Annals of the Institute of Statistical Mathematics (2020) 72:915–943
https://doi.org/10.1007/s10463-019-00713-7

Estimation of extreme conditional quantiles under a general tail-first-order condition

Laurent Gardes[1] **· Armelle Guillou**[1] **· Claire Roman**[1]

Received: 19 June 2018 / Revised: 25 February 2019 / Published online: 9 April 2019
© The Institute of Statistical Mathematics, Tokyo 2019

Abstract
We consider the estimation of an extreme conditional quantile. In a first part, we propose a new tail condition in order to establish the asymptotic distribution of an extreme conditional quantile estimator. Next, a general class of estimators is introduced, which encompasses, among others, kernel or nearest neighbors types of estimators. A unified theorem of the asymptotic normality for this general class of estimators is provided under the new tail condition and illustrated on the different well-known examples. A comparison between different estimators belonging to this class is provided on a small simulation study and illustrated on a real dataset on earthquake magnitudes.

Keywords Extreme quantile · Local estimation · Asymptotic normality

1 Introduction

To describe the dependence between a real-valued random variable Y and an explanatory random vector X of dimension $p \in \mathbb{N} \setminus \{0\}$, different approaches can be used. The most common one is perhaps provided by the conditional mean $m(X) := \mathbb{E}(Y|X)$, which gives information on the central part of the conditional distribution. However, depending on the applications in mind, it can be also of interest to consider a conditional quantile instead of $m(X)$ (e.g., median or quartile). To be more specific, denoting by $S(\cdot|x_0) := \mathbb{P}(Y > \cdot|X = x_0)$ the conditional survival function of Y given $\{X = x_0\}$ for some $x_0 \in \mathbb{R}^p$ in the support of X, the conditional quantile of level $\alpha \in [0, 1]$ of Y given $\{X = x_0\}$ is $Q(\alpha|x_0) := S^{\leftarrow}(\alpha|x_0) = \inf\{y \in \mathbb{R};\ S(y|x_0) \leq \alpha\}$ with

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-019-00713-7) contains supplementary material, which is available to authorized users.

✉ Armelle Guillou
armelle.guillou@math.unistra.fr

[1] Institut de Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg & CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France

the convention $\inf\{\emptyset\} = +\infty$. This conditional quantile presents the advantage to be more robust than the classical conditional mean.

Given n independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of (X, Y), one question of interest is of course the estimation of the conditional quantile $Q(\alpha|x_0)$ in a nonparametric way. There exist numerous estimation methods in the literature. The most common one is the *indirect* method: starting from a suitable estimator $\widehat{S}_n(\cdot|x_0)$ of $S(\cdot|x_0)$, the associated estimator of $Q(\alpha|x_0)$ is given by

$$\widehat{Q}_n(\alpha|x_0) := \widehat{S}_n^{\leftarrow}(\alpha|x_0) = \inf\{y \in \mathbb{R};\ \widehat{S}_n(y|x_0) \leq \alpha\}. \tag{1}$$

Estimator (1) is called indirect since, as pointed by Racine and Li (2017), "one estimates a conditional survival function, and then, one 'backs out' the inferred quantile via inversion."

An alternative way to estimate a conditional quantile is by using the so-called check function defined for $\alpha \in [0, 1]$ by $\rho_\alpha(v) := v[\alpha - \mathbb{I}_{(-\infty,0]}(v)]$ where for any $A \subset \mathbb{R}$, $\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise. Indeed, since the conditional quantile is also defined by

$$Q(\alpha|x_0) = \arg\min_{\tau\in[0,1]} \mathbb{E}\left[\rho_\alpha(Y - \tau)|\, X = x_0\right],$$

the estimation of $Q(\alpha|x_0)$ can be achieved by replacing the conditional expectation by a suitable estimator and then by solving the minimization problem. This method of estimation was investigated among others by Koenker and Bassett (1978), Koenker et al. (1994) and He and Ng (1999). In this paper, we focus on the so-called indirect method.

In some applications, we are interested in the tail of the conditional distribution rather than on its central part. In this case, instead of looking at the conditional quantile of level $\alpha \in [0, 1]$, we consider an extreme conditional quantile, i.e., a conditional quantile of level α_n where $\alpha_n \to 0$ as the sample size n increases.

To obtain the asymptotic distribution of an indirect conditional quantile estimator, the following two-step procedure can be used. First, we establish the asymptotic distribution of the associated conditional survival function estimator. Next, a delta-type method is used to deduce the result on the conditional quantile estimator from this first step. This requires an additional condition on the conditional survival function. When the level α is fixed, this condition is simply that $S(\cdot|x_0)$ is continuously differentiable. However, in case of an extreme level, this condition is much more complicated. In this work, we introduce a new general condition, called *tail-first-order condition*, which is the cornerstone to obtain the asymptotic distribution of any indirect conditional quantile estimator. As we will see, this condition is more flexible than the one classically used in extreme value theory.

To understand where the tail-first-order condition comes from, the main ingredients of the proof of the asymptotic normality in case of a fixed level α and of an extreme level α_n is outlined in Sect. 2. In Sect. 3, this condition is specified and illustrated on many well-known examples of conditional distributions. Section 4 is devoted to the study of a general class of extreme conditional quantile estimators. In particular, a unified theorem for the asymptotic normality is established. A simulation study is provided in Sect. 5 where several examples of estimators belonging to this class, among them, the kernel and nearest neighbors type estimators, are compared. Their performance is finally illustrated in Sect. 6 on a real dataset on earthquake magnitudes. The proofs of the main results are postponed to Sect. 7, whereas those of the technical results are postponed to the supplementary material.

2 Description of the methodology

The aim of this paper is to show the asymptotic normality of a general class of indirect type of conditional quantile estimators when the level is extreme. This requires a condition, which is not usual in the case of a fixed level α. To understand where this condition comes from we briefly start to present the simple case where the level is fixed, and then, we outline the main differences when it is assumed to be extreme, and we introduce the required condition in that context.

Case where the level is fixed—When the level α is fixed, the asymptotic distribution of (1) can be deduced from the one of the conditional survival function estimator $\widehat{S}_n(\cdot|x_0)$. More precisely, if we assume that for some $y \in \mathbb{R}$, there exists a sequence $v_n(x_0) \to \infty$ such that for all sequence $\varepsilon_n \to 0$

$$v_n(x_0)\left(\widehat{S}_n(y+\varepsilon_n|x_0) - S(y+\varepsilon_n|x_0)\right) \xrightarrow{d} \Lambda, \tag{2}$$

where Λ is some non-degenerate distribution, then if $S(\cdot|x_0)$ is a continuously differentiable function with $S[Q(\alpha|x_0)|x_0] = \alpha$

$$v_n(x_0)\left(\widehat{Q}_n(\alpha|x_0) - Q(\alpha|x_0)\right) \xrightarrow{d} \frac{1}{f(Q(\alpha|x_0)|x_0)}\Lambda, \tag{3}$$

where $f(\cdot|x_0)$ is the probability density function of Y given $X = x_0$ with $f(Q(\alpha|x_0)|x_0) \neq 0$. The proof of (3) is based on the following remark: for all $z \in \mathbb{R}$, letting $\sigma_n(x_0) := v_n(x_0) f(Q(\alpha|x_0)|x_0)$, one has

$$\mathbb{P}\left[\sigma_n(x_0)\left(\widehat{Q}_n(\alpha|x_0) - Q(\alpha|x_0)\right) \leq z\right] = \mathbb{P}[Z_n(x_0) \leq z_n(x_0)], \tag{4}$$

where,

$$Z_n(x_0) := v_n(x_0)\left(\widehat{S}_n(Q(\alpha|x_0) + z\sigma_n^{-1}(x_0)|x_0) - S(Q(\alpha|x_0) + z\sigma_n^{-1}(x_0)|x_0)\right)$$

and $z_n(x_0) := v_n(x_0)[\alpha - S(Q(\alpha|x_0) + z\sigma_n^{-1}(x_0)|x_0)]$. From (2) with $y = Q(\alpha|x_0)$, $Z_n(x_0) \xrightarrow{d} \Lambda$ and since $S(\cdot|x_0)$ is continuously differentiable, $z_n(x_0) \to z$ as $n \to \infty$ proving (3). Note that the asymptotic distribution of indirect estimators for a fixed level α has been treated for instance by Berlinet et al. (2001).

Case of an extreme level—We consider the situation where the level of the conditional quantile is a sequence α_n where $\alpha_n \to 0$ as the sample size n increases.

