REVSTAT – Statistical Journal
Volume 17, Number 1, January 2019, 109–144

AN INTEGRATED FUNCTIONAL WEISSMAN ESTIMATOR FOR CONDITIONAL EXTREME QUANTILES

Authors: Laurent Gardes
– University of Strasbourg, CNRS, IRMA 7501,
F-67000 Strasbourg, France
gardes@unistra.fr

Gilles Stupfler
– School of Mathematical Sciences, The University of Nottingham,
University Park, Nottingham NG7 2RD, United Kingdom
Gilles.Stupfler@nottingham.ac.uk

Received: September 2016 Revised: January 2017 Accepted: February 2017

Abstract:

- It is well-known that estimating extreme quantiles, namely, quantiles lying beyond the range of the available data, is a nontrivial problem that involves the analysis of tail behavior through the estimation of the extreme-value index. For heavy-tailed distributions, on which this paper focuses, the extreme-value index is often called the tail index and extreme quantile estimation typically involves an extrapolation procedure. Besides, in various applications, the random variable of interest can be linked to a random covariate. In such a situation, extreme quantiles and the tail index are functions of the covariate and are referred to as conditional extreme quantiles and the conditional tail index, respectively. The goal of this paper is to provide classes of estimators of these quantities when there is a functional (*i.e.* possibly infinite-dimensional) covariate. Our estimators are obtained by combining regression techniques with a generalization of a classical extrapolation formula. We analyze the asymptotic properties of these estimators, and we illustrate the finite-sample performance of our conditional extreme quantile estimator on a simulation study and on a real chemometric data set.

Key-Words:

- *heavy-tailed distribution; functional random covariate; extreme quantile; tail index; asymptotic normality.*

AMS Subject Classification:

- 62G05, 62G20, 62G30, 62G32.

1. INTRODUCTION

Studying extreme events is relevant in numerous fields of statistical applications. In hydrology for example, it is of interest to estimate the maximum level reached by seawater along a coast over a given period, or to study extreme rainfall at a given location; in actuarial science, a major problem for an insurance firm is to estimate the probability that a claim so large that it represents a threat to its solvency is filed. When analyzing the extremes of a random variable, a central issue is that the straightforward empirical estimator of the quantile function is not consistent at extreme levels; in other words, direct estimation of a quantile exceeding the range covered by the available data is impossible, and this is of course an obstacle to meaningful estimation results in practice.

In many of the aforementioned applications, the problem can be accurately modeled using univariate heavy-tailed distributions, thus providing an extrapolation method to estimate extreme quantiles. Roughly speaking, a distribution is said to be heavy-tailed if and only if its related survival function decays like a power function with negative exponent at infinity; its so-called tail index γ is then the parameter which controls its rate of convergence to 0 at infinity. If Q denotes the underlying quantile function, this translates into: $Q(\delta) \approx [(1-\beta)/(1-\delta)]^{\gamma} Q(\beta)$ when β and δ are close to 1. The quantile function at an arbitrarily high extreme level can then be consistently deduced from its value at a typically much smaller level provided γ can be consistently estimated. This procedure, suggested by Weissman [42], is one of the simplest and most popular devices as far as extreme quantile estimation is concerned.

The estimation of the tail index γ, an excellent overview of which is given in the recent monographs by Beirlant *et al.* [2] and de Haan and Ferreira [27], is therefore a crucial step to gain understanding of the extremes of a random variable whose distribution is heavy-tailed. In practical applications, the variable of interest Y can often be linked to a covariate X. For instance, the value of rainfall at a given location depends on its geographical coordinates; in actuarial science, the claim size depends on the sum insured by the policy. In this situation, the tail index and quantiles of the random variable Y given $X = x$ are functions of x to which we shall refer as the conditional tail index and conditional quantile functions. Their estimation has been considered first in the "fixed design" case, namely when the covariates are nonrandom. Smith [36] and Davison and Smith [12] considered a regression model while Hall and Tajvidi [28] used a semi-parametric approach to estimate the conditional tail index. Fully nonparametric methods have been developed using splines (see Chavez-Demoulin and Davison [6]), local polynomials (see Davison and Ramesh [11]), a moving window approach (see Gardes and Girard [19]) and a nearest neighbor approach (see Gardes and Girard [20]), among others.

Despite the great interest in practice, the study of the random covariate case has been initiated only recently. We refer to the works of Wang and Tsai [41], based on a maximum likelihood approach, Daouia *et al.* [9] who used a fixed number of non parametric conditional quantile estimators to estimate the conditional tail index, later generalized in Daouia *et al.* [10] to a regression context with conditional response distributions belonging to the general max-domain of attraction, Gardes and Girard [21] who introduced a local generalized Pickands-type estimator (see Pickands [33]), Goegebeur *et al.* [25], who studied a non-parametric regression estimator whose strong uniform properties are examined in Goegebeur *et al.* [26]. Some generalizations of the popular moment estimator of Dekkers *et al.* [13] have been proposed by Gardes [18], Goegebeur *et al.* [23, 24] and Stupfler [37, 38]. In an attempt to obtain an estimator behaving better in finite-sample situations, Gardes and Stupfler [22] worked on a smoothed local Hill estimator (see Hill [29]) related to the work of Resnick and Stărică [34]. A different approach, that has been successful in recent years, is to combine extreme value theory and quantile regression: the pioneering paper is Chernozhukov [7], and we also refer to the subsequent papers by Chernozhukov and Du [8], Wang *et al.* [39] and Wang and Li [40].

The goal of this paper is to introduce integrated estimators of conditional extreme quantiles and of the conditional tail index for random, possibly infinite-dimensional, covariates. In particular, our estimator of the conditional tail index, based on the integration of a conditional log-quantile estimator, is somewhat related to the one of Gardes and Girard [19]. Our aim is to examine the asymptotic properties of our estimators, as well as to examine the applicability of our conditional extreme quantile estimator on numerical examples and on real data. Our paper is organized as follows: we define our estimators in Section 2. Their asymptotic properties are stated in Section 3. A simulation study is provided in Section 4 and we revisit a set of real chemometric data in Section 5. All the auxiliary results and proofs are deferred to the Appendix.

2. FUNCTIONAL EXTREME QUANTILE: DEFINITION AND ESTIMATION

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be n independent copies of a random pair (X, Y) taking its values in $\mathcal{E} \times \mathbb{R}_+$ where $(\mathcal{E}, d)$ is a (not necessarily finite-dimensional) Polish space endowed with a semi-metric d. For instance, $\mathcal{E}$ can be the standard p-dimensional space $\mathbb{R}^p$, a space of continuous functions over a compact metric space, or a Lebesgue space $L^p(\mathbb{R})$, to name a few. For $y > 0$, we denote by $S(y|X)$ a regular version of the conditional probability $\mathbb{P}(Y > y|X)$. Note that since $\mathcal{E}$ is a Polish space, such conditional probabilities always exist, see Jiřina [30].

In this paper, we focus on the situation where the conditional distribution of Y given X is heavy-tailed. More precisely, we assume that there exists a positive function $\gamma(\cdot)$, called the conditional tail index, such that

$$\lim_{y\to\infty} \frac{S(\lambda y|x)}{S(y|x)} = \lambda^{-1/\gamma(x)}, \tag{2.1}$$

for all $x \in \mathcal{E}$ and all $\lambda > 0$. This is the adaptation of the standard extreme-value framework of heavy-tailed distributions to the case when there is a covariate. The conditional quantile function of Y given $X = x$ is then defined for $x \in \mathcal{E}$ by $Q(\alpha|x) := \inf\{y > 0 \,|\, S(y|x) \leq \alpha\}$. If $x \in \mathcal{E}$ is fixed, our final aim is to estimate the conditional extreme quantile $Q(\beta_n|x)$ of order $\beta_n \to 0$. As we will show below, this does in fact require estimating the conditional tail index $\gamma(x)$ first.

2.1. Estimation of a functional extreme quantile

Recall that we are interested in the estimation of $Q(\beta_n|x)$ when $\beta_n \to 0$ as the sample size increases. The natural empirical estimator of this quantity is given by

$$\widehat{Q}_n(\beta_n|x) := \inf\left\{y > 0 \,|\, \widehat{S}_n(y|x) \leq \beta_n\right\}, \tag{2.2}$$

where

$$\widehat{S}_n(y|x) = \sum_{i=1}^n \mathbb{I}\{Y_i > y\}\mathbb{I}\{d(x, X_i) \leq h\} \Bigg/ \sum_{i=1}^n \mathbb{I}\{d(x, X_i) \leq h\}$$

and where $h = h(n)$ is a nonrandom sequence converging to 0 as $n \to \infty$. Unfortunately, denoting by $m_x(h) := n\mathbb{P}(d(x, X) \leq h)$ the average number of observations whose covariates belong to the ball $B(x, h) = \{x' \in \mathcal{E} \,|\, d(x, x') \leq h\}$ with center x and radius h, it can be shown (see Proposition 6.1) that the condition $m_x(h)\beta_n \to \infty$ is required to obtain the consistency of $\widehat{Q}_n(\beta_n|x)$. This means that at the same time, sufficiently many observations should belong to the ball $B(x, h)$ and β_n should be so small that the quantile $Q(\beta_n|x)$ is covered by the range of this data, and therefore the order β_n of the functional extreme quantile cannot be chosen as small as we would like. We thus need to propose another estimator adapted to this case. To this end, we start by remarking (see Bingham *et al.* [4, Theorem 1.5.12]) that (2.1) is equivalent to

$$\lim_{\alpha\to 0} \frac{Q(\lambda\alpha|x)}{Q(\alpha|x)} = \lambda^{-\gamma(x)}, \tag{2.3}$$

for all $\lambda > 0$. Hence, for $0 < \beta < \alpha$ with α small enough, we obtain the extrapolation formula $Q(\beta|x) \approx Q(\alpha|x)(\alpha/\beta)^{\gamma(x)}$ which is at the heart of Weissman's extrapolation method [42]. In order to borrow more strength from the available information in the sample, we note that, if μ is a probability measure on the interval $[0,1]$, another similar, heuristic approximation holds:

$$Q(\beta|x) \approx \int_{[0,1]} Q(\alpha|x) \left(\frac{\alpha}{\beta}\right)^{\gamma(x)} \mu(d\alpha).$$

If we have at our disposal a consistent estimator $\widehat{\gamma}_n(x)$ of $\gamma(x)$ (an example of such an estimator is given in Section 2.2), an idea is to estimate $Q(\beta_n|x)$ by:

$$\check{Q}_n(\beta_n|x) = \int_{[0,1]} \widehat{Q}_n(\alpha|x) \left(\frac{\alpha}{\beta_n}\right)^{\widehat{\gamma}_n(x)} \mu(d\alpha). \tag{2.4}$$

In order to obtain a consistent estimator of the extreme conditional quantile, the support of the measure μ, denoted by $\operatorname{supp}(\mu)$, should be located around 0. To be more specific, we assume in what follows that $\operatorname{supp}(\mu) \subset [\tau u, u]$ for some $\tau \in (0,1)$ and $u \in (0,1)$ small enough. For instance, taking μ to be the Dirac measure at u leads to $\check{Q}_n(\beta_n|x) = \widehat{Q}_n(u|x)\,(u/\beta_n)^{\widehat{\gamma}_n(x)}$, which is a straightforward adaptation to our conditional setting of the classical Weissman estimator [42]. If on the contrary μ is absolutely continuous, estimator (2.4) is a properly integrated and weighted version of Weissman's estimator. Due to the fact that it takes more of the available data into account, we can expect such an estimator to perform better than the simple adaptation of Weissman's estimator, a claim we investigate in our finite-sample study in Section 4.

2.2. Estimation of the functional tail index

To provide an estimator of the functional tail index $\gamma(x)$, we note that equation (2.3) warrants the approximation $\gamma(x) \approx \log[Q(\alpha|x)/Q(u|x)]/\log(u/\alpha)$ for $0 < \alpha < u$ when u is small enough. Let $\Psi(\cdot,u)$ be a measurable function defined on $(0,u)$ such that $0 < \left|\int_0^u \log(u/\alpha)\Psi(\alpha,u)d\alpha\right| < \infty$. Multiplying the aforementioned approximation by $\Psi(\cdot,u)$, integrating between 0 and 1 and replacing $Q(\cdot|x)$ by the classical estimator $\widehat{Q}_n(\cdot|x)$ defined in (2.2) leads to the estimator:

$$\widehat{\gamma}_n(x,u) := \int_0^u \Psi(\alpha,u) \log\frac{\widehat{Q}_n(\alpha|x)}{\widehat{Q}_n(u|x)} d\alpha \Bigg/ \int_0^u \log(u/\alpha)\Psi(\alpha,u)d\alpha. \tag{2.5}$$

Without loss of generality, we shall assume in what follows that

$$\int_0^u \log(u/\alpha)\Psi(\alpha,u)d\alpha = 1.$$

Particular choices of the function $\Psi(\cdot,u)$ actually yield generalizations of some well-known tail index estimators to the conditional framework. Let $k_x := uM_x(h)$, where $M_x(h)$ is the total number of covariates whose distance to x is not greater than h:

$$M_x(h) = \sum_{i=1}^n \mathbb{I}\{d(x,X_i) \le h\}.$$

The choice $\Psi(\cdot, u) = 1/u$ leads to the estimator:

$$\widehat{\gamma}_n^H(x) = \frac{1}{k_x} \sum_{i=1}^{\lfloor k_x \rfloor} \log \frac{\widehat{Q}_n((i-1)/M_x(h)|x)}{\widehat{Q}_n(k_x/M_x(h)|x)}, \tag{2.6}$$

which is the straightforward conditional adaptation of the classical Hill estimator (see Hill [29]). Now, taking $\Psi(\cdot, u) = u^{-1}(\log(u/\cdot) - 1)$ leads, after some algebra, to the estimator:

$$\widehat{\gamma}_n^Z(x) = \frac{1}{k_x} \sum_{i=1}^{\lfloor k_x \rfloor} i \log\left(\frac{k_x}{i}\right) \log \frac{\widehat{Q}_n((i-1)/M_x(h)|x)}{\widehat{Q}_n(i/M_x(h)|x)}.$$

This estimator can be seen as a generalization of the Zipf estimator (see Kratz and Resnick [31], Schultze and Steinebach [35]).

