selection of an appropriate order  $\alpha_n$ . To simplify the discussion we set  $\alpha_n$  at  $k/n_{xh}^*$ , where the integer  $k$  varies between 1 and  $n_{xh}^* - 1$ , for each  $h \in \mathcal{H}$ . We also consider the value  $J = 3$  for which  $\hat{\gamma}_n^{RP,1}(x) \equiv \hat{\gamma}_n^{RP,2}(x) := \hat{\gamma}_n^{RP}(x; h, k)$  and  $\tilde{q}_n^{RP,1}(\beta_n|x) \equiv \tilde{q}_n^{RP,2}(\beta_n|x) := \tilde{q}_n^{RP}(\beta_n|x; h, k)$ , with  $r = 1/J$ . An empirical way to decide what values of  $(h, k)$  should one use to compute the estimates in practice could be the automatic *ad hoc* data driven-rule employed in [6]. The main idea is to

evaluate first the estimates, for each  $x$  in a chosen grid of values, and then to select the parameters where the variation of the results is the smallest. This can be achieved in two ways:

### Selecting $h$ and $k$ separately.

*Step 1.* Select a data-driven global bandwidth  $h$ , for example,  $h_{cv}$  or  $h_{yj}$ .

*Step 2.* Evaluate  $\tilde{q}_n^{RP}(\beta_n|x; h, k)$  at  $k = 1, \dots, n_{xh}^* - 1$ . Then compute the standard deviation of the estimates over a ‘window’ of (say  $[\sqrt{n_{xh}^*}]$ ) successive values of  $k$ . The value of  $k$  where this standard deviation is minimal defines the desired parameter.

The same considerations evidently apply to  $\hat{\gamma}_n^{RP}(x)$  and to the ‘benchmark’ estimators  $\hat{\gamma}_n^{GP}(x) := \hat{\gamma}_n^{GP}(x; h, k)$  and  $\hat{q}_n^{GP}(\beta_n|x) := \hat{q}_n^{GP}(\beta_n|x; h, k)$ , defined in Subsection 4.2, with the covariate dependent threshold being  $u_x := Y_{(n_{xh}^* - k)}^x$ , and  $Y_{(1)}^x \leq \dots \leq Y_{(n_{xh}^*)}^x$  being the sequence of ascending order statistics corresponding to the  $Y_i$ ’s such that  $|X_i - x| \leq h$ .

The main difficulty when employing such a separate choice of  $h$  and  $k$  is that both  $\tilde{q}_n^{RP}(\beta_n|x; h_{cv}, k)$  and  $\hat{q}_n^{GP}(\beta_n|x; h_{cv}, k)$ , respectively  $\hat{\gamma}_n^{RP}(x; h_{cv}, k)$  and  $\hat{\gamma}_n^{GP}(x; h_{cv}, k)$ , as functions of  $k$  may be so unstable that reasonable values of  $k$  (which would correspond to the true value of  $q(\beta_n|x)$ , respectively  $\gamma(x)$ ) may be hidden in the graphs. In result, the estimators may exhibit considerable volatility as functions of  $x$  itself.

A typical realization is shown in Figure 2 when the bandwidths  $h_{cv}$  (left panels) and  $h_{yj}$  (right panels) are used in Step 1. It may be seen that the method affords reasonable estimates in both Gaussian and Beta error models regarding the difficult curvature of the extreme quantile regression and the very small sample size  $n = 200$ . However, it seems that the method fails in the case of Student noise, where the superiority of  $\tilde{q}_n^{RP}(\beta_n|\cdot)$  over both  $\hat{q}_n^{GP}(\beta_n|\cdot)$  and  $\hat{q}_n(\beta_n|\cdot)$ , demonstrated via the Monte Carlo study, is clearly sacrificed. This failure is probably due to the arbitrary choice  $(3, 1/3)$  of the parameters  $(J, r)$  in  $\tilde{q}_n^{RP}(\beta_n|\cdot)$ . It might also be seen that, apart from the student error model, the three estimators  $\tilde{q}_n^{RP}(\beta_n|\cdot)$ ,  $\hat{q}_n(\beta_n|\cdot)$  and  $\hat{q}_n^{GP}(\beta_n|\cdot)$  point toward similar results.

### Selecting $h$ and $k$ simultaneously.

*Step 1.* For each  $h \in \mathcal{H}$ , proceed to Step 2 described in the separate parameters’ selection. Set the value of  $k$  where the standard deviation is minimal to be  $k_{xh}$  and calculate the corresponding estimate  $\tilde{q}_n^{RP}(\beta_n|x; h, k_{xh})$ .

*Step 2.* Compute the standard deviation of the estimates  $\tilde{q}_n^{RP}(\beta_n|x; h, k_{xh})$  over a window of (say 10) successive values of  $h$ . Select the bandwidth where the standard deviation is minimal and then evaluate the corresponding estimate.

In our simulations, we used a refined grid  $\mathcal{H}$  of 50 points between  $\min(h_{cv}, h_{yj} - h_{cv})$  and  $h_{yj} + 2h_{cv}$ . Any other limit bounds of  $\mathcal{H}$  could of course be chosen near  $h_{cv}$  below and near  $h_{yj}$  above. See Figure 3 for a typical realization in each simulated scenario. Here also the method is not without disadvantage as can be seen from the case of Student noise, where good results require a large sample size.

## 4.4 Concluding remarks

*Monte Carlo evidence.* The experiments indicate that  $\hat{q}_n(\beta_n|x)$  is efficient relative to the modern smoothing estimator  $\hat{q}_n^{GP}(\beta_n|x)$  first introduced in [1, 2]. The simulations also indicate that the performance of the alternative estimator  $\tilde{q}_n^{RP,1}(\beta_n|x)$  is quite remarkable in comparison with its analog  $\tilde{q}_n^{RP,2}(\beta_n|x)$ , at least in terms of MSE. In comparison with  $\hat{q}_n(\beta_n|x)$  and  $\hat{q}_n^{GP}(\beta_n|x)$ , the variability and the bias of both  $\tilde{q}_n^{RP,1}(\beta_n|x)$  and  $\tilde{q}_n^{RP,2}(\beta_n|x)$  are quite respectable and it seems that the heavier is the conditional tail, the better the estimators  $\tilde{q}_n^{RP}(\beta_n|x)$  are. It should be also clear that  $\tilde{q}_n^{RP,1}(\beta_n|x)$  and  $\tilde{q}_n^{RP,2}(\beta_n|x)$  can be improved appreciably by tuning the choice of the parameters  $J$  and  $r$ .

*Tuning parameters selection in practice.* The simulations worked reasonably well for our ‘ad hoc’