Replacing the level α by a sequence α_n does not change (at least if α_n does not converge too fast to 0) the estimation procedure. We still estimate $Q(\alpha_n|x_0)$ as in (1) just by replacing α by α_n. The difference lies in the assumptions required to obtain the asymptotic distribution of $\widehat{Q}_n(\alpha_n|x_0)$. First, instead of (2), the following kind of result for the conditional survival function estimator is required: for some well-chosen sequence $y_n(x_0) \to y^*(x_0) := Q(0|x_0)$, there exists a sequence $v_n(x_0) \to \infty$ such that

$$v_n(x_0)\left(\widehat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0)\right) \xrightarrow{d} \Lambda, \tag{5}$$

for some non-degenerate distribution Λ. Of course, the sequence $v_n(x_0)$ depends on the sequence $y_n(x_0)$. Since $y^*(x_0)$ is the right endpoint, convergence (5) focuses on the asymptotic behavior of $\widehat{S}_n(\cdot|x_0)$ in the right tail of the conditional distribution. To obtain the asymptotic distribution of $\widehat{Q}_n(\alpha_n|x_0)$, we start again with (4) where α is replaced by α_n. In the extreme level case, the main difficulty is to deal with the non-random sequence $z_n(x_0)$. More specifically, assuming that $S[Q(\alpha|x_0)|x_0] = \alpha$ at least for α small enough, we need to find a general condition on the conditional distribution ensuring that for a well-chosen sequence $\sigma_n(x_0)$ and for a sequence $v_n(x_0)$ satisfying (5) with $y_n(x_0) = Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)$

$$z_n(x_0) = -\alpha_n v_n(x_0)\left[\frac{S[y_n(x_0)|x_0]}{S[Q(\alpha_n|x_0)|x_0]} - 1\right] \to z, \tag{6}$$

as $n \to \infty$ for all $z \in \mathbb{R}$. Obviously, assuming that $S(\cdot|x_0)$ is a continuously differentiable function is not relevant here and the sequence $\sigma_n(x_0)$ is not necessarily equal to $v_n(x_0)f(Q(\alpha_n|x_0)|x_0)$. Since $Q(\alpha_n|x_0) \to y^*(x_0)$, a natural general condition leading to (6) is to assume that for some open interval $I_{x_0} = I \subset \mathbb{R}$ containing 0, there exist positive functions $d_{x_0} \equiv d$ and $\Psi_{x_0} \equiv \Psi$ such that for all $t \in I$,

$$\lim_{y \uparrow y^*(x_0)} \Psi(y)\left(\frac{S[y + td(y)|x_0]}{S(y|x_0)} - 1\right) \to \phi_{x_0}^{-1}(t), \tag{7}$$

where $\phi_{x_0}^{-1} \equiv \phi^{-1}$ is the inverse of a continuous and strictly decreasing function $\phi_{x_0} \equiv \phi$ such that $\phi(t)/t \to -1$ as $t \to 0$.

Indeed, taking $\sigma_n(x_0) = \alpha_n v_n(x_0)/[\Psi(Q(\alpha_n|x_0))d(Q(\alpha_n|x_0))]$ and $t_n^{-1}(x_0) := \sigma_n(x_0)d[Q(\alpha_n|x_0)]$, we obtain

$$z_n(x_0) = -\frac{\Psi[Q(\alpha_n|x_0)]}{t_n(x_0)}\left(\frac{S[Q(\alpha_n|x_0) + zt_n(x_0)d[Q(\alpha_n|x_0)]|x_0]}{S[Q(\alpha_n|x_0)|x_0]} - 1\right).$$

Under (7) and assuming that $t_n(x_0) \to 0$, we can show that $z_n(x_0) \to z$ (see Sect. 3, Proposition 1). Next, the random sequence $Z_n(x_0)$ is treated by (5). To sum up, in the extreme level case, a natural condition on $S(\cdot|x_0)$ to establish the asymptotic distribution of the conditional quantile estimator is (7). Condition (7) is referred in what follows to as the tail-first-order condition. Under this condition and if (5) holds with

$y_n(x_0) := Q(\alpha_n|x_0) + z\sigma_n^{-1}(x_0)$, we have $\sigma_n(x_0)(\widehat{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)) \xrightarrow{d} \Lambda$. We show in Sect. 3 that this tail-first-order condition is satisfied by a larger class of conditional distributions than the one satisfying the condition classically used in extreme value theory. Note that while on the fixed level case, the rate of convergence of $\widehat{Q}_n(\alpha|x_0)$ is proportional to $v_n(x_0)$ this is no longer the case when estimating an extreme conditional quantile.

3 The tail-first-order condition

The tail-first-order condition is related to the conditional distribution of Y given $\{X = x_0\}$ for some $x_0 \in \mathbb{R}^p$ in the support of X. Since x_0 is fixed, the dependence on x_0 can be omitted. This is what we do in all this section. For a given (conditional) survival function S, we denote by $Q = S^{\leftarrow}$ the associated quantile and by $x^* = S^{\leftarrow}(0)$ the right endpoint.

Definition 1 A survival function S satisfies the tail-first-order (TFO) condition if for some open interval $I \subset \mathbb{R}$ containing 0, there exist positive functions d and Ψ such that for all $t \in I$,

$$\lim_{x \uparrow x^*} \Psi(x) \left(\frac{S[x + td(x)]}{S(x)} - 1 \right) = \phi^{-1}(t), \tag{8}$$

where ϕ^{-1} is the inverse of a continuous and strictly decreasing function $\phi : J \to I$ such that $\phi(t)/t \to -1$ as $t \to 0$.

Note that convergence (8) entails that for all $t \in I$ and for x large enough, $x + td(x) < x^*$. Consequently, the function Ψ is such that $\Psi(x)/S(x) \to \infty$ as $x \uparrow x^*$. Finally, it is easy to check that $\phi^{-1}(t)/t \to -1$ as $t \to 0$. As a consequence of Dini's theorem, we obtain the useful properties gathered in the next proposition.

Proposition 1 *If S satisfies the TFO condition, the following statements are true:*

1. *Convergence in (8) holds locally uniformly on I.*
2. *For all $t_0 \in I$,*

$$\lim_{(t,x) \to (t_0, x^*)} \frac{\Psi(x)}{t} \left(\frac{S[x + td(x)]}{S(x)} - 1 \right) = \lim_{t \to t_0} \frac{\phi^{-1}(t)}{t}.$$

We give in the next result some equivalent reformulations of the TFO condition.

Proposition 2 *The following statements are equivalent:*

1. *The survival function S satisfies the TFO condition.*
2. *There exist positive functions a and g such that for all $t \in J$,*

$$\lim_{\alpha \to 0} \frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(\alpha^{-1})} = \phi(t). \tag{9}$$

3. *There exist sequences* $a_n > 0$, $b_n \in \mathbb{R}$ *and* $c_n > 0$ *with* $nc_n \to \infty$ *such that for all* $t \in I$,

$$\lim_{n\to\infty} [nc_n S(a_n t + b_n) - c_n] = \phi^{-1}(t). \tag{10}$$

Remark (1) The relations between the auxiliary functions involved in (8) and (9) are: $d(\cdot) = a(1/S(\cdot))$ and $\Psi(\cdot) = S(\cdot)/g(S(\cdot))$.

(2) A possible choice for the sequences a_n, b_n and c_n in (10) is $a_n = a(n)$, $b_n = Q(1/n)$ and $c_n = 1/[ng(1/n)]$. It is also easy to check that necessarily $g(\alpha) \to 0$ as $\alpha \to 0$.

(3) An interpretation of condition (9) is based on the following remark: from the second statement of Proposition 1,

$$\frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{tg(\alpha)} \sim -\frac{a(\alpha^{-1})}{g(\alpha)},$$

as $(t, \alpha) \to (0, 0)$. Hence, one can see the function $-a(\alpha^{-1})/g(\alpha)$ as the derivative of Q near 0 and in the direction of $g(\alpha)$. This heuristic is confirmed by the next result which provides a sufficient condition for the TFO condition.

Proposition 3 *Assume that* Q *is a differentiable function and that for some open interval* $J \subset \mathbb{R}$ *containing 0, there exists a positive function* g *such that for all* $t \in J$,

$$\lim_{\alpha\to 0} \frac{Q'[\alpha + tg(\alpha)]}{Q'(\alpha)} = \Theta(t). \tag{11}$$

If for all $t \in J$, $\int_0^t \Theta(s)\mathrm{d}s =: \theta(t) \in \mathbb{R}$ *where* θ *is an increasing function on* J *such that* $\theta(t)/t \to 1$ *as* $t \to 0$ *then condition* (9) *holds with* $\phi(t) = -\theta(t)$ *and* $a(\alpha^{-1}) = -Q'(\alpha)g(\alpha)$.

We conclude this section by giving examples of distributions satisfying the TFO condition.

Maximum domain of attraction—In extreme value theory, in order to make inference on the tail of a distribution S, we classically assume that there exist sequences $a_n > 0$ and b_n and a non-degenerate distribution function G for which

$$\lim_{n\to\infty} [1 - S(a_n x + b_n)]^n = G(x), \tag{12}$$

for all point of continuity of G. Fisher and Tippett (1928) and Gnedenko (1943) show that $G(x) = G_\gamma(ax + b)$ for some $a > 0$ and $b \in \mathbb{R}$ where

$$G_\gamma(x) = \exp\left[-(1 + \gamma x)^{-1/\gamma}\right],$$

for all x such that $1 + \gamma x > 0$. A survival function S satisfying (12) is said to belong to the maximum domain of attraction of the extreme value distribution G_γ.