3. MAIN RESULTS

Our aim is now to establish asymptotic results for our estimators. We assume in all what follows that $Q(\cdot|x)$ is continuous and decreasing. Particular consequences of this condition include that $S(Q(\alpha|x)|x) = \alpha$ for any $\alpha \in (0,1)$ and that given $X = x$, Y has an absolutely continuous distribution with probability density function $f(\cdot|x)$.

Recall that under (2.1), or equivalently (2.3), the conditional quantile function may be written for all $t > 1$ as follows:

$$Q(t^{-1}|x) = c(t|x) \exp\left(\int_1^t \frac{\Delta(v|x) - \gamma(x)}{v} dv\right),$$

where $c(\cdot|x)$ is a positive function converging to a positive constant at infinity and $\Delta(\cdot|x)$ is a measurable function converging to 0 at infinity, see Bingham *et al.* [4, Theorem 1.3.1]. We assume in what follows that

(H_{SO}) $c(\cdot|x)$ is a constant function equal to $c(x) > 0$, the function $\Delta(\cdot|x)$ has ultimately constant sign at infinity and there exists $\rho(x) < 0$ such that for all $\lambda > 0$,

$$\lim_{y \to \infty} \left| \frac{\Delta(\lambda y|x)}{\Delta(y|x)} \right| = \lambda^{\rho(x)}.$$

The constant $\rho(x)$ is called the conditional second-order parameter of the distribution. These conditions on the function $\Delta(\cdot|x)$ are commonly used when studying tail index estimators and make it possible to control the error term in convergence (2.3). In particular, it is straightforward to see that for all $z > 0$,

$$\lim_{t \to \infty} \frac{1}{\Delta(t|x)} \left(\frac{Q((tz)^{-1}|x)}{Q(t^{-1}|x)} - z^{\gamma(x)}\right) = z^{\gamma(x)} \frac{z^{\rho(x)} - 1}{\rho(x)}, \tag{3.1}$$

which is the conditional analogue of the second-order condition of de Haan and Ferreira [27] for heavy-tailed distributions, see Theorem 2.3.9 therein.

Finally, for $0 < \alpha_1 < \alpha_2 < 1$, we introduce the quantity:

$$\omega(\alpha_1, \alpha_2, x, h) = \sup_{\alpha \in [\alpha_1, \alpha_2]} \sup_{x' \in B(x,h)} \left| \log \frac{Q(\alpha|x')}{Q(\alpha|x)} \right|,$$

which is the uniform oscillation of the log-quantile function in its second argument. Such a quantity is also studied in Gardes and Stupfler [22], for instance. It acts as a measure of how close conditional distributions are for two neighboring values of the covariate.

These elements make it possible to state an asymptotic result for our conditional extreme quantile estimator:

Theorem 3.1. *Assume that conditions (2.3) and (H_{SO}) are satisfied and let $u_{n,x} \in (0,1)$ be a sequence converging to 0 and such that $\operatorname{supp}(\mu) \subset [\tau u_{n,x}, u_{n,x}]$ with $\tau \in (0,1]$. Assume also that $m_x(h) \to \infty$ and that there exists $a(x) \in (0,1)$ such that:*

$$c_1 \le \liminf_{n\to\infty} u_{n,x}[m_x(h)]^{a(x)} \le \limsup_{n\to\infty} u_{n,x}[m_x(h)]^{a(x)} \le c_2 \tag{3.2}$$

for some constants $0 < c_1 \le c_2$, $z^{1-a(x)}\Delta^2(z^{a(x)}|x) \to \lambda(x) \in \mathbb{R}$ as $z \to \infty$ and

$$[m_x(h)]^{1-a(x)}\omega^2\Big([m_x(h)]^{-1-\delta}, 1-[m_x(h)]^{-1-\delta}, x, h\Big) \to 0 \tag{3.3}$$

for some $\delta > 0$. If moreover $[m_x(h)]^{(1-a(x))/2}(\widehat{\gamma}_n(x) - \gamma(x)) \xrightarrow{d} \Gamma$ with Γ a non-degenerate distribution, then, provided we have that $\beta_n[m_x(h)]^{a(x)} \to 0$ and $[m_x(h)]^{a(x)-1}\log^2([m_x(h)]^{-a(x)}/\beta_n) \to 0$, it holds that

$$\frac{[m_x(h)]^{(1-a(x))/2}}{\log([m_x(h)]^{-a(x)}/\beta_n)}\left(\frac{\breve{Q}_n(\beta_n|x)}{Q(\beta_n|x)} - 1\right) \xrightarrow{d} \Gamma.$$

Note that $[m_x(h)]^{1-a(x)} \to \infty$ depends on the average number of available data points that can be used to compute the estimator. More precisely, under condition (3.2), this quantity is essentially proportional to $u_{n,x}m_x(h)$, which is the average number of data points actually used in the estimation. In particular, the conditions in Theorem 3.1 are analogues of the classical hypotheses in the estimation of an extreme quantile. Besides, condition (3.3) ensures that the distribution of Y given $X = x'$ is close enough to that of Y given $X = x$ when x' is in a sufficiently small neighborhood of x. Finally, taking μ to be the Dirac measure at $u_{n,x}$ makes it possible to obtain the asymptotic properties of the functional adaptation of the standard Weissman extreme quantile estimator. In particular, as in the unconditional univariate case, the asymptotic distribution of the conditional extrapolated estimator depends crucially on the asymptotic properties of the conditional tail index estimator used.

We proceed by stating the asymptotic normality of the estimator $\widehat{\gamma}_n(x,u)$ in (2.5). To this end, an additional hypothesis on the weighting function $\Psi(\cdot,u)$ is required.

(H_Ψ) The function $\Psi(\cdot,u)$ satisfies for all $u \in (0,1]$ and $\beta \in (0,u]$:

$$\frac{u}{\beta}\int_0^\beta \Psi(\alpha,u)d\alpha = \Phi(\beta/u) \quad \text{and} \quad \sup_{0<v\le 1/2}\int_0^v |\Psi(\alpha,v)|d\alpha < \infty,$$

where Φ is a nonincreasing probability density function on $(0,1)$ such that $\Phi^{2+\kappa}$ is integrable for some $\kappa > 0$. In addition, there exists a positive continuous function g defined on $(0,1)$ such that for any $k > 1$ and $i \in \{1,2,...,k\}$,

$$|i\Phi\,(i/k) - (i-1)\Phi\,((i-1)/k)| \le g\,(i/(k+1))\,, \tag{3.4}$$

and the function $g(\cdot)\max(\log(1/\cdot),1)$ is integrable on $(0,1)$.

Note that for all $t \in (0,1)$, $0 \le t\Phi(t) \le \int_0^{t/2}|\Psi(\alpha,1/2)|d\alpha$. Since the right-hand side converges to 0 as $t\downarrow 0$, we may extend the definition of the map $t \mapsto t\,\Phi(t)$ by saying it is 0 at $t = 0$. Hence, inequality (3.4) is meaningful even when $i = 1$.

Condition (H_Ψ) on the weighting function $\Psi(\cdot,u)$ is similar in spirit to a condition introduced in Beirlant *et al.* [1]. This condition is satisfied for instance by the functions $\Psi(\cdot,u) = u^{-1}$ and $\Psi(\cdot,u) = u^{-1}(\log(u/\cdot) - 1)$ with $g(\cdot) = 1$ for the first one and, for the second one, $g(\cdot) = 1 - \log(\cdot)$. In particular, our results shall then hold for the adaptations of the Hill and Zipf estimators mentioned at the end of Section 2.2.

The asymptotic normality of our family of estimators of $\gamma(x)$ is established in the following theorem.

Theorem 3.2. *Assume that conditions (2.3), (H_{SO}) and (H_Ψ) are satisfied, that $m_x(h) \to \infty$ and $u = u_{n,x} \to 0$. Assume that there exists $a(x) \in (0,1)$ such that $z^{1-a(x)}\Delta^2(z^{a(x)}|x) \to \lambda(x) \in \mathbb{R}$ as $z \to \infty$, condition (3.3) holds and that there are two ultimately decreasing functions $\varphi_1 \le \varphi_2$ such that $z^{1-a(x)}\varphi_2^2(z) \to 0$ as $z \to \infty$ and $\varphi_1(m_x(h)) \le u_{n,x}[m_x(h)]^{a(x)} - 1 \le \varphi_2(m_x(h))$. Then, $[m_x(h)]^{(1-a(x))/2}\,(\widehat{\gamma}_n(x,u_{n,x}) - \gamma(x))$ converges in distribution to*

$$\mathcal{N}\left(\lambda(x)\int_0^1 \Phi(\alpha)\alpha^{-\rho(x)}d\alpha, \gamma^2(x)\int_0^1 \Phi^2(\alpha)d\alpha\right).$$

Our asymptotic normality result thus holds under generalizations of the common hypotheses on the standard univariate model, provided the conditional distributions of Y at two neighboring points are sufficiently close. We close this section by pointing out that our main results are also similar in spirit to results obtained in the literature for other conditional tail index or conditional extreme-value index estimators, see *e.g.* Gardes and Stupfler [22] and Stupfler [37, 38].

4. SIMULATION STUDY

4.1. Hyperparameters selection

The aim of this paragraph is to propose a selection procedure of the hyperparameters involved in the estimator $\check{Q}_n(\beta_n|x)$ of the extreme conditional quantile and in the estimator $\widehat{\gamma}_n(x,u)$ of the functional tail index. Assuming that the measure μ used in (2.4) is such that $\mathrm{supp}(\mu) \subset [\tau u, u]$ for some $\tau \in (0,1)$ fixed by the user (a discussion of the performance of the estimator as a function of τ is included in Section 4.2 below), these hyperparameters are: the bandwidth h controlling the smoothness of the estimators and the value $u \in (0,1)$ which selects the part of the tail distribution considered in the estimation procedure. The criterion used in our selection procedure is based on the following remark: for any positive and integrable weight function $W : [0,1] \to [0,\infty)$,

$$E_W := \mathbb{E}\left[\int_0^1 W(\alpha)\left(\mathbb{I}\{Y > Q(\alpha|X)\} - \alpha\right)^2 d\alpha\right] = \int_0^1 W(\alpha)\alpha(1-\alpha)d\alpha.$$

The sample analogue of E_W is given by

$$\frac{1}{n}\sum_{i=1}^n \int_0^1 W(\alpha)\left(\mathbb{I}\{Y_i > Q(\alpha|X_i)\} - \alpha\right)^2 d\alpha,$$

and for a good choice of h and u, this quantity should of course be close to the known quantity E_W. Let then $W_n^{(1)}$ and $W_n^{(2)}$ be two positive and integrable weight functions. Replacing the unobserved variable $Q(\alpha|X_i)$ by the statistic $\widehat{Q}_{n,i}(\alpha|X_i)$ which is the estimator (2.2) computed without the observation (X_i, Y_i) leads to the following estimator of $E_{W_n^{(1)}}$:

$$\widehat{E}^{(1)}_{W_n^{(1)}}(h) := \frac{1}{n}\sum_{i=1}^n \int_0^1 W_n^{(1)}(\alpha)\left(\mathbb{I}\{Y_i > \widehat{Q}_{n,i}(\alpha|X_i)\} - \alpha\right)^2 d\alpha.$$

Note that $\widehat{E}^{(1)}_{W_n^{(1)}}(h)$ only depends on the hyperparameter h. In the same way, one can also replace $Q(\alpha|X_i)$ by the statistic $\check{Q}_{n,i}(\alpha|X_i)$ which is the estimator (2.4) computed without the observation (X_i, Y_i). An estimator of $E_{W_n^{(2)}}$ is then given by:

$$\widehat{E}^{(2)}_{W_n^{(2)}}(u,h) := \frac{1}{n}\sum_{i=1}^n \int_0^1 W_n^{(2)}(\alpha)\left(\mathbb{I}\{Y_i > \check{Q}_{n,i}(\alpha|X_i)\} - \alpha\right)^2 d\alpha.$$

Obviously, this last quantity depends both on u and h. We propose the following two-stage procedure to choose the hyperparameters u and h. First, we compute our selected bandwidth h^{opt} by minimizing with respect to h the function

$$\text{CV}^{(1)}(h) := \left[\widehat{E}^{(1)}_{W_n^{(1)}}(h) - \int_0^1 W_n^{(1)}(\alpha)\alpha(1-\alpha)d\alpha\right]^2.$$

Next, our selected sample fraction u^{opt} is obtained by minimizing with respect to u the function $\text{CV}^{(2)}(u, h^{\text{opt}})$ where

$$\text{CV}^{(2)}(u,h) := \left[\widehat{E}^{(2)}_{W_n^{(2)}}(u,h) - \int_0^1 W_n^{(2)}(\alpha)\alpha(1-\alpha)d\alpha\right]^2.$$

Note that the functions $\text{CV}^{(1)}$ and $\text{CV}^{(2)}$ can be seen as adaptations to the problem of conditional extreme quantile estimation of the cross-validation function introduced in Li *et al.* [32].