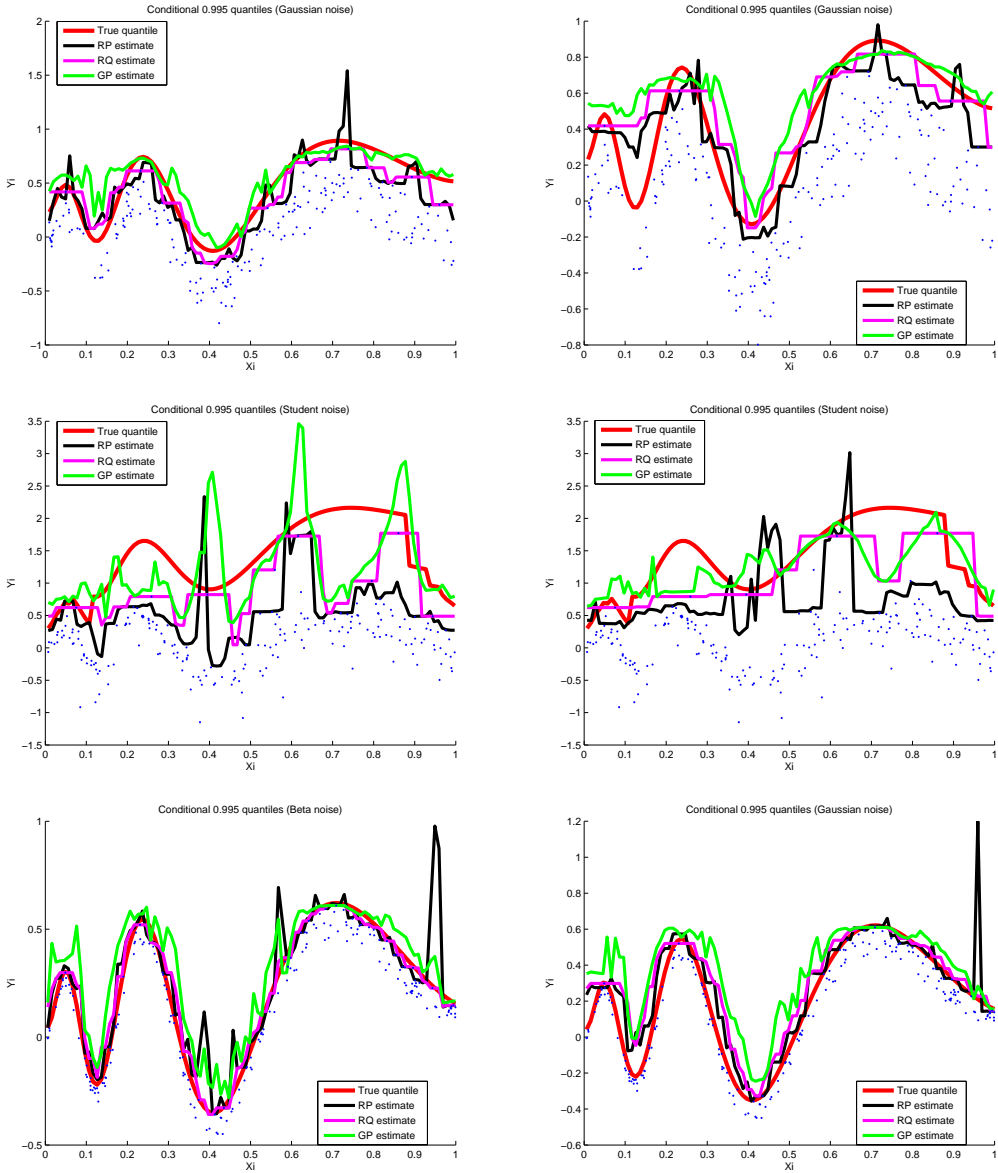


Figure 2: *Separate parameters' selection using the same simulated samples as in Figure 1. From left to right, the used bandwidth is  $h_{cv}$ ,  $h_{yj}$ . From top to bottom,  $Y|X$  is Gaussian, Student, Beta. The true quantile function  $q(\beta_n|\cdot)$  in red with  $\beta_n = 1/n$ . Its estimators  $\hat{q}_n(\beta_n|\cdot)$  in magenta,  $\tilde{q}_n^{RP,1}(\beta_n|\cdot) \equiv \tilde{q}_n^{RP,2}(\beta_n|\cdot)$  in black with  $r = 1/J$  and  $J = 3$ , and  $\hat{q}_n^{GP}(\beta_n|\cdot)$  in green. The observations  $(X_i, Y_i)$  are depicted as blue points.*

selection methods except for the heavy-tailed case, corresponding to generally severe events. A sensible practice would be to verify whether the resulting ‘RP’, ‘GP’ and ‘RQ’ estimators point toward similar conclusions: the hard question of how to pick out the smoothing parameters  $(h, k)$  simultaneously in an optimal way might thus become less urgent. In contrast, if the estimators look clearly different, this might diagnose a heavy-tailed conditional distribution with a great variability in severity: thereby our technique might be viewed as an exploratory tool, rather than as a method for final analysis. Doubtlessly, further work to define a concept of selecting appropriate values for the crucial parameters  $(J, r)$  in the  $\tilde{q}_n^{RP}(\beta_n|x)$  estimators will yield new refinements.

*The case of multiple covariates.* We have discussed the asymptotic distributional properties of both

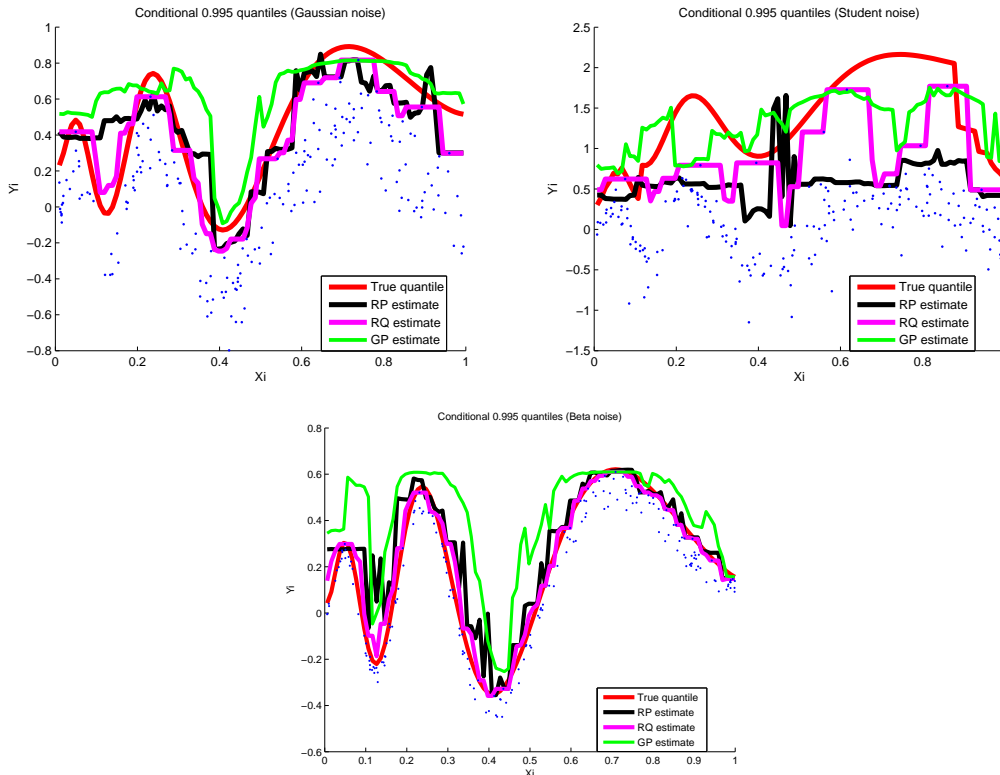


Figure 3: *Simultaneous parameters’ selection using the same simulated samples as above. From left to right and from top to bottom,  $Y|X$  is Gaussian, Student, Beta. The true quantile function  $g(\beta_n|\cdot)$  in red with  $\beta_n = 1/n$ . Its estimators  $\hat{q}_n(\beta_n|\cdot) = \hat{q}_n(\beta_n|\cdot, h_{cv})$  in magenta,  $\tilde{q}_n^{RP,1}(\beta_n|\cdot) \equiv \tilde{q}_n^{RP,2}(\beta_n|\cdot)$  in black with  $(r, J) = (1/3, 3)$ , and  $\hat{q}_n^{GP}(\beta_n|\cdot)$  in green. The  $(X_i, Y_i)$ ’s are depicted as blue points.*

estimators  $\hat{q}_n(\beta_n|x)$  and  $\tilde{q}_n^{RP}(\beta_n|x)$  in detail in Sections 2-3 for multiple regressors  $X \in \mathbb{R}^p$ , but our contributions are probably only of a theoretical value in the case  $p > 1$ . Indeed, as in the ordinary setting where the quantile order does not depend on the sample size, the kernel-smoothing method suffers from the ‘curse of dimensionality’. In our setting of extreme quantile regression, the curse is exacerbated by several degrees of magnitude and drastically increases in higher dimensions. To overcome this vexing defect, one can use dimension reduction techniques such as ADE (Average Derivative Estimator), see for instance [21]. Nevertheless, the theoretical properties of such methods are not yet established in the extreme-value framework.

## 5 Data example

Data on 123 American electric utility companies were collected and the aim is to investigate the economic efficiency of these companies (see, *e.g.*, [18]). A possible way to measure this efficiency is by looking at the maximum level of produced goods which is attainable for a given level of inputs-usage. From a statistical point of view, this problem translates into studying the upper boundary of the set of possible inputs  $X$  and outputs  $Y$ , the so-called cost/econometric frontier in production theory. Hendricks and Koenker [22] stated: “In the econometric literature on the estimation of production technologies, there has been considerable interest in estimating so called frontier production models that correspond closely to models for extreme quantiles of a stochastic production surface”. The present paper may be viewed as the first ‘purely’ nonparametric work to actually investigate theoretically the idea of Hendricks and Koenker.

For our illustration purposes, we used the measurements on the variable  $Y = \log(Q)$ , with  $Q$

being the production output of a firm, and the variable  $X = \log(C)$ , with  $C$  being the total cost involved in the production. Figure 4 shows the  $n = 123$  observations, together with estimated extreme conditional quantiles  $\hat{q}_n(\beta_n|x)$ ,  $\hat{q}_n^{GP}(\beta_n|x)$  and  $\tilde{q}_n^{\text{RP},1}(\beta_n|x) = \tilde{q}_n^{\text{RP},2}(\beta_n|x)$  at  $r = 1/J$  with  $J = 3$ . Given the small sample size, it was enough to use  $\beta_n = 1/n$  in describing the conditional distribution tails. For selecting the window width  $h$  and the number  $k$  of extremes, we maintained the automatic empirical data-driven rules described above. It appears that the extreme conditional quantile estimates are similar for both simultaneous (top) and separate (bottom) selection methods. Following their evolution, it may be seen that the American electric utility data do not correspond to the situation hoped for by the practitioners of a heavily short-tailed production process. Indeed, one may distinguish between two different behaviors of the extreme regression quantiles: They indicate a short-tailed conditional distribution for companies working at (transformed) input-factors larger than, say, 2. In contrast, the tail distribution for the smallest companies with inputs  $X_i < 2$  seems to be moderately heavy. Therefore, the theoretical economic assumption that producers should operate on the upper boundary of the joint support of  $(X, Y)$  rather than on its interior is clearly not fulfilled here, revealing a certain lack of production performance in this sector of activity. The estimated graph of  $\tilde{q}_n^{\text{RP},1}(\beta_n|x)$ ,  $\hat{q}_n(\beta_n|x)$  or  $\hat{q}_n^{GP}(\beta_n|x)$  might be interpreted as the set of the most efficient firms. It is then clear that the firms achieve significantly lesser output than that predicted by the extremal quantile frontiers. This indicates a relative economic inefficiency especially in the population of the (sparse) smallest companies in terms of inputs.