The parameter $\gamma \in \mathbb{R}$ is called the extreme value index. As established in de Haan and Ferreira (2006, Theorem 1.1.6), condition (12) is equivalent to assume the existence of a positive auxiliary function a and a non-constant function ϕ for which

$$\lim_{\alpha \to 0} \frac{Q(t\alpha) - Q(\alpha)}{a(\alpha^{-1})} = \phi(t). \tag{13}$$

From de Haan and Ferreira (2006, Theorem B.2.1), the function ϕ in (13) is necessarily of the form $\phi(t) = c(t^{-\gamma} - 1)/\gamma$ for some $c \neq 0$ and where $\gamma \in \mathbb{R}$ is always the extreme value index.

The aim of the next result is to show that the TFO condition introduced in this paper (see Definition 1) is weaker than (12).

Proposition 4 *If S satisfies the TFO condition with an auxiliary function g in (9) such that $\alpha/g(\alpha) \to c \geq 0$ as $\alpha \to 0$ (with g continuous and strictly increasing if $c = 0$) then S satisfies (12).*

As a consequence of this result, if a survival function S satisfies the TFO condition with a function g as in Proposition 4, then S also satisfies the TFO condition with $g(\alpha) = \alpha$ and in this case the TFO condition coincides with the classical extreme value condition. Remark also that in this situation (i.e., $g(\alpha) = \alpha$), condition (11) is equivalent to assume that

$$\lim_{\alpha \to 0} \frac{Q'(t\alpha)}{Q'(\alpha)} = t^{-\gamma-1},$$

for some $\gamma \in \mathbb{R}$. This condition coincides with condition (1.1.33) in de Haan and Ferreira (2006, Corollary 1.1.10).

At this step, a natural question is: "Can we find survival functions that satisfy the TFO condition but not the classical extreme value one ?" Roughly speaking, this is equivalent to find survival functions S such that (9) holds with a function g such that $\alpha/g(\alpha) \to \infty$. An example of such survival functions is given by super heavy-tailed distributions.

Super heavy-tailed distributions—The term *super heavy-tailed* is often attached in the literature to a distribution with a slowly varying survival function S, i.e., such that for all $t > 0$,

$$\lim_{x \to \infty} \frac{S(tx)}{S(x)} = 1. \tag{14}$$

It can be shown that these survival functions do not satisfy the classical first-order condition (12). Note that a heavy-tailed distribution corresponds to a survival function satisfying for all $t > 0$, $S(tx)/S(x) \to t^{-1/\gamma}$ as $x \to \infty$, for some $\gamma > 0$. Hence, roughly speaking, a super heavy-tailed distribution is a heavy-tailed distribution with $\gamma = +\infty$.

Unfortunately, condition (14) is not precise enough for the study of super heavy-tailed distribution. To define more precisely the class of super heavy-tailed distribution, we start by remarking that for heavy-tailed distributions, there exists $\gamma > 0$ such that for all $s > -1$,

$$\lim_{\alpha \to 0} \frac{Q[(1+s)\alpha]}{Q(\alpha)} = (1+s)^{-\gamma}.$$

Since super heavy-tailed distribution can be seen as a heavy-tailed distribution with $\gamma = +\infty$, we propose to replace in the previous limit γ by $\gamma(\alpha)$ where $\gamma(\alpha) \to \infty$ as $\alpha \to 0$, and s by $t/\gamma(\alpha)$ with $t \in \mathbb{R}$ to obtain a non-degenerate limit:

$$\lim_{\alpha \to 0} \frac{Q[(1+t/\gamma(\alpha))\alpha]}{Q(\alpha)} = \mathrm{e}^{-t}.$$

The class of super heavy-tailed distributions can thus be defined by the set of distributions for which there exists a positive function g with $g(\alpha)/\alpha \to 0$ as $\alpha \to 0$ such that for all $t \in \mathbb{R}$

$$\lim_{\alpha \to 0} \frac{Q[\alpha + tg(\alpha)]}{Q(\alpha)} = \mathrm{e}^{-t}. \tag{15}$$

It appears that convergence (15) coincides with the TFO condition with $a(\alpha^{-1}) = Q(\alpha)$ and $\phi(t) = \mathrm{e}^{-t} - 1$. As shown in Proposition 5 below, this definition is equivalent to the one introduced for instance in Fraga Alves et al. (2009) where the class of super heavy-tailed distributions is defined as the set of distributions for which there exists a positive function b such that

$$\lim_{x \to \infty} \frac{U[x + tb(x)]}{U(x)} = \mathrm{e}^{t} \tag{16}$$

with $U(\cdot) := Q(1/\cdot)$. Note that according to Fraga Alves et al. (2009, Lemma 4.1), condition (16) implies (14). Furthermore, the function b is such that $b(x)/x \to 0$ as $x \to \infty$. Remark finally that the right endpoint of a distribution satisfying (16) is necessarily infinite. As examples of super heavy-tailed distribution satisfying (16), one can cite the standard log-Pareto distribution given by $S(x) = [\log(x)]^{-\xi}$ with $\xi > 0$ and the log-Weibull distribution for which $S(x) = \exp(-\xi \log^{\theta} x)$, with $\xi > 0$ and $\theta \in (0, 1)$. For these two distributions, one can take $b \sim U/U'$.

Proposition 5 *Conditions* (15) *and* (16) *are equivalent. The relation between the involved functions is* $b(x) = x^2 g(x^{-1})$.

4 Extreme conditional quantile estimation

Let (X, Y) be a random vector taking its values in $\mathbb{R}^p \times \mathbb{R}$. In all what follows, we assume that (X, Y) admits a probability density function (PDF). The marginal PDF of X is denoted by f. As in the introduction, for all $x_0 \in \mathbb{R}^p$, let $S(\cdot|x_0)$ and $Q(\cdot|x_0)$ be the survival function and the quantile function of the conditional distribution of Y

given $\{X = x_0\}$, respectively. Given n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y), the first part of this section is dedicated to the presentation of a large class of estimators of $Q(\cdot|x_0)$. In the second part, we show that under the TFO condition, the proposed estimators computed with an extreme level $\alpha_n \to 0$ are asymptotically Gaussian.

4.1 A class of conditional quantile estimators

As mentioned in the introduction we focus in this paper on indirect estimators of $Q(\cdot|x_0)$. The first step is thus the estimation of the conditional survival function $S(\cdot|x_0)$. We consider estimators of the form

$$\widehat{S}_n(y|x_0) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{(y,\infty)}(Y_i). \tag{17}$$

The set of weights $\{\mathcal{W}_{n,i}(x_0),\ 1 \le i \le n\}$ is a triangular array of positive random variables depending on the data $X_1, \ldots X_n$ as well as on x_0 such that

$$\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) = 1.$$

These properties on the random weights ensure that $\widehat{S}_n(\cdot|x_0)$ is a well-defined distribution function. This is crucial to estimate the conditional quantile by inverting estimator (17). This class of estimators encompasses various classical estimators of the conditional distribution function, see below for some examples. The indirect estimator of the conditional quantile of level $\alpha \in (0, 1)$ is thus defined as in (1) by

$$\widehat{Q}_n(\alpha|x_0) := \widehat{S}_n^{\leftarrow}(\alpha|x_0) = \inf\{y \in \mathbb{R};\ \widehat{S}_n(y|x_0) \le \alpha\}.$$

Of course, the main feature of the weights $\{\mathcal{W}_{n,i}(x_0),\ 1 \le i \le n\}$ is to select a set of data around x_0. For this reason, estimator of form (17) is called *weighted local* estimators.

The kernel-based estimator introduced by Nadaraya (1964) and Watson (1964) is a classical example of weighted local estimator. This estimator is obtained by using the following random weights in (17):

$$\mathcal{W}_{n,i}^{\mathrm{NW}}(x_0, h_n) := K\left(\frac{X_i - x_0}{h_n}\right) \bigg/ \sum_{j=1}^{n} K\left(\frac{X_j - x_0}{h_n}\right), \tag{18}$$

where K is a density on $\mathbb{R}^p$ and h_n is a positive non-random sequence satisfying $h_n \to 0$ as $n \to \infty$. Typically, the probability density function K has a unique mode at 0 in order to give the largest values of the weights for the observations close to x_0.