4.2. Results

The behavior of the extreme conditional quantile estimator (2.4), when the estimator (2.5) of the functional tail index is used together with our selection procedure of the hyperparameters, is tested on some random pairs $(X,Y) \in \mathcal{C}^1[-1,1] \times (0,\infty)$, where $\mathcal{C}^1[-1,1]$ is the space of continuously differentiable functions on $[-1,1]$. We generate $n = 1000$ independent copies $(X_1,Y_1),...,(X_n,Y_n)$ of (X,Y) where X is the random curve defined for all $t \in [-1,1]$ by $X(t) := \sin[2\pi t U] + (V+2\pi)t + W$, where U, V and W are independent random variables drawn from a standard uniform distribution. Note that this random covariate was used for instance in Ferraty *et al.* [16]. Regarding the conditional distribution of Y given $X = x$, $x \in \mathcal{C}^1[-1,1]$, two distributions are considered. The first one is the Fréchet distribution, for which the conditional quantile is given for all $\alpha \in (0,1)$ by $Q(\alpha|x) = [-\log(1-\alpha)]^{-\gamma(x)}$. The second one is the Burr distribution with parameter $r > 0$, for which $Q(\alpha|x) = (\alpha^{-r\gamma(x)} - 1)^{1/r}$. For these distributions, letting x' be the first derivative of x and

$$z(x) = \frac{2}{3}\left[\int_{-1}^1 x'(t)[1-\cos(\pi t)]dt - \frac{23}{2}\right],$$

the functional tail index is given by

$$\gamma(x) = \exp\left[-\frac{\log(3)}{9}z^2(x)\right]\mathbb{I}\{|z(x)| < 3\} + \frac{1}{3}\mathbb{I}\{|z(x)| \geq 3\}.$$

In this setup, it is straightforward to show that $z(x) \in [-3.14, 3.07]$ approximately, and therefore the range of values of $\gamma(x)$ is the full interval $[1/3, 1]$. Let us also mention that the second order parameter $\rho(x)$ appearing in condition (H_{SO}) is then $\rho(x) = -1$ for the Fréchet distribution and $\rho(x) = -r\gamma(x)$ for the Burr distribution; in the latter case, the range of values of $\rho(x)$ is therefore $[-r, -r/3]$.

The space $\mathcal{C}^1[-1,1]$ is endowed with the semi-metric d given for all x_1, x_2 by

$$d(x_1, x_2) = \left[\int_{-1}^{1} \left(x_1'(t) - x_2'(t)\right)^2 dt\right]^{1/2},$$

i.e. the L^2-distance between first derivatives. To compute $\widehat{\gamma}_n(x,u)$, we use the weight function $\Psi(\cdot, u) = u^{-1}(\log(u/\cdot) - 1)$, and the measure μ used in the integrated conditional quantile estimator is assumed to be absolutely continuous with respect to the Lebesgue measure, with density

$$p_{\tau,u}(\alpha) = \frac{1}{u(1-\tau)} \mathbb{I}\{\alpha \in [\tau u, u]\}.$$

In what follows, this estimator is referred to as the Integrated Weissman Estimator (IWE). Other absolutely continuous measures μ, with different densities with respect to the Lebesgue measure, have been tested, with different values of τ. It appears that the impact of the choice of the parameter τ is more important than the one of the measure μ. We thus decided to present in this simulation study the results for the aforementioned value of the measure μ only, but with different tested values for τ.

The hyperparameters are selected using the procedure described in Section 4.1. Since we are interested in the tail of the conditional distribution, the supports of the weight functions $W_n^{(1)}$ and $W_n^{(2)}$ should be located around 0. More specifically, for $i \in \{1,2\}$, we take

$$W_n^{(i)}(\alpha) := \log\left(\frac{\alpha}{\beta_{n,1}^{(i)}}\right) \mathbb{I}\{\alpha \in [\beta_{n,1}^{(i)}, \beta_{n,2}^{(i)}]\},$$

where $\beta_{n,1}^{(1)} = \lfloor 2\sqrt{n}\log n \rfloor / n$, $\beta_{n,2}^{(1)} = \lfloor 3\sqrt{n}\log n \rfloor / n$, $\beta_{n,1}^{(2)} = \lfloor 5 \log n \rfloor / n$ and $\beta_{n,2}^{(2)} = \lfloor 10 \log n \rfloor / n$. The cross-validation function $\mathrm{CV}^{(1)}(h)$ is minimized over a grid $\mathcal{H}$ of 20 points evenly spaced between 1/2 and 10 to obtain the optimal value h^{opt}, while the value u^{opt} is obtained by minimizing over a grid $\mathcal{U}$ of 26 points evenly spaced between 0.005 and 0.255 the function $\mathrm{CV}^{(2)}(u, h^{\mathrm{opt}})$.

For the Fréchet distribution and two Burr distributions (one with $r = 2$ and one with $r = 1/20$), the conditional extreme quantile estimator (2.4) is computed with the values u^{opt} and h^{opt} obtained by our selection procedure. The quality of the estimator is measured by the Integrated Squared Error given by:

$$ISE := \frac{1}{n} \sum_{i=1}^{n} \int_{\beta_{n,1}^{(2)}}^{\beta_{n,2}^{(2)}} \log^2 \frac{\breve{Q}_{n,i}(\alpha|X_i)}{Q(\alpha|X_i)} d\alpha.$$

This procedure is repeated $N = 100$ times. To give a graphical idea of the behavior of our estimator (2.4), we first depict, in Figure 1, boxplots of the N obtained replications of this estimator, computed with $\tau = 9/10$, for the Fréchet distribution and for some values of the quantile order β_n and of the covariate x.

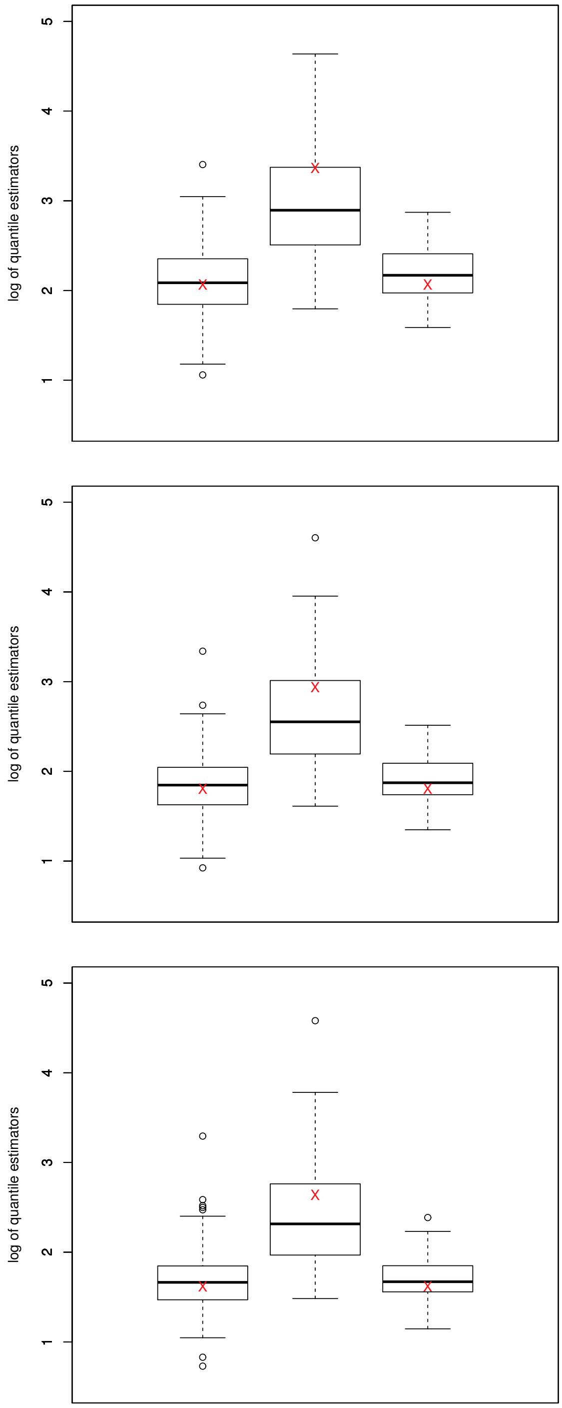

Figure 1: For the Fréchet distribution, boxplots of (the logarithm of) estimator (2.4) for $\beta_n = \beta^{(2)}_{n,1}$ (top), $\beta_n = (\beta^{(2)}_{n,1} + \beta^{(2)}_{n,2})/2$ (middle) and $\beta_n = \beta^{(2)}_{n,2}$ (bottom). In each picture, the covariate x is respectively (from left to right) such that $z(x) = -2$ ($\gamma(x) \approx 0.64$), $z(x) = 0$ ($\gamma(x) = 1$) and $z(x) = 2$ ($\gamma(x) \approx 0.64$). In each case, the true value of the conditional quantile to be estimated is represented by a cross.

More precisely, we take here $\beta_n \in \{\beta_{n,1}^{(2)}, (\beta_{n,1}^{(2)} + \beta_{n,2}^{(2)})/2, \beta_{n,2}^{(2)}\}$ and three values of the covariate are considered: $x = x_1$ with $z(x_1) = -2$ (and then $\gamma(x_1) \approx 0.64$), $x = x_2$ with $z(x_2) = 0$ (giving $\gamma(x_2) = 1$) and $x = x_3$ with $z(x_3) = 2$ (which entails $\gamma(x_3) \approx 0.64$). As expected, the quality of the estimation is strongly impacted by the quantile order β_n but also by the actual position of the covariate and, of course, by the value of the true conditional tail index $\gamma(x)$.

Next, the median and the first and third quartiles of the N values of the Integrated Squared Error are gathered in Table 1. The proposed estimator is compared to the adaptation of the Weissman estimator obtained by taking for the measure μ in (2.4) the Dirac measure at u. This estimator is referred to as the Weissman Estimator (WE) in Table 1. In the WE estimator, the functional tail index $\gamma(x)$ is estimated either by (2.6) or by the generalized Hill-type estimator of Gardes and Girard [21]: for $J \geq 2$, this estimator is given by

$$\widehat{\gamma}^{\mathrm{GG}}(x,u) = \sum_{j=1}^{J} \left(\log \widehat{Q}_n(u/j^2|x) - \log \widehat{Q}_n(u|x)\right) \Bigg/ \sum_{j=1}^{J} \log(j^2) \ .$$

Following their advice, we set $J = 10$. Again, the median and the first and third quartiles of the N values of the Integrated Squared Error of these two estimators are given in Table 1. In this Table, optimal median errors among the five tested estimators are marked in boldface characters. It appears that the IWEs outperform the two WEs in the case of the Fréchet and Burr (with $r = 1/20$) distributions.

Table 1: Comparison of the Integrated Squared Errors of the following extreme conditional quantile estimators: IWE with $\tau \in \{1/10, 1/2, 9/10\}$ (lines 1 to 3), WE when $\gamma(x)$ is estimated by (2.6) (line 4) and WE when $\gamma(x)$ is estimated by the Hill-type estimator (line 5). Results are given in the following form: `[first quartile  median  third quartile]`.

	Fréchet dist.	Burr dist. ($r = 2$)
IWE ($\tau = 1/10$)	[0.0060 0.0077 0.0132]	[0.0063 0.0099 0.0147]
IWE ($\tau = 1/2$)	[0.0060 0.0077 0.0112]	[0.0058 0.0095 0.0128]
IWE ($\tau = 9/10$)	[0.0058 **0.0076** 0.0107]	[0.0059 0.0093 0.0119]
WE (with (2.6))	[0.0054 0.0078 0.0115]	[0.0054 **0.0088** 0.0137]
WE (Hill-type)	[0.0068 0.0094 0.0120]	[0.0071 0.0103 0.0137]

	Burr dist. ($r = 1/20$)
IWE ($\tau = 1/10$)	[0.6427 0.9504 1.3982]
IWE ($\tau = 1/2$)	[0.6040 **0.8343** 1.2018]
IWE ($\tau = 9/10$)	[0.8010 1.0870 1.2725]
WE (with (2.6))	[0.5848 0.8909 1.3372]
WE (Hill-type)	[0.7679 1.1314 1.4599]

It also seems that the choice of τ has some influence on the quality of the estimator but, unfortunately, an optimal choice of τ apparently depends on the unknown underlying distribution. It is interesting though to note that the optimal IWE estimator among the three tested here always enjoys a smaller variability than the WE estimator: for instance, in the case of the Burr distribution with $r = 2$, even though the IWE with $\tau = 9/10$ does not outperform the WE (with $\gamma(x)$ estimated by (2.6)) in terms of median ISE, the interquartile range of the ISE is 27.7% lower for the IWE compared to what it is for the WE. Finally, as expected, the value of $\rho(x)$ has a strong impact on the estimation procedure: a value of $\rho(x)$ close to 0 leads to large values of the Integrated Squared Error.

5. REAL DATA EXAMPLE

In this section, we showcase our extreme quantile Integrated Weissman Estimator on functional chemometric data. This data, obtained by considering $n = 215$ pieces of finely chopped meat, consists of pairs of observations (x_n, z_n), where x_i is the absorbance curve of the ith piece of meat, obtained at 100 regularly spaced wavelengths between 850 and 1050 nanometers (this is also called the spectrometric curve), and z_i is the percentage of fat content in this piece of meat. The data, openly available at `http://lib.stat.cmu.edu/datasets/tecator`, is for instance considered in Ferraty and Vieu [14, 15]. Figure 2 is a graph of all 215 absorbance curves.