## 6 Appendix: Proofs

### 6.1 Preliminary results

We begin with a homogeneous property of the quantile function.

**Lemma 1.** *Suppose (A.1) holds. If  $\alpha_n \rightarrow 0$  as  $n \rightarrow \infty$ , then,*

$$\lim_{n \rightarrow \infty} \frac{q(\xi\alpha_n|x)}{q(\alpha_n|x)} = \xi^{-(\gamma(x) \vee 0)},$$

for all  $\xi > 0$ .

*Proof.* From (4), we have

$$\frac{q(\xi\alpha_n|x)}{q(\alpha_n|x)} = 1 + K_{\gamma(x)}(1/\xi) \frac{a(q(\alpha_n|x)|x)}{q(\alpha_n|x)}(1 + o(1))$$

and the conclusion follows using (5). ■

The following lemma states that the convergence in (3) is uniform.

**Lemma 2.** *Under (A.1), if  $z_n(x) \uparrow y_F(x)$  as  $n \rightarrow \infty$ , then for all sequence of functions  $t_n(x)$  such that  $t_n(x) \rightarrow t_0(x)$  as  $n \rightarrow \infty$  where  $t_0(x)$  is such that there exists  $\eta > 0$  for which  $1 + \gamma(x)t_0(x) \geq \eta$  then,*

$$\lim_{n \rightarrow \infty} \frac{\bar{F}(z_n(x) + t_n(x)a(z_n(x)|x)|x)}{\bar{F}(z_n(x)|x)} = \frac{1}{K_{\gamma(x)}^{-1}(t_0(x))}.$$

*Proof.* Since  $t_n(x) \rightarrow t_0(x)$  as  $n \rightarrow \infty$ , for all  $\varepsilon_1 > 0$  such that  $|\gamma(x)|\varepsilon_1 < \eta$ , there exists  $N_1 \geq 0$  such that for all  $n \geq N_1$ ,  $t_0(x) - \varepsilon_1 \leq t_n(x) \leq t_0(x) + \varepsilon_1$ . Since  $a(z_n(x)|x) > 0$  and  $\bar{F}(\cdot|x)$  is a

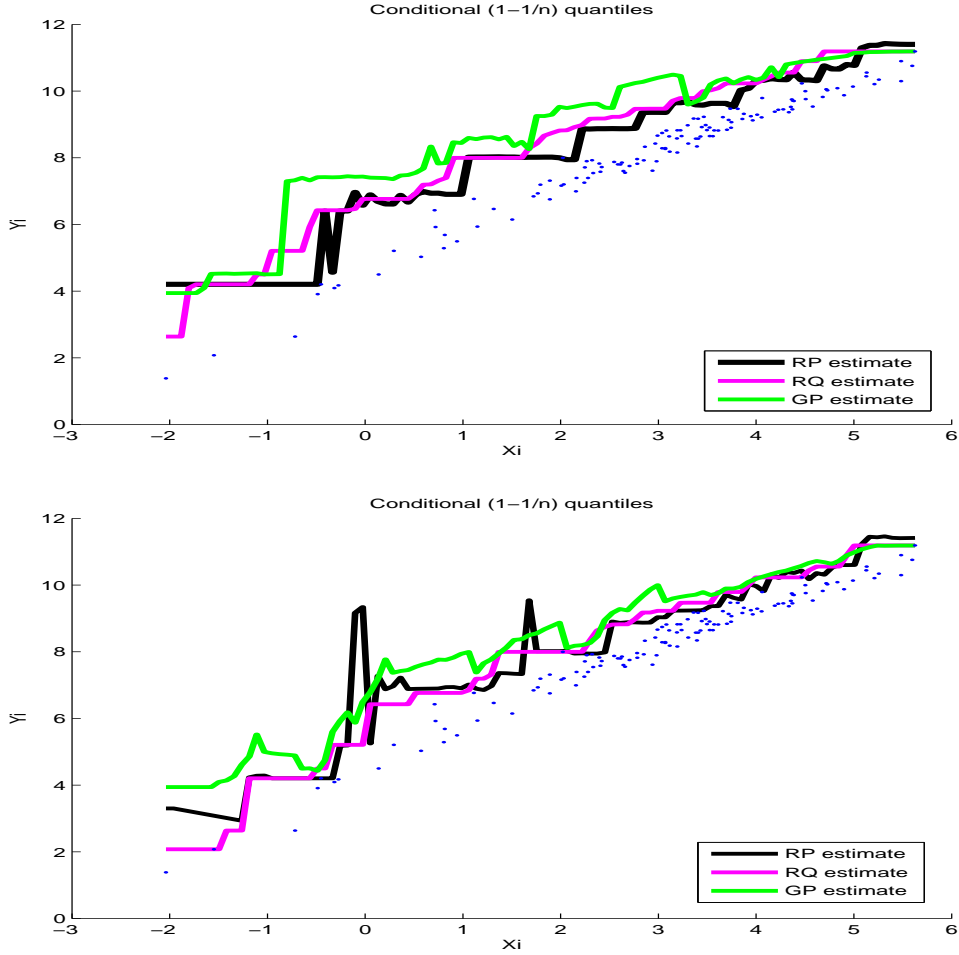


Figure 4: Scatterplot of the American electric utility data (blue points). Results obtained via the simultaneous (top) and separate (bottom) selection methods for  $\beta_n = 1/n$ : the estimator  $\hat{q}_n(\beta_n|\cdot)$  is drawn in magenta with  $h_{yj}$  (top) and  $h_{cv}$  (bottom),  $\hat{q}_n^{GP}(\beta_n|\cdot)$  in green and  $\hat{q}_n^{RP,1}(\beta_n|\cdot)$  in black with  $(r, J) = (1/3, 3)$ .

decreasing function, we have:

$$\begin{aligned} \frac{\bar{F}(z_n(x) + (t_0(x) + \varepsilon_1)a(z_n(x)|x)|x)}{\bar{F}(z_n(x)|x)} &\leq \frac{\bar{F}(z_n(x) + t_n(x)a(z_n(x)|x)|x)}{\bar{F}(z_n(x)|x)} \\ &\leq \frac{\bar{F}(z_n(x) + (t_0(x) - \varepsilon_1)a(z_n(x)|x)|x)}{\bar{F}(z_n(x)|x)}. \end{aligned}$$

Now, since  $|\gamma(x)|\varepsilon_1 < \eta$ , it is easy to check that  $1 + \gamma(x)(t_0(x) + \varepsilon_1) \wedge 1 + \gamma(x)(t_0(x) - \varepsilon_1) > 0$ . Hence, under **(A.1)**, for all  $\varepsilon_2 > 0$ , there exists  $N_2 \geq 0$  such that for all  $n \geq N_2$

$$\frac{1 - \varepsilon_2}{K_{\gamma(x)}^{-1}(t_0(x) + \varepsilon_1)} \leq \frac{\bar{F}(z_n(x) + t_n(x)a(z_n(x)|x)|x)}{\bar{F}(z_n(x)|x)} \leq \frac{1 + \varepsilon_2}{K_{\gamma(x)}^{-1}(t_0(x) - \varepsilon_1)}.$$

Since  $\varepsilon_1$  and  $\varepsilon_2$  can be chosen arbitrarily small, this concludes the proof.  $\blacksquare$

Let us remark that the kernel estimator (1) can be rewritten as  $\hat{F}_n(y|x) = \hat{\psi}_n(y, x)/\hat{g}_n(x)$  where

$$\hat{\psi}_n(y, x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \mathbb{I}\{Y_i > y\}$$



is an estimator of  $\psi(y, x) = \bar{F}(y|x)g(x)$  and  $\hat{g}_n(x)$  is the classical kernel estimator of the pdf  $g(x)$  defined by:

$$\hat{g}_n(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i).$$

Lemma 3 gives standard results on the kernel estimator (see [26] for a proof) whereas Lemma 4 is dedicated to the asymptotic properties of  $\hat{\psi}_n(y, x)$ .

**Lemma 3.** *Suppose (A.2), (A.3) hold. If  $nh^p \rightarrow \infty$ , then, for all  $x \in \mathbb{R}^p$  such that  $g(x) > 0$ ,*

- (i)  $\mathbb{E}(\hat{g}_n(x) - g(x)) = O(h)$ ,
- (ii)  $\text{var}(\hat{g}_n(x)) = \frac{g(x)\|K\|_2^2}{nh^p}(1 + o(1))$ .

Therefore, under the assumptions of the above lemma,  $\hat{g}_n(x)$  converges to  $g(x)$  in probability.