Another possibility to select the observations is to take the k_n observations which are closest to the reference position x_0. This approach is called the k_n-nearest neighbors (k_n-NN) method. More specifically, for some norm $\|\cdot\|$ on $\mathbb{R}^p$, let $\{D_i(x_0) := \|X_i - x_0\|,\ i = 1,\dots,n\}$ be the distances between each observation and x_0 and let $D_{(1)}(x_0) \le \dots \le D_{(n)}(x_0)$ the corresponding order statistics. Denoting by $\{r(i),\ i = 1,\dots,n\}$ the ranks of these distances (i.e., $D_{(i)}(x_0) = D_{r(i)}(x_0)$ for $i = 1,\dots,n$), the k_n-NN estimator is obtained by using the following random weights in (17):

$$\mathcal{W}_{n,i}^{\mathrm{NN}}(x_0, k_n) := \left[(k_n - r(i) + 1)_+\right]^{\ell} \Big/ \sum_{j=1}^{k_n} j^{\ell}, \tag{19}$$

where $(\cdot)_+$ stands for the positive part function and $\ell \in \mathbb{N}$. For instance, by taking $\ell = 0$ (with the convention $0^0 = 0$), we affect the same weight to the k_n closest observations. The corresponding weights are referred to as uniform k_n-NN weights. The choice $\ell = 1$ (resp., $\ell = 2$) leads to triangular k_n-NN weights (resp., quadratic k_n-NN weights).

Roughly speaking, the main difference between these two sets of weights is that the kernel-based estimator averages over all observations which are within a fixed distance, whereas the k_n-NN approach averages over a fixed number of observations which might be arbitrarily far away. Of course, one can also think about a linear combination (LC) of (18) and (19). For instance, we can consider the random weights defined for $\tau \in (0, 1)$ by

$$\mathcal{W}_{n,i}^{\mathrm{LC}}(x_0, \tau, h_n, k_n) := \frac{\tau}{M_n}\mathbb{I}_{[0,1]}\left(\left\|\frac{X_i - x_0}{h_n}\right\|_\infty\right) + \frac{1-\tau}{k_n}\mathbb{I}_{[0,1]}\left(\frac{r(i)}{k_n}\right), \tag{20}$$

where M_n is the random number of random variables among $\{X_1,\dots,X_n\}$ that belong to $\mathcal{B}_{x_0}(h_n)$, the closed ball with respect to $\|\cdot\|_\infty$ centered at x_0 and with radius h_n.

4.2 Main results

Under general conditions on the random weights $\{\mathcal{W}_{n,i}(x_0),\ i = 1,\dots,n\}$, we want to establish the convergence in distribution of a normalized version of $\widehat{Q}_n(\alpha_n|x_0)$ for a level α_n converging to 0 as $n \to \infty$. As outlined in Sect. 2, we first need to find a sequence $v_n(x_0) \to \infty$ and a non-degenerate distribution Λ such that (under some additional assumptions)

$$v_n(x_0)\left(\widehat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0)\right) \xrightarrow{d} \Lambda,$$

for some sequence $y_n(x_0) \uparrow y^*(x_0)$. This is done in Theorem 1 where the following notation is used

$$n_{x_0} := \left(\sum_{i=1}^{n} \mathcal{W}_{n,i}^2(x_0)\right)^{-1}.$$

Note that the random variable n_{x_0} corresponds, roughly speaking, to the number of observations used in the estimation procedure. For instance, for the Nadaraya–Watson (NW) weights with the uniform kernel $K(\cdot) \propto \mathbb{I}_{[0,1]}(\|\cdot\|_\infty)$, it is easy to check that n_{x_0} is exactly the number of points in $\mathcal{B}_{x_0}(h_n)$. For the uniform k_n-NN weights, one has $n_{x_0} = k_n$, the number of nearest neighbors.

Theorem 1 *Let $x_0 \in \mathbb{R}^p$ such that $f(x_0) > 0$ and let $y_n(x_0)$ be a sequence converging to the right endpoint $y^*(x_0)$ of the conditional distribution of Y given that $X = x_0$. Assume that there exists a sequence $m_n(x_0)$ such that $n_{x_0}/m_n(x_0) \xrightarrow{a.s.} 1$ and let $v_n^2(x_0) := m_n(x_0)/S(y_n(x_0)|x_0)$. Under the conditions*

$$v_n(x_0) \max_{1\le i\le n} \mathcal{W}_{n,i}(x_0) \xrightarrow{a.s.} 0 \tag{21}$$

and

$$v_n(x_0) \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\, |S(y_n(x_0)|X_i) - S(y_n(x_0)|x_0)| \xrightarrow{\mathbb{P}} 0, \tag{22}$$

we have that $v_n(x_0)\left(\widehat{S}_n\,(y_n(x_0)|x_0) - S\,(y_n(x_0)|x_0)\right) \xrightarrow{d} \mathcal{N}(0, 1)$.

To understand the usefulness of conditions (21) and (22), we provide below the main ideas of the proof of Theorem 1, the complete proof being postponed to Sect. 7. Let $Y_i^{x_0} := Q(U_i|x_0)$ where $U_1, U_2, \ldots$ are independent standard uniform random variables, independent of the X_i. The random vectors $\{(X_i, Q(U_i|X_i)),\ i = 1, \ldots, n\}$ are thus independent and distributed as (X, Y), which implies that

$$\widehat{S}_n(y_n(x_0)|x_0) \stackrel{d}{=} \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{(y_n(x_0),\infty)}(Q(U_i|X_i)).$$

In other words, one can work as if $Y_i = Q(U_i|X_i)$. The starting point of the proof is the decomposition

$$\begin{aligned}\widehat{S}_n\,(y_n(x_0)|x_0) - S\,(y_n(x_0)|x_0) &= \left[\widehat{S}_n^{x_0}\,(y_n(x_0)) - S\,(y_n(x_0)|x_0)\right]\\ &\quad + \left[\widehat{S}_n\,(y_n(x_0)|x_0) - \widehat{S}_n^{x_0}\,(y_n(x_0))\right],\end{aligned}$$

where for all $y \in \mathbb{R}$,

$$\widehat{S}_n^{x_0}(y) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{(y,\infty)}(Y_i^{x_0}).$$

Since $\mathbb{E}[\widehat{S}_n^{x_0}(y_n(x_0))] = S(y_n(x_0)|x_0)$, the first term corresponds to the *variance term* and the second one to the *bias term*.

The first part of the proof consists in establishing the asymptotic normality of the normalized variance term given by:

$$v_n(x_0)\left[\widehat{S}_n^{x_0}\left(y_n(x_0)\right) - S\left(y_n(x_0)|x_0\right)\right],$$

see Sect. 7, Proposition 7. This is obtained mainly by applying the Lindeberg theorem, and only condition (21) is required. This condition is in fact equivalent to the Lindeberg condition.

In the second part of the proof, we show that the bias term given by

$$B_n(x_0) := v_n(x_0)\left[\widehat{S}_n\left(y_n(x_0)|x_0\right) - \widehat{S}_n^{x_0}\left(y_n(x_0)\right)\right]$$

converges to 0 in probability (see Sect. 7, Proposition 8). The proof is based on the following remark. Let $\mathcal{W}_{n,x_0}$ be the discrete random measure define for all $A \in \mathcal{B}(\mathbb{R}^p)$ by

$$\mathcal{W}_{n,x_0}(A) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\delta_{X_i}(A).$$

Straightforward calculation leads to

$$\widehat{S}_n(y_n(x_0)|x_0) - \widehat{S}_n^{x_0}(y_n(x_0)) = \int \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{(y_n(x_0),\infty)}(Q(U_i|\cdot))\left(\mathrm{d}\mathcal{W}_{n,x_0} - \mathrm{d}\delta_{x_0}\right).$$

To control the bias term, we need to measure the discrepancy between the two probability measures $\mathcal{W}_{n,x_0}$ and δ_{x_0}. A useful distance between probability measures is the Wasserstein distance defined for all probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ by $W_1(\mathbb{P}_1, \mathbb{P}_2) = \inf\{[\mathbb{E}(|X_1 - X_2|)],\ X_1 \sim \mathbb{P}_1,\ X_2 \sim \mathbb{P}_2\}$. Condition (22) can in fact be written in term of the Wasserstein distance as follows:

$$v_n(x_0) W_1(\mathcal{W}^*_{n,x_0}, \delta^*_{x_0}) \xrightarrow{\mathbb{P}} 0, \tag{23}$$

where $\mathcal{W}^*_{n,x_0}$ and $\delta^*_{x_0}$ are the pushforward measures of $\mathcal{W}_{n,x_0}$ and δ_{x_0} by the measurable function $S(y_n(x_0)|\cdot)$.

We have now all the ingredients to establish the asymptotic distribution of the conditional quantile estimator of level α_n obtained by inverting the estimator $\widehat{S}_n(\cdot|x_0)$. This requires the following first order condition on the conditional distribution of Y given $X = x_0$.

(H) The conditional survival function $S(\cdot|x_0)$ satisfies the TFO condition with positive auxiliary functions $\Psi_{x_0} \equiv \Psi$ and $d_{x_0} \equiv d$.