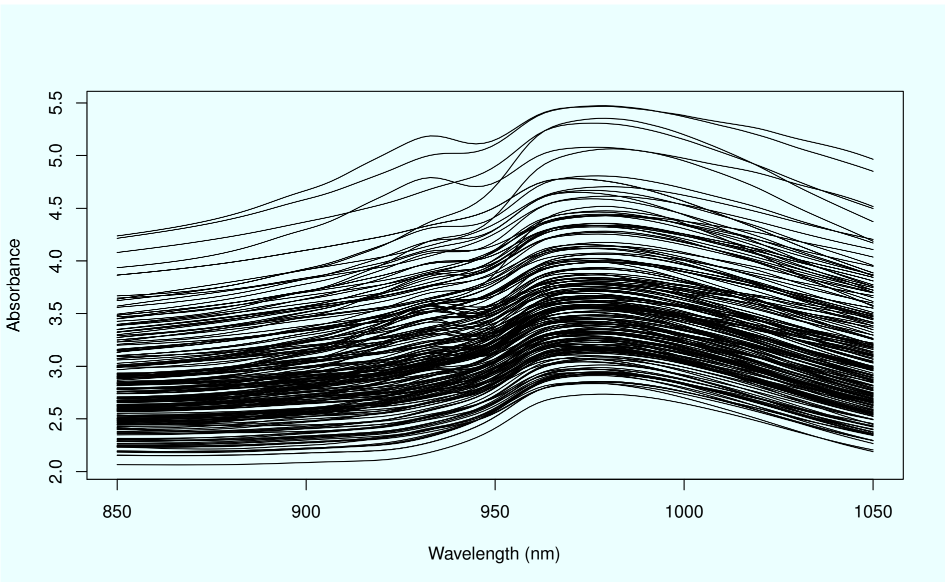

Figure 2: Spectrometric curves for the data.

Because the percentage of fat content z_i obviously belongs to $[0, 100]$, it has a finite-right endpoint and therefore cannot be conditionally heavy-tailed as required by model (2.1). We thus consider the "inverse fat content" $y_i = 100/z_i$ in this analysis. The top panel of Figure 3 shows the Hill plot of the sample $(y_1, ..., y_n)$ without integrating covariate information. It can be seen in this figure that the Hill plot seems to be stabilizing near the value 0.4 for a sizeable portion of the left of the graph, thus indicating the plausible presence of a heavy right tail in the data $(y_1, ..., y_n)$, see for instance Theorem 3.2.4 in de Haan and Ferreira [27]. The other panels in Figure 3 show exponential QQ-plots for the log-data points whose covariates lie in a fixed-size neighborhood of certain pre-specified points in the covariate space. It is seen in these subfigures that these plots are indeed roughly linear towards their right ends, which supports our conditional heavy tails assumption.

On these grounds, we therefore would like to analyze the influence of the covariate information, which is the absorbance curve, upon the inverse fat content. While of course the absorbance curves obtained are in reality made of discrete data because of the discretization of this curve, the precision of this discretization arguably makes it possible to consider our data as in fact functional. This, in our opinion, fully warrants the use of our estimator in this case.

Because the covariate space is functional, one has to wonder about how to measure the influence of the covariate and then about how to represent the results. A nice account of the problem of how to represent results when considering functional data is given in Ferraty and Vieu [15]. Here, we look at the variation of extreme quantile estimates in two different directions of the covariate space. To this end, we consider the semi-metric

$$d(x_1, x_2) = \left[\int_{850}^{1050} \left(x_1''(t) - x_2''(t) \right)^2 dt \right]^{1/2},$$

also advised by Ferraty and Vieu [14], and we compute:

- A typical pair of covariates, *i.e.* a pair $(x_1^{\text{med}}, x_2^{\text{med}})$ such that

$$d(x_1^{\text{med}}, x_2^{\text{med}}) = \text{median}\{d(x_i, x_j),\ 1 \le i, j \le n,\ i \ne j\};$$

- A pair of covariates farthest from each other, *i.e.* a pair $(x_1^{\max}, x_2^{\max})$ such that

$$d(x_1^{\max}, x_2^{\max}) = \max\{d(x_i, x_j),\ 1 \le i, j \le n,\ i \ne j\}.$$

For the purpose of comparison, we also compute the "average covariate" $\overline{x} = n^{-1} \sum_{i=1}^n x_i$. In particular, we represent on Figure 4 our two pairs of covariates together with the average covariate, the same scale being used on the y-axis in both figures. Recall that since the semi-metric d is the L^2-distance between second-order derivatives, it acts as a measure of how much the shapes of two covariate curves are different, rather than measuring how far apart they are.

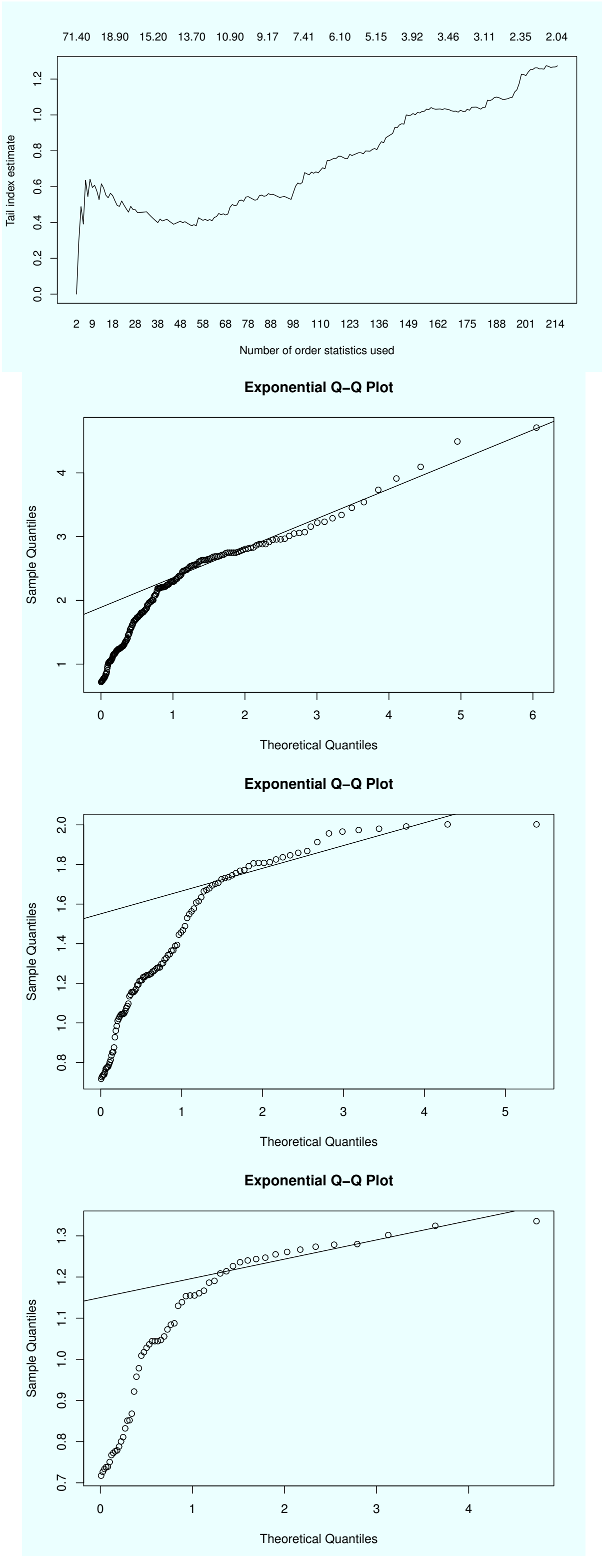

Figure 3: Top panel: Hill plot for the sample $(y_1, ..., y_n)$. On the x-axis at the top of the panel is the value of the lower threshold for the computation of the Hill estimator, *i.e.* the lowest order statistic. Other panels: local exponential QQ-plots for the log-data points whose covariates belong to a neighborhood of certain pre-specified points in the covariate space.

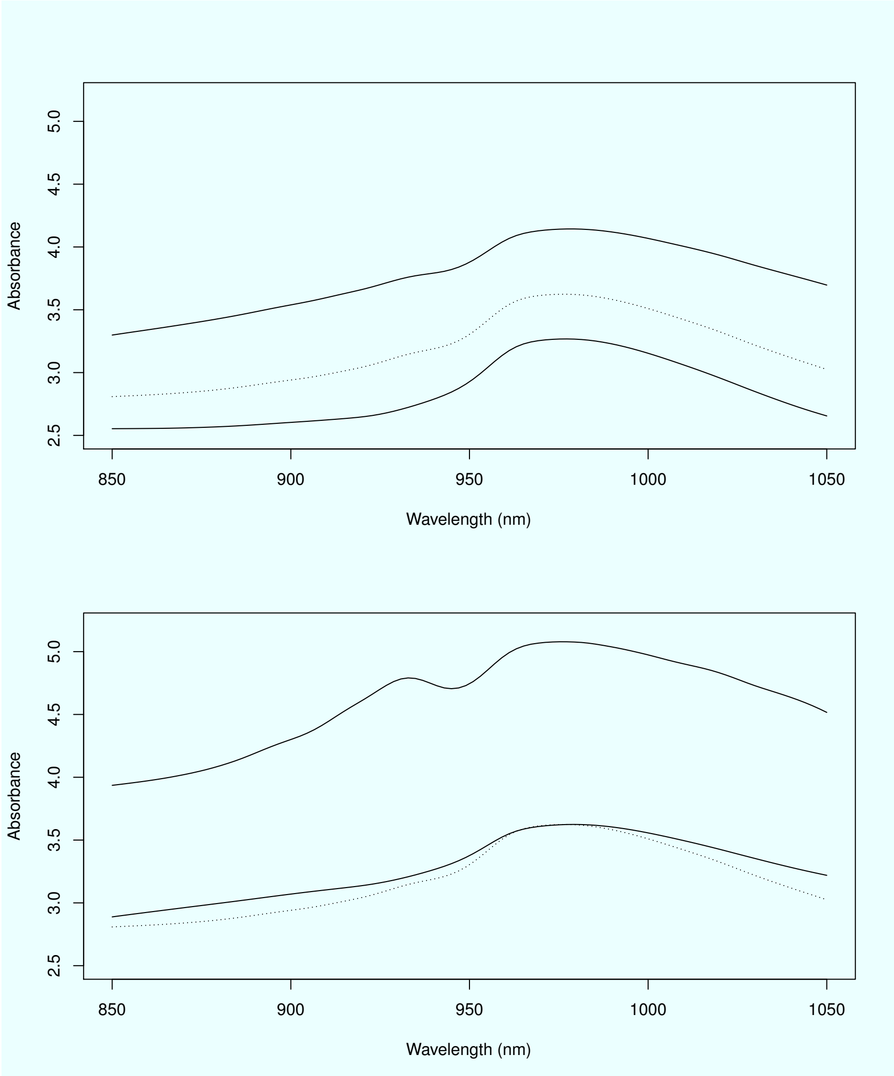

Figure 4: Top picture, solid lines: a pair of typical covariates. Bottom picture, solid lines: the pair of covariates farthest from each other. In both pictures the dotted line is the average covariate.

We compute our conditional extreme quantile estimator at the levels $5/n$ and $1/n$, using the methodology given in Section 4.2. In particular, the selection parameters $\beta_{n,1}^{(1)}$, $\beta_{n,2}^{(1)}$, $\beta_{n,1}^{(2)}$ and $\beta_{n,2}^{(2)}$ used in the cross-validation methodology were the exact same ones used in the simulation study, namely 0.437, 0.655, 0.035 and 0.069, respectively. The bandwidth h is selected in the interval $[0.00316, 0.0116]$, the lower bound in this interval corresponding to the median of all distances $d(x_i, x_j)$ $(i \neq j)$ and the upper bound corresponding to 90% of the maximum of all distances $d(x_i, x_j)$, for a final selected value of 0.00717. The value of the parameter u is selected exactly as in the simulation study, and the selection procedure gives the value 0.185. Finally, we set $\tau = 0.9$ in our Integrated Weissman Estimator.

Results are given in Figure 5; namely, we compute the extreme quantile estimates $\breve{Q}_n(\beta|x)$, for $\beta \in \{5/n, 1/n\}$, and x belonging to either the line $[x_1^{\text{med}}, x_2^{\text{med}}] = \{tx_1^{\text{med}} + (1-t)x_2^{\text{med}},\ t \in [0,1]\}$ or to the line $[x_1^{\max}, x_2^{\max}]$. It can be seen in these figures that the estimates in the direction of a typical pair of covariates are remarkably stable; they are actually essentially indistinguishable from the estimates at the average covariate, which are 42.41 for $\beta = 5/n$ and 93.86 for $\beta = 1/n$.