Let us introduce some further notations. In the following,  $y_n(x)$  is a sequence such that  $y_n(x) \uparrow y_F(x)$  and  $y_{n,j}(x) = y_n(x) + K_{\gamma(x)}(1/\tau_j)a(y_n|x)(1 + o(1))$  for all  $j = 1, \dots, K$ . Recall that  $0 < \tau_J < \dots < \tau_2 < \tau_1 \leq 1$ . Moreover, the oscillations of the csf are controlled by

$$\omega_n(x) := \max_{j=1, \dots, J} \sup_{x' \in B(x, h)} \left| \frac{\bar{F}(y_{n,j}(x)|x')}{\bar{F}(y_{n,j}(x)|x)} - 1 \right|.$$

**Lemma 4.** *Suppose (A.1), (A.2) and (A.3) hold and let  $x \in \mathbb{R}^p$  such that  $g(x) > 0$ . If  $\omega_n(x) \rightarrow 0$  and  $nh^p \bar{F}(y_n(x)|x) \rightarrow \infty$  then,*

- (i)  $\mathbb{E}(\hat{\psi}_n(y_{n,j}(x), x)) = \psi(y_{n,j}(x), x) (1 + O(\omega_n(x)) + O(h))$ , for  $j = 1, \dots, J$ .
- (ii) *The random vector*

$$\left\{ \sqrt{nh^p \psi(y_n(x), x)} \left( \frac{\hat{\psi}_n(y_{n,j}(x), x) - \mathbb{E}(\hat{\psi}_n(y_{n,j}(x), x))}{\psi(y_{n,j}(x), x)} \right) \right\}_{j=1, \dots, J}$$

*is asymptotically Gaussian, centered, with covariance matrix  $\|K\|_2^2 V$  where  $V_{j,j'} = \tau_{j \wedge j'}^{-1}$  for  $(j, j') \in \{1, \dots, J\}^2$ .*

*Proof.* (i) Since the  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  are identically distributed, we have

$$\mathbb{E}(\hat{\psi}_n(y_{n,j}(x), x)) = \int_{\mathbb{R}^p} K_h(x - t) \bar{F}(y_{n,j}(x)|t) g(t) dt = \int_S K(u) \bar{F}(y_{n,j}(x)|x - hu) g(x - hu) du,$$

under (A.3). Let us now consider

$$|\mathbb{E}(\hat{\psi}_n(y_{n,j}(x), x)) - \psi(y_{n,j}(x), x)| \leq \bar{F}(y_{n,j}(x)|x) \int_S K(u) |g(x - hu) - g(x)| du \quad (8)$$

$$+ \bar{F}(y_{n,j}(x)|x) \int_S K(u) \left| \frac{\bar{F}(y_{n,j}(x)|x - hu)}{\bar{F}(y_{n,j}(x)|x)} - 1 \right| g(x - hu) du. \quad (9)$$

Under (A.2), and since  $g(x) > 0$ , we have

$$(8) \leq \bar{F}(y_{n,j}(x)|x) c_g h \int_S d(u, 0) K(u) du = \psi(y_{n,j}(x), x) O(h), \quad (10)$$

and, in view of (10),

$$\begin{aligned}
(9) &\leq \bar{F}(y_{n,j}(x)|x)\omega_n(x) \int_S K(u)g(x-hu)du = \bar{F}(y_{n,j}(x)|x)\omega_n(x)g(x)(1+o(1)) \\
&= \psi(y_{n,j}(x), x)\omega_n(x)(1+o(1)).
\end{aligned} \tag{11}$$

Combining (10) and (11) concludes the first part of the proof.

(ii) Let  $\beta \neq 0$  in  $\mathbb{R}^J$ ,  $\Lambda_n(x) = (nh^p\psi(y_n(x), x))^{-1/2}$ , and consider the random variable

$$\begin{aligned}
\Psi_n &= \sum_{j=1}^J \beta_j \left( \frac{\hat{\psi}_n(y_{n,j}(x), x) - \mathbb{E}(\hat{\psi}_n(y_{n,j}(x), x))}{\Lambda_n(x)\psi(y_{n,j}(x), x)} \right) \\
&= \sum_{i=1}^n \frac{1}{n\Lambda_n(x)} \left\{ \sum_{j=1}^J \frac{\beta_j K_h(x - X_i) \mathbb{I}\{Y_i \geq y_{n,j}(x)\}}{\psi(y_{n,j}(x), x)} - \mathbb{E} \left( \sum_{j=1}^J \frac{\beta_j K_h(x - X_i) \mathbb{I}\{Y_i \geq y_{n,j}(x)\}}{\psi(y_{n,j}(x), x)} \right) \right\} \\
&:= \sum_{i=1}^n Z_{i,n}.
\end{aligned}$$

Clearly,  $\{Z_{i,n}, i = 1, \dots, n\}$  is a set of centered, independent and identically distributed random variables with variance

$$\text{var}(Z_{i,n}) = \frac{1}{n^2 h^{2p} \Lambda_n^2(x)} \text{var} \left( \sum_{j=1}^J \beta_j K \left( \frac{x - X_i}{h} \right) \frac{\mathbb{I}\{Y_i \geq y_{n,j}(x)\}}{\psi(y_{n,j}(x), x)} \right) = \frac{1}{n^2 h^p \Lambda_n^2(x)} \beta^t B \beta,$$

where  $B$  is the  $J \times J$  covariance matrix with coefficients defined for  $(j, j') \in \{1, \dots, J\}^2$  by

$$\begin{aligned}
B_{j,j'} &= \frac{A_{j,j'}}{\psi(y_{n,j}(x), x)\psi(y_{n,j'}(x), x)}, \\
A_{j,j'} &= \frac{1}{h^p} \text{cov} \left( K \left( \frac{x - X}{h} \right) \mathbb{I}\{Y \geq y_{n,j}(x)\}, K \left( \frac{x - X}{h} \right) \mathbb{I}\{Y \geq y_{n,j'}(x)\} \right) \\
&= \|K\|_2^2 \mathbb{E} \left( \frac{1}{h^p} Q \left( \frac{x - X}{h} \right) \mathbb{I}\{Y \geq y_{n,j}(x) \vee y_{n,j'}(x)\} \right) \\
&\quad - h^p \mathbb{E}(K_h(x - X) \mathbb{I}\{Y \geq y_{n,j}(x)\}) \mathbb{E}(K_h(x - X) \mathbb{I}\{Y \geq y_{n,j'}(x)\}),
\end{aligned}$$

with  $Q(\cdot) := K^2(\cdot)/\|K\|_2^2$  also satisfying assumption **(A.3)**. As a consequence, the three above expectations are of the same nature. Thus, remarking that, for  $n$  large enough,  $y_{n,j}(x) \vee y_{n,j'}(x) = y_{n,j \vee j'}(x)$ , part (i) of the proof implies

$$\begin{aligned}
A_{j,j'} &= \|K\|_2^2 \psi(y_{n,j \vee j'}(x), x) (1 + O(\omega_n(x)) + O(h)) \\
&\quad - h^p \psi(y_{n,j}(x), x) \psi(y_{n,j'}(x), x) (1 + O(\omega_n(x)) + O(h))
\end{aligned}$$

leading to

$$\begin{aligned}
B_{j,j'} &= \frac{\|K\|_2^2}{\psi(y_{n,j \wedge j'}(x), x)} (1 + O(\omega_n(x)) + O(h)) - h^p (1 + O(\omega_n(x)) + O(h)) \\
&= \frac{\|K\|_2^2}{\psi(y_{n,j \wedge j'}(x), x)} (1 + o(1)),
\end{aligned}$$

since  $\psi(y_{n,j \wedge j'}(x), x) \rightarrow 0$  as  $n \rightarrow \infty$ . Now, from Lemma 2,

$$\lim_{n \rightarrow \infty} \frac{\psi(y_{n,j \wedge j'}(x), x)}{\psi(y_n(x), x)} = \frac{1}{K_{\gamma(x)}^{-1}(K_{\gamma(x)}(1/\tau_{j \wedge j'}))} = \tau_{j \wedge j'}$$

entailing

$$B_{j,j'} = \frac{\|K\|_2^2 V_{j,j'}}{\psi(y_n(x), x)} (1 + o(1)),$$

and therefore,  $\text{var}(Z_{i,n}) \sim \|K\|_2^2 \beta^t V \beta / n$ , for all  $i = 1, \dots, n$ . As a preliminary conclusion, the variance of  $\Psi_n$  converges to  $\|K\|_2^2 \beta^t V \beta$ . Consequently, Lyapounov criteria for the asymptotic normality of sums of triangular arrays reduces to  $\sum_{i=1}^n \mathbb{E} |Z_{i,n}|^3 = n \mathbb{E} |Z_{1,n}|^3 \rightarrow 0$ . Remark that  $Z_{1,n}$  is a bounded random variable:

$$|Z_{1,n}| \leq \frac{2\|K\|_\infty \sum_{j=1}^J |\beta_j|}{n \Lambda_n(x) h^p \psi(y_{n,J}, x)} = \frac{2}{\tau_J} \|K\|_\infty \sum_{j=1}^J |\beta_j| \Lambda_n(x) (1 + o(1))$$

and thus,

$$\begin{aligned} n \mathbb{E} |Z_{1,n}|^3 &\leq \frac{2}{\tau_J} \|K\|_\infty \sum_{j=1}^J n |\beta_j| \Lambda_n(x) \text{var}(Z_{1,n}) (1 + o(1)) \\ &= \frac{2}{\tau_J} \|K\|_\infty \|K\|_2^2 \sum_{j=1}^J |\beta_j| \beta^t V \beta \Lambda_n(x) (1 + o(1)) \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ . As a conclusion,  $\Psi_n$  converges in distribution to a centered Gaussian random variable with variance  $\|K\|_2^2 \beta^t V \beta$  for all  $\beta \neq 0$  in  $\mathbb{R}^p$ . The result is proved.  $\blacksquare$

Let us now focus on the estimation of small tail probabilities  $\bar{F}(y_n(x)|x)$  when  $y_n(x) \uparrow y_F(x)$  as  $n \rightarrow \infty$ . The following result provides sufficient conditions for the asymptotic normality of  $\hat{\bar{F}}_n(y_n(x)|x)$ .