Let $a(1/\cdot) \equiv a_{x_0}(1/\cdot) = d[Q(\cdot|x_0)]$ and $g(\cdot) \equiv g_{x_0}(\cdot) = \cdot/\Psi[Q(\cdot|x_0)]$. From Proposition 2, condition **(H)** is equivalent to assume that for some open interval $J_{x_0} = J \subset \mathbb{R}$ containing 0, one has for all $t \in J$

$$\lim_{\alpha\to 0} \frac{Q(\alpha + tg(\alpha)|x_0) - Q(\alpha|x_0)}{a(\alpha^{-1})} = \phi_{x_0}(t),$$

where $\phi_{x_0} \equiv \phi$ is a continuous and strictly decreasing function such that $\phi(t)/t \to -1$ as $t \to 0$.

Theorem 2 *Let* $x_0 \in \mathbb{R}^p$ *such that* $f(x_0) > 0$ *and assume that condition* **(H)** *holds. Assume that there exists a sequence* $m_n(x_0)$ *such that* $n_{x_0}/m_n(x_0) \xrightarrow{a.s.} 1$ *and let* $v_n^2(x_0) := m_n(x_0)/\alpha_n$. *If* $\alpha_n m_n(x_0) \to \infty$, $v_n(x_0) g(\alpha_n) \to \infty$,

$$v_n(x_0) \max_{1\le i\le n} \mathcal{W}_{n,i}(x_0) \xrightarrow{a.s.} 0$$

$$\textit{and} \quad [\alpha_n m_n(x_0)]^{1/2} \sup_{|\beta/\alpha_n - 1|\le \xi} \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \left| \frac{S[Q(\beta|x_0)|X_i]}{\beta} - 1 \right| \xrightarrow{\mathbb{P}} 0,$$

for some $\xi \in (0, 1)$ *then*

$$v_n(x_0) \frac{g(\alpha_n) Q(\alpha_n|x_0)}{a(\alpha_n^{-1})} \left(\frac{\widehat{Q}_n(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

Recall that if $g(\alpha) = \alpha$ (or equivalently $\Psi \equiv 1$), condition **(H)** coincides with the classical first-order condition (13) used in extreme value theory. In this case, $\phi(t) \propto (t^{-\gamma(x_0)} - 1)/\gamma(x_0)$ where the function γ is referred to as the conditional extreme value index. Under (13) and if the conditions of Theorem 2 are satisfied,

$$[\alpha_n m_n(x_0)]^{1/2} \frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1}|x_0)} \left(\frac{\widehat{Q}_n(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

Moreover, we know from de Haan and Ferreira (2006, Lemma 1.2.9) that under (13), $Q(\alpha_n|x_0)/a(\alpha_n^{-1}) \to 1/\gamma_+(x_0)$, where $\gamma_+(x_0) = \max(\gamma(x_0), 0)$. So, under the first-order condition (13), the worst rate of convergence is achieved when $\gamma(x_0) > 0$. This was expected since the case $\gamma(x_0) > 0$ corresponds to heavy-tailed distributions.

Let us now focus on the rate of convergence in Theorem 2 for conditional super heavy-tailed distribution. Taking the definition of super heavy-tailed distributions given in Fraga Alves et al. (2009) into account, we have in this case $a(\alpha^{-1}) = Q(\alpha|x_0)$ and $g(\alpha)/\alpha \to 0$ as $\alpha \to 0$. Hence, for these distributions,

$$[\alpha_n m_n(x_0)]^{1/2} \frac{g(\alpha_n)}{\alpha_n} \left(\frac{\widehat{Q}_n(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

Not surprisingly, this rate is worse than the one for heavy-tailed distributions.

Theorem 2 is proved under general conditions on the random weights used to define the conditional survival estimator (17). We close this section by applying Theorem 2 to particular weights.

Nadaraya–Watson weights—Taking the weights defined in (18) leads to the well-known NW estimator of the conditional survival function:

$$\widehat{S}_n^{\mathrm{NW}}(y|x_0) := \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h_n}\right)\mathbb{I}_{(y,\infty)}(Y_i) \Big/ \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h_n}\right). \tag{24}$$

The corresponding conditional quantile estimator is denoted by $\widehat{Q}_n^{\mathrm{NW}}(\alpha_n|x_0)$. In order to apply Theorem 2, we need to check that the NW weights satisfy the required conditions. To this aim, we assume the following on the kernel function K:

(K) the kernel K is either an indicator function on a cell of $\mathbb{R}^p$ or such that $K(x) = L(\|x\|)$ where L is of bounded variation, continuous on $(0, \infty)$ and with support $[0, 1]$.

It is very easy to check that **(K)** is satisfied for a large range of usual kernels such as the uniform kernel ($K(t) \propto \mathbb{I}_{[0,1]}(\|t\|_\infty)$), triangular (with $L(t) \propto 1 - t$), Epanechnikov kernel ($L(t) \propto 1 - t^2$), biweight kernel ($L(t) \propto (1 - t^2)^2$), etc.
We can now state the convergence in distribution of the conditional survival estimator (24). Recall that f is the PDF of X.

Corollary 1 *Let $x_0 \in \mathbb{R}^p$ such that f is continuous at x_0 and $f(x_0) > 0$ and let K be a kernel satisfying* **(K)**. *Under* **(H)**, *for sequences $h_n \to 0$ and $\alpha_n \in (0, 1)$ such that $nh_n^p[\alpha_n \wedge (\log\log n)^{-1}] \to \infty$, $\alpha_n^{-1} nh_n^p g^2(\alpha_n) \to \infty$ and*

$$\sup_{\substack{|\beta/\alpha_n - 1| \leq \xi \\ \|x - x_0\| \leq h_n}} \left|\frac{S[Q(\beta|x_0)|x]}{\beta} - 1\right|^2 = o\left(\frac{1}{nh_n^p \alpha_n}\right), \tag{25}$$

for some $\xi \in (0, 1)$ we have

$$\frac{g(\alpha_n)}{\alpha_n}\frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1})}\left(nh_n^p\alpha_n\right)^{1/2}\left(\frac{\widehat{Q}_n^{\mathrm{NW}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\|K\|_2^2}{f(x_0)}\right).$$

Note that under the classical first-order condition (13) (i.e., when $g(\alpha_n) = \alpha_n$ in **(H)**, see (9) and the remarks below Proposition 2), the asymptotic normality of the NW conditional quantile estimator has already been obtained in Daouia et al. (2013, Corollary 1). This last result also requires the use of condition (25) which controls the oscillations of the function $Q(\alpha_n|\cdot)$. Of course, the proof of Daouia et al. (2013, Corollary 1) uses arguments adapted to the NW estimator while Theorem 2 can be used for a large range of weighted conditional survival estimators. As a consequence, conditions on h_n and α_n involved in Daouia et al. (2013, Corollary 1) and in our Corollary 1 are slightly different. More precisely, the conditions in Daouia et al. (2013, Corollary 1) are $nh_n^p\alpha_n \to \infty$ and $nh_n^{p+2}\alpha_n \to 0$ while in our Corollary 1 it is required that $nh_n^p\alpha_n \to \infty$ and $nh_n^p(\log\log n)^{-1} \to \infty$. Hence, if $\alpha_n \log\log n \to 0$ as $n \to \infty$ (i.e., for large quantiles), conditions on the sequences h_n and α_n are weaker in Corollary 1 than in Daouia et al. (2013, Corollary 1).

Nearest neighbors approach—Now, let us consider the k_n-NN random weights defined in (19) and leading to the conditional survival function estimator

$$\widehat{S}_n^{KNN}(y|x_0) := \sum_{i=1}^{n} [(k_n - r(i) + 1)_+]^{\ell} \mathbb{I}_{(y,\infty)}(Y_i) \Big/ \sum_{j=1}^{k_n} j^{\ell},$$

with $k_n \in \{1, \dots, n\}$, $\ell \in \mathbb{N}$ and $r(i)$ is the rank of $\|X_i - x_0\|$ among the random variables $X_1, \dots, X_n$. The asymptotic normality of the k_n-NN conditional quantile estimator $\widehat{Q}_n^{\mathrm{KNN}}(\alpha_n|x_0)$ is established in the following result.

Corollary 2 *Let $x_0 \in \mathbb{R}^p$ such that $f(x_0) > 0$. Under* **(H)***, for sequences $k_n \to \infty$ and $\alpha_n \in (0,1)$ such that $k_n\alpha_n \to \infty$, $\alpha_n^{-1} k_n g^2(\alpha_n) \to \infty$ and*

$$(k_n\alpha_n) \sup_{\substack{|\beta/\alpha_n - 1| \leq \xi \\ \|x - x_0\| \leq D_{(k_n)}(x_0)}} \left| \frac{S[Q(\beta|x_0)|x]}{\beta} - 1 \right|^2 \xrightarrow{\mathbb{P}} 0,$$

for some $\xi \in (0,1)$, we have

$$\frac{g(\alpha_n)}{\alpha_n} \frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1})} (k_n\alpha_n)^{1/2} \left(\frac{\widehat{Q}_n^{\mathrm{KNN}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{(\ell+1)^2}{2\ell+1}\right).$$

The asymptotic variance $(\ell+1)^2/(2\ell+1)$ is an increasing function of ℓ, and thus, the best choice (at least in term of variance) seems to be $\ell = 0$, i.e., when the same weight $1/k_n$ is affected to the k_n observations closest to x_0.