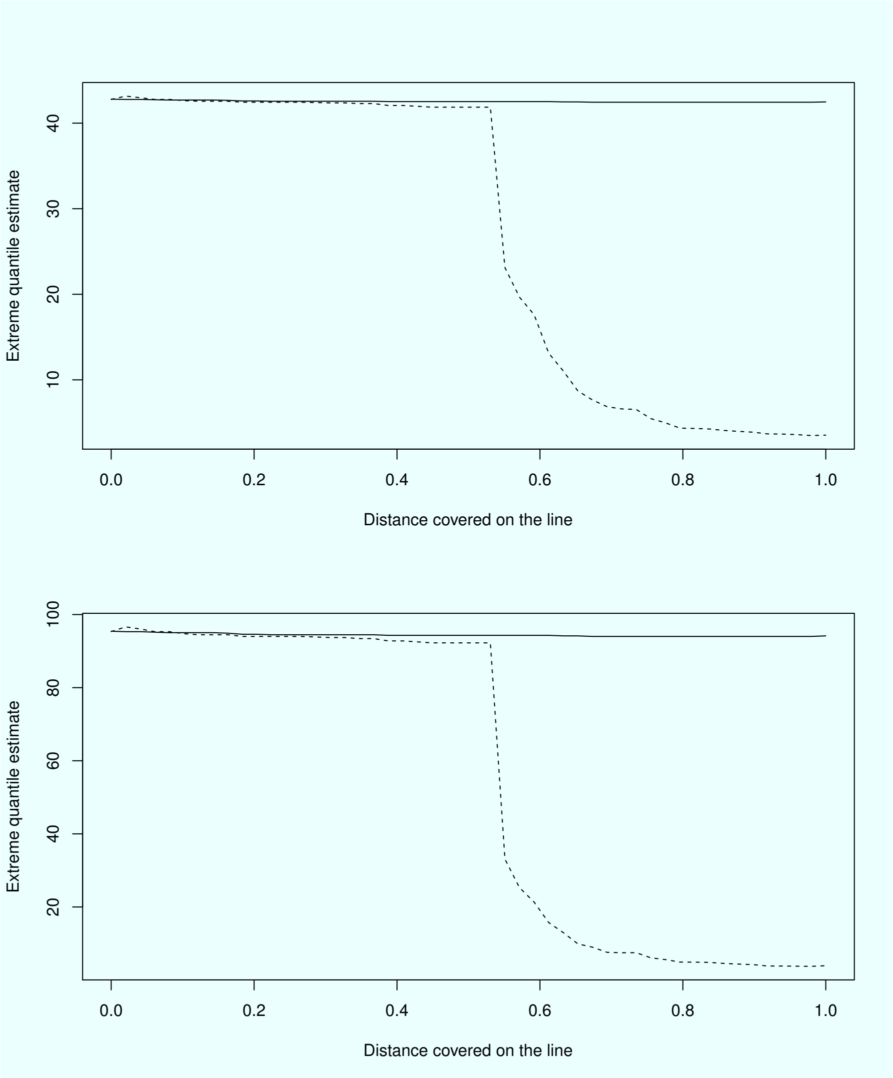

Figure 5: Solid line: extreme quantile estimate in the direction of a typical pair of covariates, dashed line: extreme quantile estimate in the direction of a pair of covariates farthest from each other. Top picture: case $\beta = 5/n$, bottom picture: $\beta = 1/n$.

By contrast, the estimates on the line $[x_1^{\max}, x_2^{\max}]$, while roughly stable for 60% of the line and approximately equal to the value of the estimated quantiles at the average covariate, very sharply drop afterwards, the reduction factor being close to 10 from the beginning of the line to its end in the case $\beta = 5/n$.

This conclusion suggests that while in typical directions of the covariate space the tail behavior of the fat content is very stable, there may be certain directions in which this is not the case. In particular, there appear to be certain values of the covariate for which thresholds for the detection of unusual levels of fat should differ from those of more standard cases.

6. PROOFS OF THE MAIN RESULTS

Before proving the main results, we recall two useful facts. The first one is a classical equivalent of

$$M_x(h) := \sum_{i=1}^{n} \mathbb{I}\{d(X_i, x) \leq h\}.$$

If $m_x(h) \to \infty$ as $n \to \infty$ then, for any $\delta \in (0,1)$:

$$[m_x(h)]^{(1-\delta)/2} \left| \frac{M_x(h)}{m_x(h)} - 1 \right| \xrightarrow{\mathbb{P}} 0 \quad \text{as} \quad n \to \infty, \tag{6.1}$$

see Lemma 1 in Stupfler [37]. For the second one, let $\{Y_i^*,\ i = 1, ..., M_x(h)\}$ be the response variables whose associated covariates $\{X_i^*,\ i = 1, ..., M_x(h)\}$ are such that $d(X_i^*, x) \leq h$. Lemma 4 in Gardes and Stupfler [22] shows that the random variables $V_i = 1 - F(Y_i^*|X_i^*)$ are such that, for all $u_1, ..., u_p \in [0,1]$,

$$\mathbb{P}\left(\bigcap_{i=1}^{p} \{V_i \leq u_i\} | M_x(h) = p \right) = u_1 ... u_p, \tag{6.2}$$

i.e. they are independent standard uniform random variables given $M_x(h)$.

6.1. Proof of Theorem 3.1

The following proposition is a uniform consistency result for the estimator $\widehat{Q}_n(\beta_n|x)$ when β_n goes to 0 at a moderate rate.

Proposition 6.1. *Assume that conditions (2.3), (H_{SO}), (3.2) and (3.3) are satisfied. If $m_x(h) \to \infty$, then*

$$\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x)}{Q(\alpha|x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}\left([m_x(h)]^{(a(x)-1)/2} \right).$$

Proof: Let $M_n := M_x(h)$, $\{U_i,\ i \geq 1\}$ be independent standard uniform random variables, $V_i := S(Y_i^*|X_i^*)$ and

$$Z_n(x) := \sup_{\alpha\in[\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x)}{Q(\alpha|x)} - 1 \right|.$$

We start with the following inequality: $Z_n(x) \leq T_n(x) + R_n^{(Q)}(x)$, with

$$T_n(x) := \sup_{\alpha\in[\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(V_{\lfloor \alpha M_n\rfloor+1, M_n}|x)}{Q(\alpha|x)} - 1 \right| \tag{6.3}$$

$$\text{and} \quad R_n^{(Q)}(x) := \sup_{\alpha\in[\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x) - Q(V_{\lfloor \alpha M_n\rfloor+1, M_n}|x)}{Q(\alpha|x)} \right|. \tag{6.4}$$

Let us first focus on the term $T_n(x)$. For any $t > 0$,

$$\mathbb{P}(v_{n,x} T_n(x) > t) = \sum_{j=0}^{n} \mathbb{P}(v_{n,x} T_n(x) > t | M_n = j)\mathbb{P}(M_n = j),$$

where $v_{n,x} := [m_x(h)]^{(1-a(x))/2}$. From (6.1), letting

$$I_n := [m_x(h)(1 - [m_x(h)]^{[a(x)/4]-1/2}), m_x(h)(1 + [m_x(h)]^{[a(x)/4]-1/2})], \tag{6.5}$$

one has $\mathbb{P}(M_n \notin I_n) \to 0$ as $n \to \infty$. Hence,

$$\mathbb{P}(v_{n,x} T_n(x) > t) \leq \sup_{p\in I_n} \mathbb{P}(v_{n,x}, T_n(x) > t | M_n = p) + \mathrm{o}(1).$$

Using Lemma A.1,

$$\sup_{p\in I_n} \mathbb{P}(v_{n,x} T_n(x) > t | M_n = p) = \sup_{p\in I_n} \mathbb{P}(v_{n,x} \underline{T}_p(x) > t),$$

where

$$\underline{T}_p(x) := \sup_{\alpha\in[\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(U_{\lfloor p\alpha\rfloor+1, p}|x)}{Q(\alpha|x)} - 1 \right|.$$

Using condition (3.2), it is clear that there are constants $d_1, d_2 > 0$ with $d_1 < d_2$ such that for n large enough, we have for all $p \in I_n$:

$$\underline{T}_p(x) \leq \sup_{\alpha\in[d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha\rfloor+1, p}|x)}{Q(\alpha|x)} - 1 \right|.$$

Thus, for all $t > 0$, $\mathbb{P}(v_{n,x} T_n(x) > t)$ is bounded above by

$$\sup_{p\in I_n} \mathbb{P}\left(v_{n,x} \sup_{\alpha\in[d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha\rfloor+1, p}|x)}{Q(\alpha|x)} - 1 \right| > t \right) + \mathrm{o}(1).$$

Furthermore, for n large enough, there exists $\kappa > 0$ such that for all $p \in I_n$, $v_{n,x} \leq \kappa p^{(1-a(x))/2}$ and thus, for all $t > 0$, $\mathbb{P}(v_{n,x} T_n(x) > t)$ is bounded above by

$$\sup_{p\in I_n} \mathbb{P}\left(\kappa p^{(1-a(x))/2} \sup_{\alpha\in[d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha\rfloor+1, p}|x)}{Q(\alpha|x)} - 1 \right| > t \right) + \mathrm{o}(1).$$

Since
$$p^{(1-a(x))/2} \sup_{\alpha\in[d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha\rfloor+1,p}|x)}{Q(\alpha|x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(1),$$
(see Lemma A.2 for a proof), it now becomes clear that $T_n(x) = \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1})$.

Let us now focus on the term $R_n^{(Q)}(x)$. As before, one can show that for all $t > 0$,
$$\mathbb{P}(v_{n,x} R_n^{(Q)}(x) > t) \le \sup_{p\in I_n} \mathbb{P}(v_{n,x} R_n^{(Q)}(x) > t | M_n = p) + \mathrm{o}(1).$$
Lemma A.1 and condition (3.3) yield for any $t > 0$ and n large enough:
$$\begin{aligned}
&\sup_{p\in I_n} \mathbb{P}(v_{n,x} R_n^{(Q)}(x) > t | M_n = p) \\
\le\ &\sup_{p\in I_n} \mathbb{P}(v_{n,x}\omega(U_{1,p}, U_{p,p}, x, h)\exp(\omega(U_{1,p}, U_{p,p}, x, h))(1 + \underline{T}_p(x)) > t) \\
\le\ &\sup_{p\in I_n} \mathbb{P}(p^{(1-a(x))/2}\omega(U_{1,p}, U_{p,p}, x, h)\exp(\omega(U_{1,p}, U_{p,p}, x, h))(1 + \underline{T}_p(x)) > t/\kappa) \\
\le\ &\sup_{p\in I_n} \left[\mathbb{P}(U_{1,p} < [m_x(h)]^{-1-\delta}) + \mathbb{P}(U_{p,p} > 1 - [m_x(h)]^{-1-\delta})\right].
\end{aligned}$$
Since for n large enough
$$\begin{aligned}
(6.6)\quad &\sup_{p\in I_n} \left[\mathbb{P}(U_{1,p} < [m_x(h)]^{-1-\delta}) + \mathbb{P}(U_{p,p} > 1 - [m_x(h)]^{-1-\delta})\right] \\
&= 2\sup_{p\in I_n} \left[1 - [1 - [m_x(h)]^{-1-\delta}]^p\right] \le 2\left(1 - [1 - [m_x(h)]^{-1-\delta}]^{2m_x(h)}\right) \to 0,
\end{aligned}$$
we thus have proven that $R_n^{(Q)}(x) = \mathrm{o}_{\mathbb{P}}(v_{n,x}^{-1})$ and the proof is complete. □

Proof of Theorem 3.1: The key point is to write
$$\check{Q}_n(\beta_n|x) = \int_{\tau u_{n,x}}^{u_{n,x}} Q(\alpha|x)\left(\frac{\alpha}{\beta_n}\right)^{\gamma(x)} \left\{ \frac{\widehat{Q}_n(\alpha|x)}{Q(\alpha|x)} \left(\frac{\alpha}{\beta_n}\right)^{\widehat{\gamma}_n(x)-\gamma(x)} \right\} \mu(d\alpha).$$
Now, by assumption $v_{n,x}(\widehat{\gamma}_n(x) - \gamma(x)) \xrightarrow{d} \Gamma$ where $v_{n,x} := [m_x(h)]^{(1-a(x))/2}$. Since $\beta_n/u_{n,x}$ is asymptotically bounded from below and above by sequences proportional to $\beta_n[m_x(h)]^{a(x)} \to 0$, one has for n large enough that
$$\sup_{\alpha\in[\tau u_{n,x}, u_{n,x}]} \left| \log\left[\left(\frac{\alpha}{\beta_n}\right)^{\widehat{\gamma}_n(x)-\gamma(x)}\right] \right| \le |\widehat{\gamma}_n(x) - \gamma(x)| \log\left(\frac{u_{n,x}}{\beta_n}\right) = \mathrm{o}_{\mathbb{P}}(1),$$
since by assumption $v_{n,x}^{-1}\log(u_{n,x}/\beta_n) \to 0$. A Taylor expansion for the exponential function thus yields
$$\left(\frac{\alpha}{\beta_n}\right)^{\widehat{\gamma}_n(x)-\gamma(x)} - 1 - \log(\alpha/\beta_n)(\widehat{\gamma}_n(x) - \gamma(x)) = \mathcal{O}_{\mathbb{P}}\left(v_{n,x}^{-1}\log^2(u_{n,x}/\beta_n)\right),$$

uniformly in $\alpha \in [\tau u_{n,x}, u_{n,x}]$. We then obtain

$$\check{Q}_n(\beta_n|x) = \int_{\tau u_{n,x}}^{u_{n,x}} Q(\alpha|x) \left(\frac{\alpha}{\beta_n}\right)^{\gamma(x)} G_{n,x}(\alpha)\mu(d\alpha)$$

where

$$G_{n,x}(\alpha) := \frac{\widehat{Q}_n(\alpha|x)}{Q(\alpha|x)} \left[1 + \log(\alpha/\beta_n)(\widehat{\gamma}_n(x) - \gamma(x)) + \mathcal{O}_{\mathbb{P}}\left(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n)\right)\right].$$

By Proposition 6.1,

$$\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x)}{Q(\alpha|x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1}),$$

and therefore:

$$G_{n,x}(\alpha) = 1 + \log(\alpha/\beta_n)(\widehat{\gamma}_n(x) - \gamma(x)) + \mathcal{O}_{\mathbb{P}}\left(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n)\right). \tag{6.7}$$

By Lemma A.3,

$$\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(\alpha|x)}{Q(\beta_n|x)} \left(\frac{\alpha}{\beta_n}\right)^{\gamma(x)} - 1 \right| = \mathcal{O}\left(\Delta(u_{n,x}^{-1}|x)\right), \tag{6.8}$$

and thus, (6.7) and (6.8) lead to

$$\begin{aligned} \frac{\check{Q}(\beta_n|x)}{Q(\beta_n|x)} - 1 \;&=\; (\widehat{\gamma}_n(x) - \gamma(x)) \int_{\tau u_{n,x}}^{u_{n,x}} \log(\alpha/\beta_n)\mu(d\alpha) \left[1 + \mathcal{O}\left(\Delta(u_{n,x}^{-1}|x)\right)\right] \\ &\quad + \mathcal{O}\left(\Delta(u_{n,x}^{-1}|x)\right) + \mathcal{O}_{\mathbb{P}}\left(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n)\right). \end{aligned}$$