**Proposition 1.** *Suppose (A.1), (A.2) and (A.3) hold and let  $x \in \mathbb{R}^p$  such that  $g(x) > 0$ . If  $nh^p \bar{F}(y_n(x)|x) \rightarrow \infty$  and  $nh^p \bar{F}(y_n(x)|x) (h \vee \omega_n(x))^2 \rightarrow 0$ , then, the random vector*

$$\left\{ \sqrt{nh^p \bar{F}(y_n(x)|x)} \left( \frac{\hat{\bar{F}}_n(y_{n,j}(x)|x)}{\bar{F}(y_{n,j}(x)|x)} - 1 \right) \right\}_{j=1, \dots, J}$$

is asymptotically Gaussian, centered, with covariance matrix  $\|K\|_2^2 / g(x) V$  where  $V_{j,j'} = \tau_{j \wedge j'}^{-1}$  for  $(j, j') \in \{1, \dots, J\}^2$ .

**Proof of Proposition 1.** Keeping in mind the notations of Lemma 4, the following expansion holds

$$\Lambda_n^{-1}(x) \sum_{j=1}^J \beta_j \left( \frac{\hat{\bar{F}}_n(y_{n,j}(x)|x)}{\bar{F}(y_{n,j}(x)|x)} - 1 \right) = \frac{T_{1,n} + T_{2,n} - T_{3,n}}{\hat{g}_n(x)}, \quad (12)$$

where

$$\begin{aligned} T_{1,n} &= g(x) \Lambda_n^{-1}(x) \sum_{j=1}^J \beta_j \left( \frac{\hat{\psi}_n(y_{n,j}(x), x) - \mathbb{E}(\hat{\psi}_n(y_{n,j}(x), x))}{\psi(y_{n,j}(x), x)} \right) \\ T_{2,n} &= g(x) \Lambda_n^{-1}(x) \sum_{j=1}^J \beta_j \left( \frac{\mathbb{E}(\hat{\psi}_n(y_{n,j}(x), x)) - \psi(y_{n,j}(x), x)}{\psi(y_{n,j}(x), x)} \right) \\ T_{3,n} &= \left( \sum_{j=1}^J \beta_j \right) \Lambda_n^{-1}(x) (\hat{g}_n(x) - g(x)). \end{aligned}$$

Let us highlight that assumptions  $nh^p\bar{F}(y_n(x)|x)\omega_n^2(x) \rightarrow 0$  and  $nh^p\bar{F}(y_n(x)|x) \rightarrow \infty$  imply that  $\omega_n(x) \rightarrow 0$  as  $n \rightarrow \infty$ . Thus, from Lemma 4(ii), the random term  $T_{1,n}$  can be rewritten as

$$T_{1,n} = g(x)\|K\|_2\sqrt{\beta^t V\beta}\xi_n, \quad (13)$$

where  $\xi_n$  converges to a standard Gaussian random variable. The nonrandom term  $T_{2,n}$  is controlled with Lemma 4(i):

$$T_{2,n} = O(\Lambda_n^{-1}(x)(h + \Delta(y_n(x), x))) = O\left((nh^p\bar{F}(y_n(x)|x))^{1/2}(h \vee \omega_n(x))\right) = o(1). \quad (14)$$

Finally,  $T_{3,n}$  is a classical term in kernel density estimation, which can be bounded by Lemma 3:

$$\begin{aligned} T_{3,n} &= O(h\Lambda_n^{-1}(x)) + O_P(\Lambda_n^{-1}(x)(nh^p)^{-1/2}) \\ &= O(nh^{p+2}\bar{F}(y_n(x)|x))^{1/2} + O_P(\bar{F}(y_n(x)|x))^{1/2} = o_P(1). \end{aligned} \quad (15)$$

Collecting (12)–(15), it follows that

$$\hat{g}_n(x)\Lambda_n^{-1}(x) \sum_{j=1}^J \beta_j \left( \frac{\hat{\bar{F}}_n(y_{n,j}(x)|x)}{\bar{F}(y_{n,j}(x)|x)} - 1 \right) = g(x)\|K\|_2\sqrt{\beta^t V\beta}\xi_n + o_P(1).$$

Finally,  $\hat{g}_n(x) \xrightarrow{P} g(x)$  yields

$$\sqrt{nh^p\bar{F}(y_n(x)|x)} \sum_{j=1}^J \beta_j \left( \frac{\hat{\bar{F}}_n(y_{n,j}(x)|x)}{\bar{F}(y_{n,j}(x)|x)} - 1 \right) = \|K\|_2\sqrt{\frac{\beta^t V\beta}{g(x)}}\xi_n + o_P(1)$$

and the result is proved.  $\blacksquare$

The last lemma establishes that  $K_{\hat{\gamma}_n(x)}(r_n)$  inherits from the convergence properties of  $\hat{\gamma}_n(x)$ .

**Lemma 5.** *Suppose  $\xi_n^{(\gamma)}(x) := \Lambda_n^{-1}(\hat{\gamma}_n(x) - \gamma(x)) = O_{\mathbb{P}}(1)$ , where  $\Lambda_n \rightarrow 0$ . Let  $r_n \geq 1$  or  $r_n \leq 1$  such that  $\Lambda_n \log(r_n) \rightarrow 0$ . Then,*

$$\Lambda_n^{-1} \left( \frac{K_{\hat{\gamma}_n(x)}(r_n) - K_{\gamma(x)}(r_n)}{K'_{\gamma(x)}(r_n)} \right) = \xi_n^{(\gamma)}(x)(1 + o_{\mathbb{P}}(1)).$$

**Proof.** Since  $\hat{\gamma}_n(x) \xrightarrow{P} \gamma(x)$ , a first order Taylor expansion yields

$$K_{\hat{\gamma}_n(x)}(r_n) = K_{\gamma(x)}(r_n) + \Lambda_n \xi_n^{(\gamma)}(x) K'_{\tilde{\gamma}_n(x)}(r_n),$$

where  $\tilde{\gamma}_n(x) = \gamma(x) + \Theta_n \Lambda_n \xi_n^{(\gamma)}(x)$  with  $\Theta_n \in (0, 1)$ . As a consequence

$$\begin{aligned} \Lambda_n^{-1} \left( \frac{K_{\hat{\gamma}_n(x)}(r_n) - K_{\gamma(x)}(r_n)}{K'_{\gamma(x)}(r_n)} \right) &= \xi_n^{(\gamma)}(x) \frac{K'_{\tilde{\gamma}_n(x)}(r_n)}{K'_{\gamma(x)}(r_n)} \\ &= \xi_n^{(\gamma)}(x) \left( 1 + \frac{\int_1^{r_n} (s^{\tilde{\gamma}_n(x) - \gamma(x)} - 1) s^{\gamma(x) - 1} \log(s) ds}{\int_1^{r_n} s^{\gamma(x) - 1} \log(s) ds} \right). \end{aligned}$$

Suppose for instance  $r_n \geq 1$ . The assumptions yield  $(\log r_n)(\tilde{\gamma}_n(x) - \gamma(x)) \xrightarrow{P} 0$  and thus, for  $n$  large enough, with high probability,

$$\sup_{s \in [1, r_n]} |(s^{\tilde{\gamma}_n(x) - \gamma(x)} - 1)| \leq 2(\log r_n)|\tilde{\gamma}_n(x) - \gamma(x)| = o_{\mathbb{P}}(1).$$

As a conclusion,

$$\Lambda_n^{-1} \left( \frac{K_{\hat{\gamma}_n(x)}(r_n) - K_{\gamma(x)}(r_n)}{K'_{\gamma(x)}(r_n)} \right) = \xi^{(\gamma)}(x)(1 + o_{\mathbb{P}}(1))$$

and the result is proved. The case  $r_n \leq 1$  is easily deduced since  $K_{\gamma(x)}(1/r_n) = -K_{-\gamma(x)}(r_n)$  and  $K'_{\gamma(x)}(1/r_n) = K'_{-\gamma(x)}(r_n)$ .  $\blacksquare$

## 6.2 Proofs of main results

**Proof of Theorem 1.** Let us introduce  $v_n = (nh^p\alpha_n^{-1})^{1/2}$ ,  $\sigma_n(x) = (v_n f(q(\alpha_n|x)|x))^{-1}$  and, for all  $j = 1, \dots, J$ ,

$$\begin{aligned} W_{n,j}(x) &= v_n \left( \hat{F}_n(q(\tau_j\alpha_n|x) + \sigma_n(x)z_j|x) - \bar{F}(q(\tau_j\alpha_n|x) + \sigma_n(x)z_j|x) \right) \\ a_{n,j}(x) &= v_n \left( \tau_j\alpha_n - \bar{F}(q(\alpha_n|x) + \sigma_n(x)z_j|x) \right) \end{aligned}$$

and  $z_j \in \mathbb{R}$ . We examine the asymptotic behavior of  $J$ -variate function defined by

$$\Phi_n(z_1, \dots, z_J) = \mathbb{P} \left( \bigcap_{j=1}^J \{ \sigma_n^{-1}(x)(\hat{q}_n(\tau_j\alpha_n|x) - q(\tau_j\alpha_n|x)) \leq z_j \} \right) = \mathbb{P} \left( \bigcap_{j=1}^J \{ W_{n,j}(x) \leq a_{n,j}(x) \} \right).$$