Linear combination of weights – We finally focus on the estimator $\widehat{Q}_n^{\mathrm{LC}}(\alpha_n|x_0)$ of $Q(\alpha_n|x_0)$ obtained by using the LC weights introduced in (20).

Corollary 3 *Let $x_0 \in \mathbb{R}^p$ such that f is continuous at x_0 and $f(x_0) > 0$. Let $h_n \to 0$, $k_n \to \infty$ and α_n be sequences such that $nh_n^p/\log\log n \to \infty$, $\ell_n\alpha_n \to \infty$ with $\ell_n := (nh_n^p \wedge k_n)$, $\alpha_n^{-1}\ell_n g^2(\alpha_n) \to \infty$ and*

$$(\ell_n\alpha_n) \sup_{\substack{|\beta/\alpha_n - 1| \leq \xi \\ \|x - x_0\| \leq (h_n \vee D_{(k_n)}(x_0))}} \left| \frac{S[Q(\beta|x_0)|x]}{\beta} - 1 \right|^2 \xrightarrow{\mathbb{P}} 0,$$

for some $\xi \in (0,1)$. Under **(H)** *and if there exists $\kappa \in [0,\infty]$ such that $k_n/(nh_n^p) \to \kappa$, we have*

$$\frac{g(\alpha_n)}{\alpha_n} \frac{Q(\alpha_n|x_0)}{a(\alpha_n^{-1})} (\ell_n\alpha_n)^{1/2} \left(\frac{\widehat{Q}_n^{\mathrm{LC}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{C^2(\kappa)}{2^p f(x_0)}\right).$$

In practice, one can take $k_n = \lfloor \kappa n h_n^p \rfloor$ with $\kappa > 0$. The parameter κ is thus a tuning parameter that has to be chosen by a data-driven procedure (see Sect. 5.1).

5 Simulation study

In this section, we are interested in the finite sample behavior of the estimator $\widehat{Q}_n(\alpha_n|x_0)$ defined in (1) for a given value of x_0. The random weights $\{\mathcal{W}_{n,1}(x_0),\ldots,\mathcal{W}_{n,n}(x_0)\}$ used in the expression of estimator (17) of the conditional survival function $S(\cdot|x_0)$ often depend on an hyperparameter $\lambda_n \in \mathbb{R}^d$, $d \in \mathbb{N}\setminus\{0\}$, useful in order to control the smoothness of the estimator. This is the case for instance for the NW weights, the k_n-NN random weights or the LC weights defined in (18), (19) and (20), where λ_n is equal to h_n, k_n and (h_n,κ), respectively. In the next section, we propose an adaptive procedure to select λ_n in practice.

5.1 Choice of the hyperparameter

For $t \in \mathbb{R}^p$, let us denote by $\widehat{Q}_n(\alpha_n|t,\lambda_n)$ an estimator of $Q(\alpha_n|t)$ depending on an hyperparameter λ_n and by $\widehat{Q}_{n,-i}(\alpha_n|t,\lambda_n)$ the estimator computed without the random pair (X_i,Y_i).

Our procedure of selection is based on the following simple remark: for a good choice of λ_n, the random value $S[\widehat{Q}_{n,-1}(\alpha_n|X_1,\lambda_n)|X_1]$ should be close to α_n at least when the observed value of X_1 is close to x_0. We thus propose to define our *optimal* value of λ_n as $\lambda_{\text{opt}} := \arg\min\{\Lambda_n^2(\lambda),\ \lambda\in\mathbb{R}^d\}$, with

$$\Lambda_n(\lambda) := \mathbb{E}\left[\frac{\mathcal{W}_{n,1}(x_0)}{\mathbb{E}\left[\mathcal{W}_{n,1}(x_0)\right]} S[\widehat{Q}_{n,-1}(\alpha_n|X_1,\lambda)|X_1]\right] - \alpha_n.$$

Note that the proximity of X_1 and x_0 is controlled by the random weight $\mathcal{W}_{n,1}(x_0)$. Of course, the function Λ_n is unknown in practice and should be estimated. We propose to use the following estimator

$$\hat{\Lambda}_n(\lambda) := \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0)\mathbb{I}_{\{Y_i>\widehat{Q}_{n,-i}(\alpha_n|X_i,\lambda)\}} - \alpha_n. \tag{26}$$

The estimated optimal value of the hyperparameter λ_n is thus given by

$$\widehat{\lambda}_{n,\text{opt}} := \arg\min\{\hat{\Lambda}_n^2(\lambda),\ \lambda\in\mathbb{R}^d\}. \tag{27}$$

Estimator (26) can be motivated by the following result.

Proposition 6 *If there exists a function $\varphi : \mathbb{R}^p \times \mathbb{R}^{p\times(n-1)} \mapsto [0,\infty)$ such that for all $i = 1,\ldots,n$, $\mathcal{W}_{n,i}(x_0) = \varphi(X_i,\mathbb{X}_{-i})$ where the matrix $\mathbb{X}_{-i}$ is given by $[X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n]$ then $\mathbb{E}[\hat{\Lambda}_n(\lambda)] = \Lambda_n(\lambda)$ for all $\lambda\in\mathbb{R}^d$.*

Note that the assumption of Proposition 6 is satisfied for the NN approach with the function φ defined for $t \in \mathbb{R}^p$ and $u = [u_1,\ldots,u_{n-1}] \in \mathbb{R}^{p\times(n-1)}$ by $\varphi(t,u) = \lambda^{-1}\mathbb{I}_{\{\|t-x_0\|<d_{(\lambda)}(x_0)\}}$, where $d_i(x_0) = \|u_i - x_0\|$, $i = 1,\ldots,n-1$ and $d_{(1)}(x_0) \le \ldots \le d_{(n-1)}(x_0)$ are the corresponding ordered values.

This is also the case for the NW weights by using the function

$$\varphi(t,u) = K[(t-x_0)/\lambda] \Big/ \left(\sum_{i=1}^{n-1} K[(u_i - x_0)/\lambda] + K[(t-x_0)/\lambda] \right).$$

5.2 Finite sample behavior

Using a sample of size n from the random vector (X, Y), we are interested in estimating an extreme conditional quantile in the situation where the quantile level α_n is not too small. We consider the situation where X is a real-valued random variable ($p = 1$). In a theoretical point of view, we assume that the conditions of Theorem 2 are satisfied for such a sequence α_n. In practice, we take $\alpha_n = 20/n$ and the quantile $Q(\alpha_n|x_0)$ is estimated using (1). Three sets of random weights are considered:

(i) the NW weights with the Epanechnikov kernel: $K(u) = \frac{3}{4}(1-u^2)\mathbb{I}_{[0,1]}(|u|)$,
(ii) the k_n-NN weights with $\ell = 1$ (triangular k_n-NN weights),
(iii) the LC weights given in (20) with $\tau = 1/2$ and $k_n = \lfloor \kappa n h_n \rfloor$.

Although the theory on our estimators is valid without any assumption on the distribution of X, from a practical point of view, the estimation is very difficult in case of unbounded distribution, especially at the border. For this reason, we illustrate our methodology in the case of a bounded distribution, namely the standard uniform distribution. The four following models have been considered for the conditional survival distribution function of Y given X:

M1—Conditional Burr distribution:

$$S(y|X) = \left(1 + y^{-\rho/\gamma(X)}\right)^{1/\rho}, y > 0,$$

where $\rho < 0$ and for all $x \in (0,1)$, $\gamma(x) = 2x(1-x)$.

It is well-known that for this model, condition (13) holds (i.e., condition **(H)** with $g(\alpha|x_0) = \alpha$, see, e.g., Embrechts et al. (1997), Table 3.4.2). The parameter ρ is referred in the literature to as the second-order parameter and it affects the bias of the estimator.

M2—Conditional Beta distribution with parameters $\theta_1 > 0$ and $\theta_2(X)$ where for all $x \in (0,1)$, $\theta_2(x) = 1/[2x(1-x)]$.

This conditional distribution satisfies condition (13) with a conditional extreme value index given by $\gamma(x) = -1/\theta_2(x) < 0$ (see, e.g., Embrechts et al. (1997), Table 3.4.3).

M3—Conditional Gaussian distribution with mean $\mu(X) = 2X(1-X)$ and variance σ^2.

Under this model, condition (13) is satisfied with $\gamma(X) = 0$ (see, e.g., Embrechts et al. (1997), Table 3.4.4).

We finally consider a model for which condition (13) does not hold.

M4—Conditional super heavy-tailed distribution:

$$S(y|X) = \exp\left\{-\xi[\ln(y)]^{\theta(X)}\right\},\ y > 1,$$

with $\xi > 0$ and $\theta(x) = 19(x + 1/2)(3/2 - x)/20 \in [0, 0.95]$.