Since $u_{n,x}/\beta_n \to 0$ and $\mu([\tau u_{n,x}, u_{n,x}]) = 1$, one has

$$\begin{aligned} \int_{\tau u_{n,x}}^{u_{n,x}} \log(\alpha/\beta_n)\mu(d\alpha) \;&=\; \int_{\tau u_{n,x}}^{u_{n,x}} \left[\log(u_{n,x}/\beta_n) + \log(\alpha/u_{n,x})\right] \mu(d\alpha) \\ &=\; \log(u_{n,x}/\beta_n)(1 + \mathrm{o}(1)), \end{aligned}$$

and thus

$$\begin{aligned} \frac{\check{Q}(\beta_n|x)}{Q(\beta_n|x)} - 1 \;&=\; (\widehat{\gamma}_n(x) - \gamma(x)) \log(u_{n,x}/\beta_n) \left[1 + \mathrm{o}(1)\right] \\ &\quad + \mathcal{O}\left(\Delta(u_{n,x}^{-1}|x)\right) + \mathcal{O}_{\mathbb{P}}\left(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n)\right). \end{aligned}$$

Using the convergence in distribution of $\widehat{\gamma}_n(x)$ completes the proof. □

6.2. Proof of Theorem 3.2

For the sake of brevity, let $v_{n,x} = [m_x(h)]^{(1-a(x))/2}$, $M_n = M_x(h)$ and $K_n = u_{n,x} M_n$. The cumulative distribution function of a normal distribution with mean $\lambda(x) \int_0^1 \Phi(\alpha)\alpha^{-\rho(x)} d\alpha$ and variance $\gamma^2(x) \int_0^1 \Phi^2(\alpha) d\alpha$ is denoted by H_x in what follows. Let $t \in \mathbb{R}$ and $\varepsilon > 0$. Denoting by $E_n(t)$ the event $\{v_{n,x}(\widehat{\gamma}_n(x,u_{n,x}) - \gamma(x)) \leq t\}$, one has

$$|\mathbb{P}[E_n(t)] - H_x(t)| \leq \sum_{p=0}^{n} \mathbb{P}(M_n = p) \left|\mathbb{P}[E_n(t)|M_n = p] - H_x(t)\right|.$$

Recall that from (6.1), $\mathbb{P}(M_n \notin I_n) \to 0$ as $n \to \infty$ where I_n is defined in (6.5). Hence, for n large enough,

$$|\mathbb{P}[E_n(t)] - H_x(t)| \leq \sup_{p \in I_n} |\mathbb{P}[E_n(t)|M_n = p] - H_x(t)| + \frac{\varepsilon}{8}. \tag{6.9}$$

Now, using the notation $V_i := S(Y_i^*|X_i^*)$ for $i = 1, ..., M_n$, let us introduce the statistics:

$$\widetilde{\gamma}_n(x,u_{n,x}) := \sum_{i=1}^{\lfloor K_n \rfloor} W_{i,n}(u_{n,x}, M_n) \log \frac{Q(V_{i,M_n}|x)}{Q(V_{\lfloor K_n \rfloor + 1, M_n}|x)}$$

$$\text{and} \quad R_n^{(\gamma)}(x) := \widehat{\gamma}_n(x,u_{n,x}) - \widetilde{\gamma}_n(x,u_{n,x}), \tag{6.10}$$

where

$$W_{i,n}(u_{n,x}, M_n) := \int_{(i-1)/M_n}^{i/M_n} \Psi(\alpha, u_{n,x}) d\alpha. \tag{6.11}$$

It is straightforward that for all $\kappa > 0$,

$$\sup_{p \in I_n} |\mathbb{P}[E_n(t)|M_n = p] - H_x(t)| \leq T_{n,x}^{(1)} + T_{n,x}^{(2)}, \tag{6.12}$$

where

$$\begin{aligned} T_{n,x}^{(1)} &:= \sup_{p \in I_n} \left|\mathbb{P}\left[E_n(t) \cap \left\{v_{n,x}|R_n^{(\gamma)}(x)| \leq \kappa\right\} |M_n = p\right] - H_x(t)\right| \\ \text{and } T_{n,x}^{(2)} &:= \sup_{p \in I_n} \mathbb{P}\left[v_{n,x}|R_n^{(\gamma)}(x)| > \kappa | M_n = p\right]. \end{aligned}$$

Let us first focus on the term $T_{n,x}^{(1)}$. Let $\widetilde{E}_n(t) := \{v_{n,x}(\widetilde{\gamma}_n(x,u_{n,x}) - \gamma(x)) \leq t\}$. For all $p \in I_n$, $\mathbb{P}[E_n(t) \cap \{v_{n,x}|R_n^{(\gamma)}(x)| \leq \kappa\}|M_n = p] \leq \mathbb{P}[\widetilde{E}_n(t+\kappa)|M_n = p]$ and

$$\begin{aligned} &\mathbb{P}\left[E_n(t) \cap \left\{v_{n,x}|R_n^{(\gamma)}(x)| \leq \kappa\right\} |M_n = p\right] \\ &\geq \mathbb{P}\left[\widetilde{E}_n(t-\kappa) \cap \left\{v_{n,x}|R_n^{(\gamma)}(x)| \leq \kappa\right\} |M_n = p\right] \\ &\geq \mathbb{P}\left[\widetilde{E}_n(t-\kappa)|M_n = p\right] - \mathbb{P}\left[v_{n,x}|R_n^{(\gamma)}(x)| > \kappa|M_n = p\right]. \end{aligned}$$

Using the inequality $|x| \leq |a| + |b|$ which holds for all $x \in [a,b]$, it is then clear that for all $\kappa > 0$,

$$\begin{aligned} T_{n,x}^{(1)} \leq &\sup_{p \in I_n} \left|\mathbb{P}\left[\widetilde{E}_n(t+\kappa)|M_n = p\right] - H_x(t+\kappa)\right| \\ &+ \sup_{p \in I_n} \left|\mathbb{P}\left[\widetilde{E}_n(t-\kappa)|M_n = p\right] - H_x(t-\kappa)\right| \\ &+ |H_x(t) - H_x(t+\kappa)| + |H_x(t) - H_x(t-\kappa)| + T_{n,x}^{(2)}. \end{aligned}$$

Since H_x is continuous, we can actually choose $\kappa > 0$ so small that

$$|H_x(t) - H_x(t+\kappa)| \le \frac{\varepsilon}{8} \text{ and } |H_x(t) - H_x(t-\kappa)| \le \frac{\varepsilon}{8}$$

and therefore

$$\begin{aligned} T_{n,x}^{(1)} \le\ & \sup_{p \in I_n} \left| \mathbb{P}\left[\widetilde{E}_n(t+\kappa)|M_n = p\right] - H_x(t+\kappa)\right| \\ & + \sup_{p \in I_n} \left| \mathbb{P}\left[\widetilde{E}_n(t-\kappa)|M_n = p\right] - H_x(t-\kappa)\right| + T_{n,x}^{(2)} + \frac{\varepsilon}{4}. \end{aligned} \tag{6.13}$$

We now focus on the two first terms in the left-hand side of the previous inequality. From Lemma A.4, the distribution of $\widetilde{\gamma}_n(x, u_{n,x})$ given $M_n = p$ is that of

$$\overline{\gamma}_p(x, u_{n,x}) := \frac{1}{pu_{n,x}} \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} \Phi\left(\frac{i}{pu_{n,x}}\right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)}.$$

Hence, for all $s \in \mathbb{R}$ and $p \in I_n$, $\mathbb{P}[\widetilde{E}_n(s)|M_n = p] = \mathbb{P}[v_{n,x}(\overline{\gamma}_p(x, u_{n,x}) - \gamma(x)) \le s]$. Furthermore, for n large enough we have

$$p/2 \le \frac{p}{1 + [m_x(h)]^{[a(x)/4]-1/2}} \le m_x(h) \le \frac{p}{1 - [m_x(h)]^{[a(x)/4]-1/2}} \le 2p$$

for all $p \in I_n$, so that for n large enough:

$$\xi^{(+)}(p) \le m_x(h) \le \xi^{(-)}(p), \tag{6.14}$$

with $\xi^{(+)}(p) := p[1 + (2p)^{[a(x)/4]-1/2}]^{-1}$ and $\xi^{(-)}(p) := p[1 - (p/2)^{[a(x)/4]-1/2}]^{-1}$. Under our assumptions on the sequence $u_{n,x}$, the previous inequalities lead to $k_1(p) \le pu_{n,x} \le k_2(p)$ where $k_1(p) := p[\xi^{(-)}(p)]^{-a(x)}[1 + \varphi_1(\xi^{(-)}(p))]$ and $k_2(p) := p[\xi^{(+)}(p)]^{-a(x)}[1 + \varphi_2(\xi^{(+)}(p))]$. Since Φ is a nonincreasing function on $(0, 1)$, we then get that:

$$\begin{aligned} \overline{\gamma}_p(x, u_{n,x}) \ &\le \ \frac{1}{k_1(p)} \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} \Phi\left(\frac{i}{\lfloor k_2(p) \rfloor + 1}\right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)} \\ &\le \ \frac{1}{k_1(p)} \sum_{i=1}^{\lfloor k_2(p) \rfloor + 1} \Phi\left(\frac{i}{\lfloor k_2(p) \rfloor + 1}\right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)} \\ &= \ \breve{\gamma}_p(x, k_1(p), k_2(p)) \end{aligned}$$

with

$$\breve{\gamma}_p(x, k, k') := \frac{1}{k} \sum_{i=1}^{\lfloor k' \rfloor} \Phi\left(\frac{i}{\lfloor k' \rfloor + 1}\right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)}. \tag{6.15}$$

A similar lower bound applies and thus $\breve{\gamma}_p(x, k_2(p), k_1(p) - 1) \le \overline{\gamma}_p(x, u_{n,x}) \le \breve{\gamma}_p(x, k_1(p), k_2(p))$ for all $p \in I_n$. As a first conclusion, using the inequality $|x| \le$

$|a|+|b|$ which holds for all $x \in [a,b]$, we have shown that for all $s \in \mathbb{R}$,

$$\begin{aligned}
&\sup_{p\in I_n}\left|\mathbb{P}\left[\widetilde{E}_n(s)|M_n=p\right]-H_x(s)\right| \\
\le\; &\sup_{p\in I_n}\left|\mathbb{P}\left[v_{n,x}(\breve{\gamma}_p(x,k_1(p),k_2(p))-\gamma(x))\le s\right]-H_x(s)\right| \\
&+\sup_{p\in I_n}\left|\mathbb{P}\left[v_{n,x}(\breve{\gamma}_p(x,k_2(p),k_1(p)-1)-\gamma(x))\le s\right]-H_x(s)\right|.
\end{aligned}$$

Since from (6.14), $[\xi^{(+)}(p)]^{(1-a(x))/2} \le v_{n,x} \le [\xi^{(-)}(p)]^{(1-a(x))/2}$ for all $p \in I_n$ and since by assumption on the φ_i,

$$\begin{aligned}
\frac{k_1(p)}{k_2(p)} &= 1+\mathcal{O}\left(p^{[a(x)/4]-1/2}\right)+\mathcal{O}\left(\varphi_1(\xi^{(+)}(p))\right)+\mathcal{O}\left(\varphi_2(\xi^{(+)}(p))\right) \\
&= 1+\mathrm{o}(p^{(a(x)-1)/2}),
\end{aligned}$$

one can apply Lemmas A.6 and A.7 to show that for n large enough

$$\begin{aligned}
&\sup_{p\in I_n}\left|\mathbb{P}\left[\widetilde{E}_n(t+\kappa)|M_n=p\right]-H_x(t+\kappa)\right| \\
&+\sup_{p\in I_n}\left|\mathbb{P}\left[\widetilde{E}_n(t-\kappa)|M_n=p\right]-H_x(t-\kappa)\right| \le \frac{\varepsilon}{2}.
\end{aligned} \tag{6.16}$$

It remains to study the term $T^{(2)}_{n,x}$. Lemma A.4 entails that

$$T^{(2)}_{n,x} \le \sup_{p\in I_n}\mathbb{P}\left[2v_{n,x}\omega(U_{1,p},U_{p,p},x,h)\int_0^{u_{n,x}}|\Psi(\alpha,u_{n,x})|d\alpha>\kappa\right].$$

From condition (H_Ψ),

$$\limsup_{u\downarrow 0}\int_0^u|\Psi(\alpha,u)|d\alpha=C<\infty$$

and thus for n large enough, using (6.6):

$$\begin{aligned}
T^{(2)}_{n,x} &\le \sup_{p\in I_n}\mathbb{P}\left[v_{n,x}\omega(U_{1,p},U_{p,p},x,h)>\frac{\kappa}{4C}\right] \\
&\le 2\left(1-[1-[m_x(h)]^{-1-\delta}]^{2m_x(h)}\right)\le\frac{\varepsilon}{8}.
\end{aligned} \tag{6.17}$$

Collecting (6.9), (6.12), (6.13), (6.16) and (6.17) concludes the proof. □

APPENDIX

The first lemma is dedicated to the statistics $T_n(x)$ and $R_n^{(Q)}(x)$ defined in the proof of Proposition 6.1, equations (6.3) and (6.4).