Let us first focus on the nonrandom term  $a_{n,j}(x)$ . For each  $j \in \{1, \dots, J\}$  there exists  $\theta_{n,j} \in (0, 1)$  such that

$$a_{n,j}(x) = v_n \sigma_n(x) z_j f(q_{n,j}(x)|x) = z_j \frac{f(q_{n,j}(x)|x)}{f(q(\alpha_n|x)|x)},$$

where

$$\begin{aligned} q_{n,j}(x) &= q(\tau_j\alpha_n|x) + \theta_{n,j}\sigma_n(x)z_j \\ &= q(\tau_j\alpha_n|x) + \theta_{n,j} \frac{z_j}{\tau_j} (nh^p\alpha_n)^{-1/2} \frac{\tau_j\alpha_n}{f(q(\tau_j\alpha_n|x)|x)} \frac{f(q(\tau_j\alpha_n|x)|x)}{f(q(\alpha_n|x)|x)} \\ &= q(\tau_j\alpha_n|x) + \theta_{n,j} z_j \tau_j^{\gamma(x)} (nh^p\alpha_n)^{-1/2} \frac{\tau_j\alpha_n}{f(q(\tau_j\alpha_n|x)|x)} (1 + o(1)), \end{aligned}$$

since  $y \mapsto f(q(y|x)|x)$  is regularly varying at 0 with index  $\gamma(x) + 1$ , see [19, Corollary 1.1.10, eq. 1.1.33]. Now, in view of [19, Theorem 1.2.6] and [19, Remark 1.2.7], a possible choice of the auxiliary function is

$$a(t|x) = \frac{\bar{F}(t|x)}{f(t|x)} (1 + o(1)), \quad (16)$$

leading to

$$q_{n,j}(x) = q(\tau_j\alpha_n|x) + \theta_{n,j} z_j \tau_j^{\gamma(x)} (nh^p\alpha_n)^{-1/2} a(q(\tau_j\alpha_n|x)|x) (1 + o(1)).$$

Applying Lemma 2 with  $z_n(x) = q(\tau_j\alpha_n|x)$ ,  $t_n(x) = \theta_{n,j} z_j \tau_j^{\gamma(x)} (nh^p\alpha_n)^{-1/2} (1 + o(1))$  and  $t_0(x) = 0$  yields

$$\frac{\bar{F}(q_{n,j}(x)|x)}{\tau_j\alpha_n} \rightarrow K_{\gamma(x)}^{-1}(0) = 1$$

as  $n \rightarrow \infty$ . Recalling that  $y \mapsto f(q(y|x)|x)$  is regularly varying, we have

$$\frac{f(q_{n,j}(x)|x)}{f(q(\alpha_n|x)|x)} \rightarrow \tau_j^{\gamma(x)+1}$$

as  $n \rightarrow \infty$  and therefore

$$a_{n,j}(x) = z_j \tau_j^{\gamma(x)+1} (1 + o(1)), \quad j = 1, \dots, J. \quad (17)$$

Let us now turn to the random term  $W_{n,j}(x)$ . Let us define  $z_{n,j}(x) = q(\tau_j\alpha_n|x) + \sigma_n(x)z_j$  for  $j = 1, \dots, J$ ,  $y_n(x) = q(\alpha_n|x)$ , and consider the expansion

$$\frac{z_{n,j}(x) - y_n(x)}{a(y_n(x)|x)} = \frac{q(\tau_j\alpha_n|x) - q(\alpha_n|x)}{a(q(\alpha_n|x)|x)} + \frac{\sigma_n(x)z_j}{a(q(\alpha_n|x)|x)}.$$

From (4), we have

$$\lim_{n \rightarrow \infty} \frac{q(\tau_j \alpha_n | x) - q(\alpha_n | x)}{a(q(\alpha_n | x) | x)} = K_{\gamma(x)}(1/\tau_j),$$

and

$$\lim_{n \rightarrow \infty} \frac{\sigma_n(x) z_j}{a(q(\alpha_n | x) | x)} = 0,$$

leading to  $z_{n,j}(x) = y_n(x) + K_{\gamma(x)}(1/\tau_j) a(y_n(x) | x) (1 + o(1))$ . Introducing  $\beta_{n,j}(x) = \bar{F}(y_{n,j}(x) | x)$ , the oscillation  $\omega_n(x)$  can be rewritten as

$$\omega_n(x) = \max_{j=1, \dots, J} \sup_{x' \in B(x, h)} \left| \frac{\bar{F}(q(\beta_{n,j}(x) | x) | x')}{\beta_{n,j}(x)} - 1 \right|.$$

For all  $\kappa \in (0, \tau_J)$  and  $j = 1, \dots, J$ , we eventually have  $z_{n,j}(x) \in [y_n(x), z_n(x)]$  where  $z_n(x) := y_n(x) + K_{\gamma(x)}(2/\kappa) a(y_n(x) | x)$  and thus  $\beta_{n,j}(x) \in [\bar{F}(z_n(x) | x), \alpha_n]$  eventually. Now, Lemma 2 implies that  $\bar{F}(z_n(x) | x) / \alpha_n \rightarrow \kappa/2$  as  $n \rightarrow \infty$  and thus, for  $n$  large enough,  $\beta_{n,j}(x) \in [\kappa \alpha_n, \alpha_n]$ . Consequently,  $\omega_n(x) \leq \Delta_\kappa(\alpha_n, x)$ . Applying Proposition 1 and Lemma 2 yields

$$W_{n,j}(x) = \frac{\bar{F}(z_{n,j}(x) | x)}{\alpha_n} \xi_{n,j} = \tau_j \xi_{n,j} (1 + o(1))$$

where  $\xi_n = (\xi_{n,1}, \dots, \xi_{n,J})^t$  converges to a centered Gaussian random vector with covariance matrix  $\|K\|_2^2 / g(x) V$ . Taking into account of (17), the results follows.  $\blacksquare$

**Proof of Corollary 1.** Let us remark that, from (16),

$$\begin{aligned} & \left\{ f(q(\alpha_n | x) | x) \sqrt{nh^p \alpha_n^{-1}} (\hat{q}_n(\tau_j \alpha_n | x) - q(\tau_j \alpha_n | x)) \right\}_{j=1, \dots, J} \\ &= \left\{ q(\alpha_n | x) \frac{f(q(\alpha_n | x) | x)}{\alpha_n} \frac{q(\tau_j \alpha_n | x)}{q(\alpha_n | x)} \sqrt{nh^p \alpha_n} \left( \frac{\hat{q}_n(\tau_j \alpha_n | x)}{q(\tau_j \alpha_n | x)} - 1 \right) \right\}_{j=1, \dots, J} \\ &= \left\{ \frac{q(\alpha_n | x)}{a(q(\alpha_n | x) | x)} \frac{q(\tau_j \alpha_n | x)}{q(\alpha_n | x)} \sqrt{nh^p \alpha_n} \left( \frac{\hat{q}_n(\tau_j \alpha_n | x)}{q(\tau_j \alpha_n | x)} - 1 \right) \right\}_{j=1, \dots, J} (1 + o(1)) \\ &= \left\{ \frac{q(\alpha_n | x)}{a(q(\alpha_n | x) | x)} \tau_j^{-\gamma(x) \vee 0} \sqrt{nh^p \alpha_n} \left( \frac{\hat{q}_n(\tau_j \alpha_n | x)}{q(\tau_j \alpha_n | x)} - 1 \right) \right\}_{j=1, \dots, J} (1 + o(1)), \end{aligned}$$

in view of Lemma 1. The result follows from Theorem 1.  $\blacksquare$

**Proof of Theorem 2.** By definition,

$$q_n(\beta_n | x) = q_n(\alpha_n | x) + (K_{\gamma(x)}(\alpha_n / \beta_n) + b(\beta_n / \alpha_n, \alpha_n)) a(q(\alpha_n | x) | x)$$

and thus, the following expansion can be easily established:

$$\begin{aligned} \Lambda_n^{-1} \left( \frac{\tilde{q}_n(\beta_n | x) - q(\beta_n | x)}{a(q(\alpha_n | x) | x) K'_{\gamma(x)}(\alpha_n / \beta_n)} \right) &= \Lambda_n^{-1} \left( \frac{\hat{q}_n(\alpha_n | x) - q(\alpha_n | x)}{a(q(\alpha_n | x) | x) K'_{\gamma(x)}(\alpha_n / \beta_n)} \right) \\ &+ \Lambda_n^{-1} \left( \frac{K_{\gamma(x)}(\alpha_n / \beta_n) - K'_{\gamma(x)}(\alpha_n / \beta_n)}{K'_{\gamma(x)}(\alpha_n / \beta_n)} \right) \frac{\hat{a}_n(x)}{a(q(\alpha_n | x) | x)} \\ &+ \Lambda_n^{-1} \frac{K_{\gamma(x)}(\alpha_n / \beta_n)}{K'_{\gamma(x)}(\alpha_n / \beta_n)} \left( \frac{\hat{a}_n(x)}{a(q(\alpha_n | x) | x)} - 1 \right) \\ &+ \Lambda_n^{-1} \frac{b(\beta_n / \alpha_n, \alpha_n)}{K'_{\gamma(x)}(\alpha_n / \beta_n)} \\ &=: T_{n,1} + T_{n,2} + T_{n,3} + T_{n,4}. \end{aligned}$$