One can check that this conditional distribution satisfies condition **(H)** with

$$a(\alpha^{-1}) = Q(\alpha|x) = \exp\left\{\left[\frac{\ln(1/\alpha)}{\xi}\right]^{1/\theta(x)}\right\} \text{ and } g(\alpha) = \alpha\theta(x)\xi\left[\frac{\ln(1/\alpha)}{\xi}\right]^{1-1/\theta(x)}.$$

For each model, $N = 500$ samples of size $n = 1000$ have been generated. The hyperparameter λ_n is chosen according to (27), and the minimization is achieved

- over a regular grid $\mathcal{H}$ of 20 points evenly spaced between 0.05 and 0.3 for the NW weights,
- over a grid $\mathcal{K}$ of 20 points evenly spaced between 100 and 600 for the NN weights,
- over the grid $\mathcal{H} \times \mathcal{F}$ where $\mathcal{F}$ is a grid of 5 evenly spaced points between 0.9 and 1.1.

The accuracy of the estimators is measured by the errors

$$\text{RMSE} := \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\widehat{Q}_n^{\bullet,\text{i}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1\right]^2} \text{ and ARE} := \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\widehat{Q}_n^{\bullet,\text{i}}(\alpha_n|x_0)}{Q(\alpha_n|x_0)} - 1\right|,$$

where $\bullet$ has to be replaced by NW, NN or LC and the index i refers to the $i-$ simulation run. The error RMSE corresponds to the root mean squared error of the ratio between the estimates and the true quantile value. The error ARE is the average over all replications of the absolute value of the relative error. The estimation of $Q(\alpha_n|x_0)$ is done at three different positions: $x_0 := x_0^{(1)} = (1 - \sqrt{1/3})/2 \approx 0.211$, $x_0 = x_0^{(2)} = 1/2$ and $x_0 = x_0^{(3)} = (1 + \sqrt{1/2})/2 \approx 0.854$. The results are gathered in Tables 1, 2, 3 and 4.

Based on these simulations, we can draw the following conclusions:

- The three methods, NW, NN and LC, perform similarly for the models **M1**–**M3**;
- Concerning model **M1**, the errors (RMSE and ARE) increase as $|\rho|$ decreases. This is expected since the estimation is much more difficult when ρ is close to 0 where a bias in the estimation appears. Also the errors increase in general when $\gamma(\cdot)$ increases;
- Concerning model **M2**, both RMSE and ARE increase with θ_2, i.e., when $\gamma(\cdot) = -1/\theta_2(\cdot)$ increases, and decreases with θ_1. Compared to the model **M1**, the RMSE and ARE are considerably smaller, but this is not surprising since the conditional extreme value index is negative in model **M2**, which means that the observations are bounded;
- Concerning model **M3**, RMSE and ARE are not too much sensitive on the values of σ, nor on x_0. In general, the orders of the errors are intermediate between those obtained in the case $\gamma(\cdot) > 0$ (model **M1**) and $\gamma(\cdot) < 0$ (model **M2**);

Table 1 RMSE (first line) and ARE (second line) of $\widehat{Q}_n(20/n|x_0)$ based on 500 samples of size $n = 1000$ according to the model **M1**, for three different values of ρ and x_0 and three different weights: Nadaraya–Watson (NW), nearest neighbors (NN) and linear combination of both (LC)

	$\rho = -2$			$\rho = -1$			$\rho = -0.5$		
	NW	NN	LC	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$	0.20	0.18	0.20	0.21	0.18	0.20	0.22	0.19	0.21
$\gamma(x_0^{(1)}) = 1/3$	0.15	0.14	0.15	0.15	0.14	0.15	0.16	0.15	0.16
$x_0 = x_0^{(2)}$	0.28	0.28	0.31	0.29	0.30	0.31	0.33	0.34	0.35
$\gamma(x_0^{(2)}) = 1/2$	0.20	0.20	0.21	0.21	0.21	0.21	0.23	0.24	0.23
$x_0 = x_0^{(3)}$	0.20	0.20	0.20	0.20	0.21	0.21	0.20	0.20	0.22
$\gamma(x_0^{(3)}) = 1/4$	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16

Table 2 RMSE (first line) and ARE (second line) of $\widehat{Q}_n(20/n|x_0)$ based on 500 samples of size $n = 1000$ according to the model **M2**, for three different values of θ_1 and x_0 and three different weights: Nadaraya–Watson (NW), nearest neighbors (NN) and linear combination of both (LC)

	$\theta_1 = 1$			$\theta_1 = 2$			$\theta_1 = 3$		
	NW	NN	LC	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$	0.07	0.07	0.07	0.04	0.04	0.05	0.03	0.03	0.03
$\theta_2(x_0^{(1)}) = 3$	0.06	0.06	0.06	0.04	0.04	0.04	0.03	0.03	0.03
$x_0 = x_0^{(2)}$	0.04	0.04	0.04	0.03	0.02	0.03	0.02	0.02	0.02
$\theta_2(x_0^{(2)}) = 2$	0.03	0.03	0.04	0.02	0.02	0.02	0.01	0.01	0.01
$x_0 = x_0^{(3)}$	0.10	0.10	0.10	0.07	0.07	0.07	0.05	0.05	0.06
$\theta_2(x_0^{(3)}) = 4$	0.08	0.08	0.08	0.05	0.06	0.06	0.04	0.04	0.04

Table 3 RMSE (first line) and ARE (second line) of $\widehat{Q}_n(20/n|x_0)$ based on 500 samples of size $n = 1000$ according to the model **M3**, for three different values of σ and x_0 and three different weights: Nadaraya–Watson (NW), nearest neighbors (NN) and linear combination of both (LC)

	$\sigma = 1/2$			$\sigma = 1$			$\sigma = 3/2$		
	NW	NN	LC	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.09	0.08
$\mu(x_0^{(1)}) = 1/3$	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06
$x_0 = x_0^{(2)}$	0.06	0.07	0.07	0.07	0.07	0.07	0.08	0.08	0.08
$\mu(x_0^{(2)}) = 1/2$	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06
$x_0 = x_0^{(3)}$	0.08	0.09	0.08	0.08	0.09	0.07	0.08	0.09	0.07
$\mu(x_0^{(3)}) = 1/4$	0.06	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06

Table 4 RMSE (first line) and ARE (second line) of $\widehat{Q}_n(20/n|x_0)$ based on 500 samples of size $n = 1000$ according to the model **M4**, for three different values of ξ and x_0 and three different weights: Nadaraya–Watson (NW), nearest neighbors (NN) and linear combination of both (LC)

	$\xi = 1/2$			$\xi = 1$			$\xi = 3/2$		
	NW	NN	LC	NW	NN	LC	NW	NN	LC
$x_0 = x_0^{(1)}$	17.3	19.1	17.1	1.04	1.29	1.01	0.48	0.50	0.45
$\theta(x_0^{(1)}) \approx 0.871$	4.02	4.75	3.93	0.58	0.65	0.58	0.32	0.32	0.30
$x_0 = x_0^{(2)}$	6.98	27.3	25.3	0.96	1.26	1.22	0.47	0.53	0.50
$\theta(x_0^{(2)}) = 0.95$	1.91	3.07	2.81	0.49	0.54	0.50	0.28	0.30	0.27
$x_0 = x_0^{(3)}$	392	295	1414	2.07	2.34	3.91	0.55	0.63	0.91
$\theta(x_0^{(3)}) = 0.83125$	35.0	30.2	102	0.81	0.88	0.97	0.37	0.39	0.41

– Concerning model **M4**, RMSE and ARE depend a lot on the value of ξ. Indeed, if ξ is too small, both RMSE and ARE increase drastically and in that case the variability of the results is probably too large to allow a more precise interpretation of the results. For larger values of ξ ($\xi = 1$ or $3/2$), the errors are more reasonable, although larger than for the others models. In that case, a slight increase in $\theta(\cdot)$ implies in general a decrease in RMSE and ARE.

To complete the simulation study, we compare in Fig. 1 the boxplots of the estimates of $Q(\alpha_n)$ with the three weights (NW, NN and LC) for model **M1** when $\rho = -1/2$, which corresponds to a difficult case, and $x_0 = x_0^{(3)}$. The horizontal line indicates the true value of the conditional quantile. As is clear from this figure, the three methods perform similarly and well, with almost no bias and a sampling distribution of the estimates symmetric. Since the boxplots for the other considered cases (model and values of x_0 and parameters) are similar, they are omitted from the paper.

6 Real data analysis

As an illustration, we consider in this section the world catalogue of earthquake from 2002 until 2017 which contains information such as the longitude, latitude and seismic moment of earthquakes. The seismic moment denoted by M_S is a physical quantity which illustrates the severity of an earthquake. It is a measure of the energy released by a seism and whose unit is the dyne-centimeters. The dataset considered in this section, of size 15000, is part of the Global Centroid Moment Tensor database, which can be uploaded freely on http://www.globalcmt.org/CMTsearch.html (Dziewonski et al. 1981; Ekström et al. 2012). Note also that this database has already been used in the extreme value framework, but on different periods, by Goegebeur et al. (2014, 2017). Being able to model accurately the tail of the earthquake energy distribution is clearly of interest since severe earthquakes may cause important damage and serious losses.