Lemma A.1. *Let $\{U_i,\ i \geq 1\}$ be independent standard uniform random variables. For $x \in \mathcal{E}$ such that $m_x(h) > 0$, the conditional distribution of $T_n(x)$ given $M_x(h) = p$ is that of*

$$\underline{T}_p(x) := \sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor+1,p}|x)}{Q(\alpha|x)} - 1 \right|$$

and, given $M_x(h) = p$, $R_n^{(Q)}(x)$ is bounded from above by

$$\omega(U_{1,p}, U_{p,p}, x, h) \exp[\omega(U_{1,p}, U_{p,p}, x, h)] \left(1 + \underline{T}_p(x)\right).$$

Proof: Recall the notation $M_n := M_x(h)$ and $V_i := S(Y_i^*|X_i^*)$. First, given $M_n = p$, equation (6.2) entails that $\{V_i,\ 1 \leq i \leq M_n\}|\{M_n = p\} \stackrel{d}{=} \{U_i,\ 1 \leq i \leq p\}$ where $U_1, ..., U_p$ are independent standard uniform variables. It thus holds that

$$\left\{Q(V_{\lfloor \alpha M_n \rfloor+1,M_n}|x),\ \alpha \in [0,1)\right\} |\{M_n = p\} \stackrel{d}{=} \left\{Q(U_{\lfloor p\alpha \rfloor+1,p}|x),\ \alpha \in [0,1)\right\}.$$

As a direct consequence

$$T_n(x)|\{M_n = p\} \stackrel{d}{=} \underline{T}_p(x). \tag{A.1}$$

Let us now focus on the term $R_n^{(Q)}(x)$. Since $Q(\cdot|x)$ is continuous and decreasing, one has, for $i = 1, ..., M_n$,

$$\begin{aligned} \log Q(V_i|x) - \omega(V_{1,M_n}, V_{M_n,M_n}, x, h) &\leq \log Y_i^* = \log Q(V_i|X_i^*) \\ &\leq \log Q(V_i|x) + \omega(V_{1,M_n}, V_{M_n,M_n}, x, h). \end{aligned}$$

It follows from Lemma 1 in Gardes and Stupfler [22] that for all $i \in \{1, ..., M_n\}$,

$$\left|\log Y^*_{M_n-i+1,M_n} - \log Q(V_{i,M_n}|x)\right| \leq \omega(V_{1,M_n}, V_{M_n,M_n}, x, h). \tag{A.2}$$

Since $\widehat{Q}_n(\alpha|x) = Y^*_{M_n-i+1,M_n}$ for all $\alpha \in [(i-1)/M_n, i/M_n)$, the mean value theorem leads to

$$\begin{aligned} &\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x)}{Q(V_{\lfloor \alpha M_n \rfloor+1,M_n}|x)} - 1 \right| \\ &\leq \omega(V_{1,M_n}, V_{M_n,M_n}, x, h) \exp\left[\omega(V_{1,M_n}, V_{M_n,M_n}, x, h)\right]. \end{aligned}$$

Hence,

$$\begin{aligned} R_n^{(Q)}(x) &= \sup_{\alpha\in[\tau u_{n,x},u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x)}{Q(V_{\lfloor \alpha M_n\rfloor+1,M_n}|x)} - 1 \right| \left| \frac{Q(V_{\lfloor \alpha M_n\rfloor+1,M_n}|x)}{Q(\alpha|x)} \right| \\ &\le \omega(V_{1,M_n}, V_{M_n,M_n}, x, h) \exp\left[\omega(V_{1,M_n}, V_{M_n,M_n}, x, h)\right](1+T_n(x)). \end{aligned}$$

Use finally (6.2) and (A.1) to complete the proof. □

The next lemma examines the convergence of $T_n(x)$, defined in the above lemma, given $M_x(h)$.

Lemma A.2. *Let $U_1, ..., U_p$ be independent standard uniform variables. Assume that (2.3) and (H_{SO}) hold. If $a(x)\in(0,1)$ is such that $p^{1-a(x)}\Delta^2(p^{a(x)}|x) \to \lambda\in\mathbb{R}$ as $p\to\infty$ then, for all $d_1, d_2 > 0$ with $d_1 < d_2$, we have:*

$$p^{(1-a(x))/2} \sup_{\alpha\in[d_1p^{-a(x)},d_2p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha\rfloor+1,p}|x)}{Q(\alpha|x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(1).$$

Proof: Recall that (H_{SO}) entails that (3.1) holds. Then, one can apply [27, Theorem 2.4.8] to the independent random variables $\{Q(U_i|x),\ i=1,...,p\}$ distributed from the conditional survival function $S(\cdot|x)$: because

$$\inf_{\alpha\in[d_1p^{-a(x)},d_2p^{-a(x)}]} \frac{\alpha}{d_2p^{-a(x)}} = \frac{d_1}{d_2} > 0,$$

it holds that

$$p^{(1-a(x))/2} \sup_{\alpha\in[d_1p^{-a(x)},d_2p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha\rfloor+1,p}|x)}{Q(d_2p^{-a(x)}|x)} - \left(\frac{\alpha p^{a(x)}}{d_2}\right)^{-\gamma(x)} \right| = \mathcal{O}_{\mathbb{P}}(1). \tag{A.3}$$

Since (3.1) must in fact hold locally uniformly in $z > 0$ (see [27, Theorem B.2.9]) and $[d_1, d_2]$ is a compact interval, it is clear that

$$p^{(1-a(x))/2} \sup_{\alpha\in[d_1p^{-a(x)},d_2p^{-a(x)}]} \left| \frac{Q(\alpha|x)}{Q(d_2p^{-a(x)}|x)} - \left(\frac{\alpha p^{a(x)}}{d_2}\right)^{-\gamma(x)} \right| = \mathcal{O}(1). \tag{A.4}$$

Combine (A.3) and (A.4) to conclude the proof. □

Lemma A.3 below controls a bias term appearing in the proof of Theorem 3.1.

Lemma A.3. *Assume that conditions (2.3) and (H_{SO}) are satisfied. If $m_x(h)\to\infty$ and $\beta_n/u_{n,x}\to 0$ we have that:*

$$\sup_{\alpha\in[\tau u_{n,x},u_{n,x}]} \left| \frac{Q(\alpha|x)}{Q(\beta_n|x)} \left(\frac{\alpha}{\beta_n}\right)^{\gamma(x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}\left(\Delta(u_{n,x}^{-1}|x)\right).$$

Proof: Recall

$$\alpha^{\gamma(x)} Q(\alpha|x) = c(x) \exp\left(\int_1^{\alpha^{-1}} \frac{\Delta(v|x)}{v} dv \right),$$

and therefore

$$\frac{Q(\alpha|x)}{Q(\beta_n|x)} \left(\frac{\alpha}{\beta_n} \right)^{\gamma(x)} = \exp\left(\int_{\beta_n^{-1}}^{\alpha^{-1}} \frac{\Delta(v|x)}{v} dv \right).$$

Furthermore, since $\alpha \leq u_{n,x}$,

$$\left| \int_{\beta_n^{-1}}^{\alpha^{-1}} \frac{\Delta(v|x)}{v} dv \right| \leq |\Delta(u_{n,x}^{-1}|x)| \int_1^{\infty} \left| \frac{\Delta(y u_{n,x}^{-1}|x)}{\Delta(u_{n,x}^{-1}|x)} \right| \frac{dy}{y}.$$

As the function $y \mapsto y^{-1}\Delta(y|x)$ is regularly varying with index $\rho(x) - 1 < -1$, we may write, according to [4, Theorem 1.5.2],

$$\left| \int_{\beta_n^{-1}}^{\alpha^{-1}} \frac{\Delta(v|x)}{v} dv \right| \leq 2|\Delta(u_{n,x}^{-1}|x)| \int_1^{\infty} y^{\rho(x)-1} dy = -\frac{2}{\rho(x)} |\Delta(u_{n,x}^{-1}|x)|.$$

Since the right-hand side converges to 0 and does not depend on α, it follows by a Taylor expansion of the exponential function that

$$\sup_{\alpha \in (\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(\alpha|x)}{Q(\beta_n|x)} \left(\frac{\alpha}{\beta_n} \right)^{\gamma(x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}\left(\Delta(u_{n,x}^{-1}|x) \right),$$

which is the required conclusion. □

The next result is dedicated to the statistics $\widetilde{\gamma}_n(x, u_{n,x})$ and $R_n^{(\gamma)}(x)$ introduced in the proof of Theorem 3.2, equation (6.10).

Lemma A.4. *Let U_i, $i \geq 1$ be independent standard uniform random variables. For any $x \in \mathcal{E}$ such that $m_x(h) > 0$, the conditional distribution of $\widetilde{\gamma}_n(x, u_{n,x})$ given $M_x(h) = p$ is that of*

$$\overline{\gamma}_p(x, u_{n,x}) = \frac{1}{p u_{n,x}} \sum_{i=1}^{\lfloor p u_{n,x} \rfloor} \Phi\left(\frac{i}{p u_{n,x}} \right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)},$$

and given $M_x(h) = p$, $R_n^{(\gamma)}(x)$ is bounded from above by

$$2\omega(U_{1,p}, U_{p,p}, x, h) \int_0^{u_{n,x}} |\Psi(\alpha, u_{n,x})| d\alpha.$$

Proof: Set again $M_n = M_x(h)$. Equation (6.2) entails that the conditional distribution of $\widehat{\gamma}_n(x, u_{n,x})$ given $M_n = p$ is that of

$$\begin{aligned} & \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} W_{i,n}(u_{n,x}, p) \log \frac{Q(U_{i,p}|x)}{Q(U_{\lfloor pu_{n,x} \rfloor+1,p}|x)} \\ = & \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} W_{i,n}(u_{n,x}, p) \sum_{j=i}^{\lfloor pu_{n,x} \rfloor} \log \frac{Q(U_{j,p}|x)}{Q(U_{j+1,p}|x)}, \end{aligned}$$

where $\{U_i,\ i \geq 1\}$ are independent standard uniform random variables, and this is equal to $\overline{\gamma}_p(x, u_{n,x})$ by switching the summation order and using assumption (H_Ψ). Now, since $\widehat{Q}_n(\alpha|x) = Y^*_{M_n-i+1,M_n}$ for all $\alpha \in [(i-1)/M_n, i/M_n)$, one has

$$\widehat{\gamma}_n(x, u_{n,x}) = \sum_{i=1}^{\lfloor u_{n,x} M_n \rfloor} W_{i,n}(u_{n,x}, M_n) \log \frac{Y^*_{M_n-i+1,M_n}}{Y^*_{M_n-\lfloor u_{n,x} M_n \rfloor, M_n}},$$

where $W_{i,n}(u_{n,x}, M_n)$ was defined in (6.11). Hence the identity

$$R_n^{(\gamma)}(x) = \sum_{i=1}^{\lfloor u_{n,x} M_n \rfloor} W_{i,n}(u_{n,x}, M_n) \log \left[\frac{Q(V_{\lfloor u_{n,x} M_n \rfloor+1,M_n}|x)}{Q(V_{i,M_n}|x)} \frac{Y^*_{M_n-i+1,M_n}}{Y^*_{M_n-\lfloor u_{n,x} M_n \rfloor, M_n}} \right].$$

Using the bound (A.2) yields to

$$\begin{aligned} R_n^{(\gamma)}(x) & \leq 2\omega(V_{1,M_n}, V_{M_n,M_n}, x, h) \sum_{i=1}^{\lfloor u_{n,x} M_n \rfloor} |W_{i,n}(u_{n,x}, M_n)| \\ & \leq 2\omega(V_{1,M_n}, V_{M_n,M_n}, x, h) \int_0^{u_{n,x}} |\Psi(\alpha, u_{n,x})| d\alpha. \end{aligned}$$

Using equation (6.2) completes the proof. □

Our next result studies some particular Riemann sums. It shall prove useful when examining the convergence of $\widetilde{\gamma}_n(x, u_{n,x})$ given $M_x(h)$, see Lemma A.6.