Introducing

$$(\xi_n^{(\gamma)}(x), \xi_n^{(a)}(x), \xi_n^{(q)}(x)) := \Lambda_n^{-1} \left( \hat{\gamma}_n(x) - \gamma(x), \frac{\hat{a}_n(x)}{a(q(\alpha_n|x)|x)} - 1, \frac{\hat{q}_n(\alpha_n|x) - q(\alpha_n|x)}{a(q(\alpha_n|x)|x)} \right),$$

from and remarking that, when  $u \rightarrow \infty$ ,

$$K'_z(u) = (1 + o(1)) \begin{cases} \frac{1}{z^2} & \text{if } z < 0, \\ \frac{\log^2(u)}{2} & \text{if } z = 0, \\ \frac{u^z \log(u)}{z} & \text{if } z > 0, \end{cases} \quad (18)$$

the first term can be rewritten as

$$T_{n,1} = \frac{\xi_n^{(q)}(x)}{K'_{\gamma(x)}(\alpha_n/\beta_n)} = (\gamma(x) \wedge 0)^2 \xi_n^{(q)}(x) (1 + o_{\mathbb{P}}(1)). \quad (19)$$

Second,  $\Lambda_n \rightarrow 0$  and (7) entail  $\hat{a}_n(x)/a(q(\alpha_n|x)|x) \xrightarrow{P} 1$  and thus

$$T_{n,2} = \Lambda_n^{-1} \left( \frac{K_{\hat{\gamma}_n(x)}(\alpha_n/\beta_n) - K_{\gamma(x)}(\alpha_n/\beta_n)}{K'_{\gamma(x)}(\alpha_n/\beta_n)} \right) (1 + o_{\mathbb{P}}(1)) = \xi_n^{(\gamma)}(x) (1 + o_{\mathbb{P}}(1)), \quad (20)$$

from Lemma 5. From (7), (18), and in view of

$$K_z(u) = (1 + o(1)) \begin{cases} -\frac{1}{z} & \text{if } z < 0, \\ \log(u) & \text{if } z = 0, \\ \frac{u^z}{z} & \text{if } z > 0. \end{cases}$$

the third term can be rewritten as

$$T_{n,3} = \xi_n^{(a)}(x) \frac{K_{\gamma(x)}(\alpha_n/\beta_n)}{K'_{\gamma(x)}(\alpha_n/\beta_n)} = -(\gamma(x) \wedge 0) \xi_n^{(a)}(x) (1 + o_{\mathbb{P}}(1)). \quad (21)$$

Finally,  $T_{n,4} = o_{\mathbb{P}}(1)$  by assumption and the conclusion follows from (7), (19), (20) and (21).  $\blacksquare$

**Proof of Theorem 3.** The proof consists in deriving asymptotic expansions for the three considered random variables. i) Let us first introduce

$$\gamma_{n,j}(x) = \frac{1}{\log r} \log \left( \frac{\hat{q}_n(\tau_j \alpha_n|x) - \hat{q}_n(\tau_{j+1} \alpha_n|x)}{\hat{q}_n(\tau_{j+1} \alpha_n|x) - \hat{q}_n(\tau_{j+2} \alpha_n|x)} \right) \quad (22)$$

such that  $\hat{\gamma}_n^{\text{RP}}(x) = \sum_{j=1}^{J-2} \pi_j \gamma_{n,j}(x)$ . From Theorem 1 and in view of (4), we have, for all  $j = 1, \dots, J$ ,

$$\hat{q}_n(\tau_j \alpha_n|x) = q(\alpha_n|x) + a(q(\alpha_n|x)|x) (K_{\gamma(x)}(1/\tau_j) + b(\tau_j, \alpha_n|x)) + \sigma_n(x) \xi_{j,n},$$

with  $\sigma_n^{-1}(x) = f(q(\alpha_n|x)|x) \sqrt{nh^p \alpha_n^{-1}}$  and where the random vector  $\xi_n = (\xi_{j,n})_{j=1, \dots, J}$  is asymptotically Gaussian, centered, with covariance matrix  $\|K\|_2^2/g(x) \Sigma(x)$ . Introducing

$$\begin{aligned} \eta_n(x) &:= \max_{j=1, \dots, J} |b(\tau_j, \alpha_n|x)| \\ \varepsilon_n &:= \sigma_n(x)/a(q(\alpha_n|x)|x) = (nh^p \alpha_n)^{-1/2} (1 + o(1)), \end{aligned}$$

see [19], it follows that

$$\begin{aligned}
\frac{\hat{q}_n(\tau_j \alpha_n | x) - \hat{q}_n(\tau_{j+1} \alpha_n | x)}{a(q(\alpha_n | x) | x)} &= \varepsilon_n(\xi_{j,n} - \xi_{j+1,n}) \\
&+ K_{\gamma(x)}(1/\tau_j) - K_{\gamma(x)}(1/\tau_{j+1}) + b(\tau_j, \alpha_n | x) - b(\tau_{j+1}, \alpha_n | x) \\
&= \varepsilon_n(\xi_{j,n} - \xi_{j+1,n}) + K_{\gamma(x)}(r)r^{-\gamma(x)j} + O(\eta_n(x)). \tag{23}
\end{aligned}$$

Replacing in (22), we obtain

$$(\log r)\gamma_{n,j}(x) = \log \left( \frac{\varepsilon_n(\xi_{j,n} - \xi_{j+1,n}) + K_{\gamma(x)}(r)r^{-\gamma(x)j} + O(\eta_n(x))}{\varepsilon_n(\xi_{j+1,n} - \xi_{j+2,n}) + K_{\gamma(x)}(r)r^{-\gamma(x)(j+1)} + O(\eta_n(x))} \right),$$

or equivalently,

$$\begin{aligned}
(\log r)(\gamma_{n,j}(x) - \gamma(x)) &= \log \left( 1 + \frac{\varepsilon_n(\xi_{j,n} - \xi_{j+1,n})r^{\gamma(x)j}}{K_{\gamma(x)}(r)} + O(\eta_n(x)) \right) \\
&- \log \left( 1 + \frac{\varepsilon_n(\xi_{j+1,n} - \xi_{j+2,n})r^{\gamma(x)(j+1)}}{K_{\gamma(x)}(r)} + O(\eta_n(x)) \right)
\end{aligned}$$

A first order Taylor expansion yields

$$(\log r)\varepsilon_n^{-1}(\gamma_{n,j}(x) - \gamma(x)) = \frac{r^{\gamma(x)j}}{K_{\gamma(x)}(r)} \left( \xi_{j,n} - (1 + r^{\gamma(x)})\xi_{j+1,n} + r^{\gamma(x)}\xi_{j+2,n} \right) + O(\varepsilon_n^{-1}\eta_n(x)) + o_{\mathbb{P}}(1)$$

and thus, under the assumption  $(nh^p\alpha_n)^{1/2}\eta_n(x) \rightarrow 0$  as  $n \rightarrow \infty$ ,

$$\sqrt{nh^p\alpha_n}(\hat{\gamma}_n^{\text{RP}}(x) - \gamma(x)) = \frac{1}{(\log r)K_{\gamma(x)}(r)} \sum_{j=1}^{J-2} \pi_j r^{\gamma(x)j} \left( \xi_{j,n} - (1 + r^{\gamma(x)})\xi_{j+1,n} + r^{\gamma(x)}\xi_{j+2,n} \right) + o_{\mathbb{P}}(1).$$

Defining for the sake of simplicity  $\pi_{-1} = \pi_0 = \pi_{J-1} = \pi_J = 0$ ,  $\beta_0^{(\gamma)} = \frac{1}{\log r}$ ,  $\beta_1^{(\gamma)} = -\frac{1+r^{-\gamma(x)}}{\log(r)}$  and  $\beta_2^{(\gamma)} = \frac{r^{-\gamma(x)}}{\log(r)}$ , we end up with

$$\begin{aligned}
\xi_n^{(\gamma)}(x) &:= \sqrt{nh^p\alpha_n}(\hat{\gamma}_n^{\text{RP}}(x) - \gamma(x)) \\
&= \frac{1}{K_{\gamma(x)}(r)} \sum_{j=1}^J r^{\gamma(x)j} \left( \beta_0^{(\gamma)}\pi_j + \beta_1^{(\gamma)}\pi_{j-1} + \beta_2^{(\gamma)}\pi_{j-2} \right) \xi_{j,n} + o_{\mathbb{P}}(1). \tag{24}
\end{aligned}$$

ii) Second, let us now consider

$$a_{n,j}(x) = \frac{r^{\hat{\gamma}_n^{\text{RP}}(x)j}(\hat{q}_n(\tau_j \alpha_n | x) - \hat{q}_n(\tau_{j+1} \alpha_n | x))}{K_{\hat{\gamma}_n^{\text{RP}}(x)}(r)}$$

such that  $\hat{a}_n(x) = \sum_{j=1}^{J-2} \pi_j a_{n,j}(x)$ . From (23), it follows that, for all  $j = 1, \dots, J$ ,

$$\frac{a_{n,j}(x)}{a(q(\alpha_n | x) | x)} = \frac{r^{\hat{\gamma}_n^{\text{RP}}(x)j}}{K_{\hat{\gamma}_n^{\text{RP}}(x)}(r)} \left( \varepsilon_n(\xi_{j,n} - \xi_{j+1,n}) + K_{\gamma(x)}(r)r^{-\gamma(x)j} + O(\eta_n(x)) \right).$$

Remarking that  $\hat{\gamma}_n^{\text{RP}}(x) = \gamma(x) + (nh^p\alpha_n)^{-1/2}\xi_n^{(\gamma)}(x)$ , Lemma 5 yields

$$\frac{a_{n,j}(x)}{a(q(\alpha_n | x) | x)} = \frac{1 + \frac{r^{\gamma(x)j}}{K_{\gamma(x)}(r)}\varepsilon_n(\xi_{j,n} - \xi_{j+1,n}) + O(\eta_n(x))}{1 + \frac{K'_{\gamma(x)}(r)}{K_{\gamma(x)}(r)}(nh^p\alpha_n)^{-1/2}\xi_n^{(\gamma)}(x)(1 + o_{\mathbb{P}}(1))} \exp \left( \xi_n^{(\gamma)}(x)j \log(r)(nh^p\alpha_n)^{-1/2} \right).$$



A first order Taylor expansion thus gives

$$\begin{aligned} \sqrt{nh^p\alpha_n} \left( \frac{a_{n,j}(x)}{a(q(\alpha_n|x)|x)} - 1 \right) &= \xi_n^{(\gamma)}(x) \left( j \log(r) - \frac{K'_{\gamma(x)}(r)}{K_{\gamma(x)}(r)} \right) + \frac{r^{\gamma(x)j}}{K_{\gamma(x)}(r)} (\xi_{j,n} - \xi_{j+1,n}) \\ &+ O\left(\sqrt{nh^p\alpha_n}\eta_n(x)\right) + o_{\mathbb{P}}(1). \end{aligned}$$

Recalling that  $\pi_{-1} = \pi_0 = \pi_{J-1} = \pi_J = 0$  and introducing

$$\begin{aligned} \mathbb{E}(\pi) &= \sum_{j=1}^J j\pi_j, & \beta_1^{(a)} &= -r^{-\gamma(x)} - (r^{-\gamma(x)} + 1) \left( \mathbb{E}(\pi) - \frac{K'_{\gamma(x)}(r)}{\log(r)K_{\gamma(x)}(r)} \right), \\ \beta_0^{(a)} &= 1 + \mathbb{E}(\pi) - \frac{K'_{\gamma(x)}(r)}{\log(r)K_{\gamma(x)}(r)}, & \beta_2^{(a)} &= r^{-\gamma(x)} \left( \mathbb{E}(\pi) - \frac{K'_{\gamma(x)}(r)}{\log(r)K_{\gamma(x)}(r)} \right) \end{aligned}$$

it follows that

$$\begin{aligned} \xi_n^{(a)}(x) &:= \sqrt{nh^p\alpha_n} \left( \frac{a_n(x)}{a(q(\alpha_n|x)|x)} - 1 \right) \\ &= \left( \mathbb{E}(\pi) \log(r) - \frac{K'_{\gamma(x)}(r)}{K_{\gamma(x)}(r)} \right) \xi_n^{(\gamma)}(x) + \frac{1}{K_{\gamma(x)}(r)} \sum_{j=1}^J r^{\gamma(x)j} (\pi_j - r^{-\gamma(x)}\pi_{j-1}) \xi_{j,n} + o_{\mathbb{P}}(1) \\ &= \frac{1}{K_{\gamma(x)}(r)} \sum_{j=1}^J r^{\gamma(x)j} \left( \beta_0^{(a)}\pi_j + \beta_1^{(a)}\pi_{j-1} + \beta_2^{(a)}\pi_{j-2} \right) \xi_{j,n} + o_{\mathbb{P}}(1) \end{aligned} \quad (25)$$

in view of (24). iii) Third, Corollary 1 states that

$$\xi_{1,n} = \frac{\sqrt{nh^p\alpha_n}}{a(q(\alpha_n|x)|x)} (\hat{q}_n(\alpha_n|x) - q(\alpha_n|x)) \quad (26)$$

is asymptotically Gaussian. Finally, collecting (24), (25) and (26),

$$(\xi_n^{(\gamma)}(x), \xi_n^{(a)}(x), \xi_{1,n})^t = \frac{1}{K_{\gamma(x)}(r)} A(x) \xi_n + o_{\mathbb{P}}(1),$$

where  $A(x)$  is the  $3 \times J$  matrix defined by

$$\begin{aligned} A_{1,j}(x) &= r^{\gamma(x)j} \left( \beta_0^{(\gamma)}\pi_j + \beta_1^{(\gamma)}\pi_{j-1} + \beta_2^{(\gamma)}\pi_{j-2} \right) \\ A_{2,j}(x) &= r^{\gamma(x)j} \left( \beta_0^{(a)}\pi_j + \beta_1^{(a)}\pi_{j-1} + \beta_2^{(a)}\pi_{j-2} \right) \\ A_{3,j}(x) &= K_{\gamma(x)}(r) \mathbb{I}\{j = 1\}, \end{aligned}$$

for all  $j = 1, \dots, J$ . It is thus clear that the random vector  $(\xi_n^{(\gamma)}(x), \xi_n^{(a)}(x), \xi_{1,n})^t$  converges in distribution to a centered Gaussian random vector with covariance matrix

$$\frac{\|K\|_2^2}{g(x)K_{\gamma(x)}^2(r)} A(x) \Sigma(x) A^t(x) =: S(x). \quad (27)$$

■

**Acknowledgements.** The authors thank the Editor and both anonymous Associate Editor and reviewer for their valuable feedback on this paper.

## References

- [1] Beirlant, J., and Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis*, 89, 97–118.
- [2] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels J. (2004). *Statistics of extremes: Theory and applications*, Wiley.
- [3] Berlinet, A., Gannoun, A. and Matzner-Lober, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles. *Statistics*, 35, 139–169.
- [4] 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.
- [5] Chernozhukov, V. (2005). Extremal quantile regression, *The Annals of Statistics*, 33(2), 806–839.
- [6] Daouia, A., Florens, J-P. and Simar, L. (2010). Frontier estimation and extreme value theory. *Bernoulli*, 16, 1039–1063.
- [7] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20, 311–333.
- [8] 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.
- [9] 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.
- [10] Drees, H. (1995). Refined Pickands estimators of the extreme value index. *The Annals of Statistics*, 23, 2059–2080.
- [11] Feigin, P.D., and Resnick, S.I. (1997). Linear programming estimators and bootstrapping for heavy tailed phenomena. *Advances in Applied Probability*, 29, 759–805.
- [12] Gannoun, A. (1990). Estimation non paramétrique de la médiane conditionnelle, médianogramme et méthode du noyau. *Publications de l'Institut de Statistique de l'Université de Paris*, XXXXVI, 11–22.
- [13] Gardes, L., and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99, 2368–2388.
- [14] Gardes, L., and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13, 177–204.
- [15] Gardes, L., Girard, S., and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101, 419–433.
- [16] Girard, S., and Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data. *Journal of Multivariate Analysis*, 99, 403–420.
- [17] Girard, S., and Menneteau, L. (2005). Central limit theorems for smoothed extreme value estimates of Poisson point processes boundaries. *Journal of Statistical Planning and Inference*, 135, 433–460.
- [18] Greene, W.H. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46, 141–163.

- [19] de Haan, L., and Ferreira, A. *Extreme value theory: An introduction*, Springer Series in Operations Research and Financial Engineering, Springer, 2006.
- [20] Hall, P., and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 15, 153–167.
- [21] Hardle, W., and Stoker, T. (1999). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84, 986–995.
- [22] Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, 87, 58–68.
- [23] Jurecková, J. (2007). Remark on extreme regression quantile. *Sankhya*, 69, Part 1, 87–100.
- [24] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of American Statistical Association*, 96, 458–468.
- [25] Park, B.U. (2001). On nonparametric estimation of data edges. *Journal of the Korean Statistical Society*, 30, 265–280.
- [26] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33, 1065–1076.
- [27] Ruppert, D., Wand, M., and R.J. Carroll (2003). *Semiparametric Regression*. Cambridge U. Press.
- [28] Samanta, T. (1989). Non-parametric estimation of conditional quantiles. *Statistics and Probability Letters*, 7, 407–412.
- [29] 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.
- [30] Stone, C.J. (1977). Consistent nonparametric regression (with discussion). *The Annals of Statistics*, 5, 595–645.
- [31] Stute, W. (1986). Conditional empirical processes. *The Annals of Statistics*, 14, 638–647.
- [32] Tsay, R.S. (2002). *Analysis of financial time series*. Wiley, New York.
- [33] Yu, K. and M.C. Jones (1998). Local linear quantile regression. *Journal of American Statistical Association*, 93, 228–237.