Lemma A.5. *Let f be an integrable function on $(0,1)$. Assume that f is nonnegative and nonincreasing. For any nonnegative continuous function g on $[0,1]$ we have that:*

$$\lim_{m\to\infty} \frac{1}{m-1} \sum_{i=1}^{m-1} f\left(\frac{i}{m}\right) g\left(\frac{i}{m}\right) = \int_0^1 f(t) g(t) dt.$$

If moreover f is square-integrable then:

$$\lim_{m\to\infty} \sqrt{m} \left| \frac{1}{m-1} \sum_{i=1}^{m-1} f\left(\frac{i}{m}\right) - \int_0^1 f(t) dt \right| = 0.$$

Proof: To prove the first statement, it suffices to show that $|S_m(f,g) - S(f,g)| \to 0$ as $m \to \infty$ where

$$S_m(f,g) := \frac{1}{m}\sum_{i=1}^{m-1} f\left(\frac{i}{m}\right) g\left(\frac{i}{m}\right) \text{ and } S(f,g) := \int_0^1 f(t)g(t)dt.$$

Note first that:

$$\begin{aligned}|S(f,g) - S_m(f,g)| \le\ & \sum_{i=1}^{m-1}\int_{(i-1)/m}^{i/m}\left|f(t)g(t) - f\left(\frac{i}{m}\right)g\left(\frac{i}{m}\right)\right|dt \\ & + \int_{(m-1)/m}^{1} f(t)g(t)dt.\end{aligned}$$

Since g is nonnegative on $[0,1]$ and f is nonincreasing, it is straightforward that for all $t \in [(i-1)/m, i/m)$

$$\begin{aligned}|f(t)g(t) - f(i/m)g(i/m)| \le\ & f(t) \sup_{|s-s'|\le 1/m} |g(s) - g(s')| \\ & + \|g\|_\infty (f(t) - f(i/m)),\end{aligned}$$

where $\|g\|_\infty$ is the finite supremum of g on $[0,1]$. The fact that f is nonincreasing yields $f(t) - f(i/m) \le f((i-1)/m) - f(i/m)$ for all $i = 2, ..., m$ and thus the previous inequality leads to

$$\begin{aligned}|S(f,g) - S_m(f,g)| \le\ & \int_0^1 f(t)dt \sup_{|s-s'|\le 1/m} |g(s) - g(s')| \\ & + \|g\|_\infty \left(\int_0^{1/m} f(t)dt - \frac{f(1)}{m}\right) \\ & + \|g\|_\infty \int_{(m-1)/m}^{1} f(t)dt \to 0 \end{aligned} \tag{A.5}$$

by the uniform continuity of g on $[0,1]$ and the fact that f is an integrable function. This proves the first statement of the result. To prove the second one, remark that:

$$\sqrt{m}\left|\frac{1}{m-1}\sum_{i=1}^{m-1} f\left(\frac{i}{m}\right) - \int_0^1 f(t)dt\right| \le \frac{\sqrt{m}}{m-1}S_m(f,1) + \sqrt{m}|S(f,1) - S_m(f,1)|.$$

Using the first statement with $g = 1$ entails that the first term of the left-hand side converges to 0 as $m \to \infty$. Now, taking $g = 1$ in (A.5) leads to

$$\sqrt{m}|S(f,1) - S_m(f,1)| \le \sqrt{m}\int_0^{1/m} f(t)dt + \sqrt{m}\int_{(m-1)/m}^{1} f(t)dt.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} & \sqrt{m}\int_0^{1/m} f(t)dt \le \left(\int_0^{1/m} f^2(t)dt\right)^{1/2} \to 0 \\ \text{and} \quad & \sqrt{m}\int_{(m-1)/m}^{1} f(t)dt \le \left(\int_{(m-1)/m}^{1} f^2(t)dt\right)^{1/2} \to 0 \end{aligned}$$

since f^2 is integrable on $(0,1)$. The proof is complete. □

The next lemma establishes the asymptotic normality of the random variable $\overline{\gamma}_p(x,k,k')$ introduced in the proof of Theorem 3.2, equation (6.15).

Lemma A.6. *Assume that conditions (2.3), (H_{SO}) and (H_Ψ) are satisfied. Let $k(p)$ and $k'(p)$ be two sequences satisfying, for some $a(x)\in(0,1)$, $p^{a(x)-1}k(p)\to 1$ and $p^{(1-a(x))/2}[k(p)/k'(p)-1]\to 0$ as $p\to\infty$. Let $U_1,...,U_p$ be independent standard uniform random variables. If $p^{1-a(x)}\Delta^2(p^{a(x)}|x)\to\lambda(x)\in\mathbb{R}$, then the random variable*

$$\breve{\gamma}_p(x,k(p),k'(p)) := \frac{1}{k(p)}\sum_{i=1}^{\lfloor k'(p)\rfloor}\Phi\left(\frac{i}{\lfloor k'(p)\rfloor+1}\right) i\log\frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)}$$

is such that $p^{(1-a(x))/2}(\breve{\gamma}_p(x,k(p),k'(p))-\gamma(x))$ converges in distribution to a normal distribution with mean $\lambda(x)\int_0^1\Phi(\alpha)\alpha^{-\rho(x)}d\alpha$ and variance $\gamma^2(x)\int_0^1\Phi^2(\alpha)d\alpha$.

Proof: For the sake of brevity, let $\breve{\gamma}_p(x) := \breve{\gamma}_p(x,k(p),k'(p))$. Let $v_p := p^{(1-a(x))/2}$ and for $j\in\{1,...,k'(p)\}$

$$\tilde{\Delta}_j(p|x) := \Delta\left(\frac{p+1}{\lfloor k'(p)\rfloor+1}\middle| x\right)\left(\frac{j}{\lfloor k'(p)\rfloor+1}\right)^{-\rho(x)}$$

Under conditions (2.3), (H_{SO}) and (H_Ψ), one can apply Theorem 3.1 in Beirlant *et al.* [1] to prove that

$$v_p\left\{\frac{k(p)}{\lfloor k'(p)\rfloor}\breve{\gamma}_p(x) - \frac{1}{\lfloor k'(p)\rfloor}\sum_{j=1}^{\lfloor k'(p)\rfloor}\Phi\left(\frac{j}{\lfloor k'(p)\rfloor+1}\right)\left[\gamma(x)+\tilde{\Delta}_j(p|x)\right]\right\}$$

converges to a centered normal distribution with variance $\sigma_\Phi^2 := \gamma^2(x)\int_0^1\Phi^2(\alpha)d\alpha$. As a direct consequence of Lemma A.5, the previous convergence can be rewritten

$$v_p\left[\frac{k(p)}{\lfloor k'(p)\rfloor}\breve{\gamma}_p(x)-\gamma(x)\right]\xrightarrow{d}\mathcal{N}\left(\lambda(x)\int_0^1\Phi(\alpha)\alpha^{-\rho(x)}d\alpha,\sigma_\Phi^2\right). \tag{A.6}$$

Finally, since

$$\begin{aligned} v_p\left[\breve{\gamma}_p(x)-\gamma(x)\right] &= v_p\left(\frac{\lfloor k'(p)\rfloor}{k(p)}-1\right)\frac{k(p)}{\lfloor k'(p)\rfloor}\breve{\gamma}_p(x) \\ &\quad + v_p\left[\frac{k(p)}{\lfloor k'(p)\rfloor}\breve{\gamma}_p(x)-\gamma(x)\right], \end{aligned}$$

a combination of convergence (A.6) and of the fact that $v_p[k(p)/k'(p)-1]\to 0$ as $p\to\infty$ concludes the proof. □

The final lemma is a technical tool we shall need to bridge the gap between the convergence of our estimators and that of their conditional versions.

Lemma A.7. *Let $\{Z_p,\ p \in \mathbb{N}\}$ be a sequence of random variables such that for all $t \in \mathbb{R}$, $\mathbb{P}(Z_p \le t) \to H(t)$ where H is a continuous cumulative distribution function. For $n \in \mathbb{N}$, let $I_n := [u_n, v_n]$ where $u_n \to \infty$ as $n \to \infty$ and let (a_n) be a sequence such that there exist two functions ξ_1 and ξ_2 converging to 1 at infinity with*

$$\sup_{p\in I_n} \frac{\xi_1(p)}{a_n} \le 1 \le \inf_{p\in I_n} \frac{\xi_2(p)}{a_n}.$$

Then, for all $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \sup_{p\in I_n} |\mathbb{P}(a_n Z_p \le t) - H(t)| = 0.$$

Proof: We start by remarking that for all $\kappa > 0$,

$$\sup_{p\in I_n} |\mathbb{P}(a_n Z_p \le t) - H(t)| \le D_{n,p} + \sup_{p\in I_n} \mathbb{P}(|(a_n - 1)Z_p| > \kappa),$$

where

$$D_{n,p} := \sup_{p\in I_n} |\mathbb{P}(\{a_n Z_p \le t\} \cap \{|(a_n - 1)Z_p| \le \kappa\}) - H(t)|.$$

Now, since H is continuous, there exists $\kappa > 0$ such that for n large enough,

$$|H(t) - H(t+\kappa)| \le \frac{\varepsilon}{6} \text{ and } |H(t) - H(t-\kappa)| \le \frac{\varepsilon}{6}.$$

Furthermore, since $\xi_1(p) \le a_n \le \xi_2(p)$ for any $p \in I_n$, using the inequality $|x| \le |a| + |b|$ which holds for all $x \in [a,b]$, one has for all $p \in I_n$ that $|a_n - 1| \le |\xi_1(p) - 1| + |\xi_2(p) - 1|$; besides, since $Z_p = \mathcal{O}_{\mathbb{P}}(1)$ and ξ_1, ξ_2 converge to 1 at infinity, we have $|\xi_1(p) - 1|Z_p = \mathrm{o}_{\mathbb{P}}(1)$ and $|\xi_2(p) - 1|Z_p = \mathrm{o}_{\mathbb{P}}(1)$. Therefore, for all $\varepsilon > 0$,

$$\sup_{p\in I_n} \mathbb{P}(|(a_n - 1)Z_p| > \kappa) \le \sup_{p\in I_n} \mathbb{P}(|\xi_1(p) - 1||Z_p| + |\xi_2(p) - 1||Z_p| > \kappa) \le \frac{\varepsilon}{6}$$

for n large enough. Now remark that for all $p \in I_n$, $\mathbb{P}(\{a_n Z_p \le t\} \cap \{|(a_n - 1)Z_p| \le \kappa\}) \le \mathbb{P}(Z_p \le t + \kappa)$ and that

$$\begin{aligned}
\mathbb{P}(\{a_n Z_p \le t\} \cap \{|(a_n - 1)Z_p| \le \kappa\}) &\ge \mathbb{P}(\{Z_p \le t - \kappa\} \cap \{|(a_n - 1)Z_p| \le \kappa\}) \\
&\ge \mathbb{P}(Z_p \le t - \kappa) - \mathbb{P}(|(a_n - 1)Z_p| > \kappa).
\end{aligned}$$

Hence, for all $\kappa > 0$, the inequality:

$$\begin{aligned}
D_{n,p} &\le \sup_{p\in I_n} |\mathbb{P}(Z_p \le t + \kappa) - H(t+\kappa)| + \sup_{p\in I_n} |\mathbb{P}(Z_p \le t - \kappa) - H(t-\kappa)| \\
&\quad + |H(t) - H(t+\kappa)| + |H(t) - H(t-\kappa)| + \frac{\varepsilon}{6} \\
&\le \sup_{p\in I_n} |\mathbb{P}(Z_p \le t + \kappa) - H(t+\kappa)| + \sup_{p\in I_n} |\mathbb{P}(Z_p \le t - \kappa) - H(t-\kappa)| + \frac{\varepsilon}{2}.
\end{aligned}$$

By assumption, for n large enough:

$$\sup_{p\in I_n} |\mathbb{P}(Z_p \le t + \kappa) - H(t+\kappa)| \le \frac{\varepsilon}{6} \text{ and } \sup_{p\in I_n} |\mathbb{P}(Z_p \le t - \kappa) - H(t-\kappa)| \le \frac{\varepsilon}{6}.$$

It is now straightforward to conclude the proof. $\square$

ACKNOWLEDGMENTS

We acknowledge the valuable suggestions from Associate Editor and referee.

REFERENCES

[1] Beirlant, J.; Dierckx, G.; Guillou, A. and Stărică, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**(2), 157–180.

[2] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J.L. (2004). *Statistics of extremes – Theory and applications*, Wiley Series in Probability and Statistics.

[3] Billingsley, P. (1979). *Probability and measure*, John Wiley and Sons.

[4] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge, U.K., Cambridge University Press.

[5] Cai, J.-J.; de Haan, L. and Zhou, C. (2013). Bias correction in extreme value statistics with index around zero, *Extremes*, **16**, 173–201.

[6] 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.

[7] Chernozhukov, V. (2005). Extremal quantile regression, *The Annals of Statistics*, **33**(2), 806–839.

[8] Chernozhukov, V. and Du, S. (2008). Extremal quantiles and Value-at-Risk, *The New Palgrave Dictionary of Economics*, second edition.

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

[10] Daouia, A.; Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, **19 5B**, 2557–2589.

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

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

[13] Dekkers, A.L.M.; Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics*, **17**(4), 1833–1855.

[14] Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data, *Computational Statistics*, **17**, 545–564.

[15] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer.

[16] Ferraty, F.; Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects, *Australian and New Zealand Journal of Statististics*, **49**(3), 267–286.

[17] Fraga Alves, M.I.; Gomes, M.I. and De Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**, 193–213.

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

[19] 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.

[20] 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.

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

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

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

[24] Goegebeur, Y.; Guillou, A. and Osmann, M. (2017). A local moment type estimator for an extreme quantile in regression with random covariates, *Communications in Statistics – Theory and Methods*, **46**(1), 319–343.

[25] Goegebeur, Y.; Guillou, A. and Schorgen, A. (2014). Nonparametric regression estimation of conditional tails — the random covariate case, *Statistics*, **48**(4), 732–755.

[26] Goegebeur, Y.; Guillou, A. and Stupfler, G. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de l'Institut Henri Poincaré (B): Probability and Statistics*, **51**(3), 1190–1213.

[27] de Haan, L. and Ferreira, A. (2006). *Extreme value theory: An introduction*, Springer.

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

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

[30] Jiřina, M. (1959). On regular conditional probabilities, *Czechoslovak Mathematical Journal*, **9**, 445–451.

[31] Kratz, M. and Resnick, S. (1996). The QQ-estimator and heavy tails, *Stochastic Models*, **12**, 699–724.

[32] Li, Q.; Lin, J. and Racine, J.S. (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions, *Journal of Business & Economic Statistics*, **31**(1), 57–65.

[33] Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119–131.

[34] Resnick, S. and Stărică, C. (1997). Smoothing the Hill estimator, *Advances in Applied Probability*, **29**, 271–293.

[35] Schultze, J. and Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353–372.

[36] 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.

[37] Stupfler, G. (2013). A moment estimator for the conditional extreme-value index, *Electronic Journal of Statistics*, **7**, 2298–2343.

[38] Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring, *Journal of Multivariate Analysis*, **144**, 1–24.

[39] Wang, H.J.; Li, D. and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions, *Journal of the American Statistical Association*, **107**(500), 1453–1464.

[40] Wang, H.J. and Li, D. (2013). Estimation of extreme conditional quantiles through power transformation, *Journal of the American Statistical Association*, **108**(503), 1062–1074.

[41] Wang, H. and Tsai, C.L. (2009). Tail index regression, *Journal of the American Statistical Association*, **104**(487), 1233–1240.

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