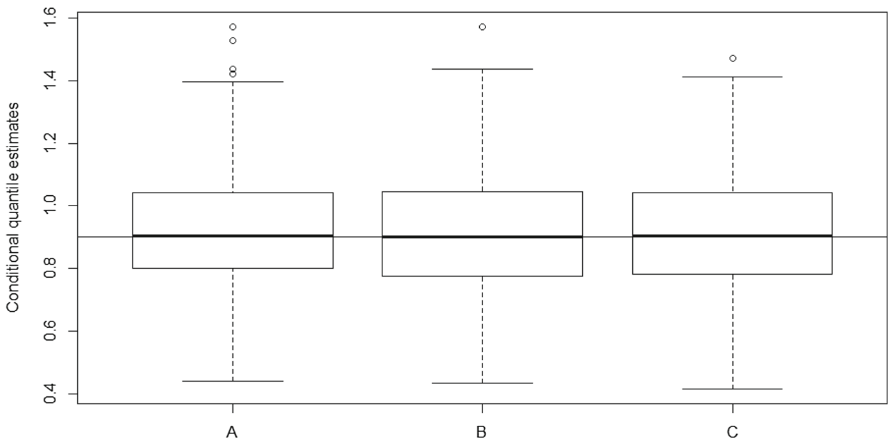

Fig. 1 Boxplots of $\widehat{Q}_n^{\mathrm{NW}}(\alpha_n|x_0)$ (A), $\widehat{Q}_n^{\mathrm{NN}}(\alpha_n|x_0)$ (B) and $\widehat{Q}_n^{\mathrm{LC}}(\alpha_n|x_0)$ (C) for the model **M1** when $\rho = -1/2$, $x_0 = (1+\sqrt{1/2})/2$, $\alpha_n = 20/n$ and $n = 1000$. The horizontal line indicates the true value of the conditional quantile

Although we want to study the tail behavior at a specific, fixed, location, the extreme conditional quantiles estimates have to take into account that earthquakes happen at a random location. Thus, this dataset is particularly suited for illustration of our local estimation method. Note that the scientists prefer to convert the seismic moment M_S into the magnitude moment M_W, defined as

$$M_W = \frac{2}{3}\log_{10}(M_S) - \frac{32}{3}$$

which is a dimensionless value. A value $M_W > 9$ indicates an extreme earthquake which may cause severe damages and losses, whereas a value $M_W < 6$ corresponds to a moderated one. Our interest is thus on the distribution of M_W given the location (in latitude and longitude) of the earthquake. The five-number summary of M_W is given below:

Min.	1st Qu.	Median	3rd Qu.	Max.
5.224	5.617	5.778	6.052	9.75

It appears that between 2002 and 2017, approximately 75% of the earthquakes can be classified as *moderate*. Concerning the points in the covariate space where we want to do our estimation, we use locations where an earthquake has already happened. In order to determine the neighborhood of these locations, we compute the distance in kilometers to every other earthquake position using the formula

$$R\ \mathrm{Arcos}\left(\cos(\phi_1)\cos(\phi_2)\cos(\phi_1-\phi_2)+\sin(\psi_1)\sin(\psi_2)\right),$$

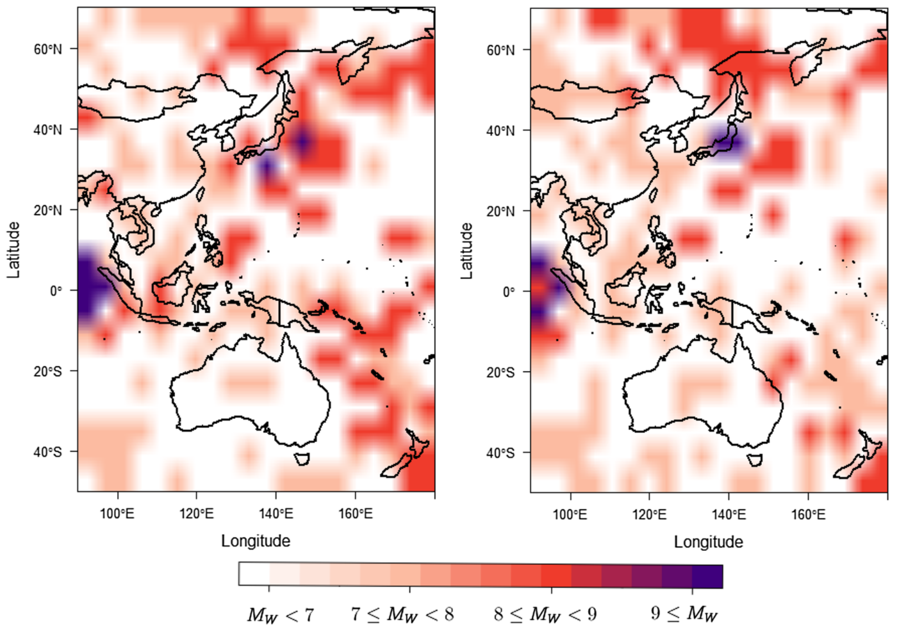

Fig. 2 Level plot of the conditional extreme quantiles of order $20/15,000$ in the Asia–Pacific region with NW weights (left panel) and NN weights (right panel)

which gives the spherical distance between two points with longitude and latitude (ϕ_1, ψ_1) and (ϕ_2, ψ_2), respectively, expressed in radian (see, e.g., Weisstein 2003). Here, it is assumed that the earth is a perfect sphere, with radius $R = 6371$km.

We estimate the extreme quantile of level $\alpha_n = 20/15000$, and the hyperparameters are selected as described in Sect. 5. The same grid as the one used in Goegebeur et al. (2014), i.e., $\mathcal{H} = \{200, 300, \ldots, 2000\}$, has been used for the NW weights, and for the NN weights, we use a grid $\mathcal{K}$ of 19 evenly spaced points between 1 and 50. Note that the LC method is not considered here since it does not outperform the others two methods as seen in Sect. 5. The level plot of our quantile estimates is given in Fig. 2 for the NW (left panel) and NN (right panel) weights, respectively. Note that this figure focuses on the Asia–Pacific region, since it is part of the well-known Ring of Fire, an area where many earthquakes and volcanic eruptions occur. The two panels of the figure are slightly different but, as expected, we can observe in both level plots that the seismic activity is intense, especially in Japan and Thailand where we can observe earthquakes with magnitude moment beyond 9. Finally, among all extreme quantile estimates of level $20/15000$ calculated with NW weights (resp. NN weights), we have a proportion of 1.5% (resp. 1.25%) for which $M_W > 9$ and 60.75% (resp. 61.25%) for which $M_W < 7$.

7 Proofs

7.1 Proof of the results given in Sect. 3

Proof of Proposition 1 1. Since S is decreasing and ϕ^{-1} is a continuous function, statement 1. is a direct consequence of Dini's theorem.

2. It suffices to remark that from the first statement, one has for all $t_0 \in I$,

$$\lim_{(t,x)\to(t_0,x^*)} \Psi(x)\left(\frac{S[x+td(x)]}{S(x)}-1\right) = \lim_{t\to t_0} \phi^{-1}(t).$$

□

Proof of Proposition 2 We first prove that condition (9) implies condition (8). From de Haan and Ferreira (2006, Lemma 1.1.1), one has, for all $t \in I$,

$$\lim_{x\to x^*} \frac{S[x+ta(1/S(x))]-S(x)}{g[S(x)]} = \phi^{-1}(t).$$

Taking $a_n = d(Q(1/n))$, $b_n = Q(1/n)$ and $c_n = \Psi(Q(1/n))$, we easily show that 1. ⇒ 3.

Finally, let us prove that 3. ⇒ 2. From de Haan and Ferreira (2006, Lemma 1.1.1), we have that for all $t \in J$,

$$\lim_{n\to\infty} \frac{Q[n^{-1}(1+tc_n)]-b_n}{a_n} = \phi(t). \tag{28}$$

Hence, since Q is decreasing and $\lfloor\alpha^{-1}\rfloor \le \alpha^{-1} < \lfloor\alpha^{-1}\rfloor + 1$,

$$Q\left(\frac{1+tc_{\lfloor\alpha^{-1}\rfloor}}{\lfloor\alpha^{-1}\rfloor}\right) \le Q[\alpha(1+tc_{\lfloor\alpha^{-1}\rfloor})] \le Q\left(\frac{1+tc_{\lfloor\alpha^{-1}\rfloor}}{\lfloor\alpha^{-1}\rfloor+1}\right). \tag{29}$$

Using (28), we know that

$$\frac{1}{a_{\lfloor\alpha^{-1}\rfloor}}\left[Q\left(\frac{1+tc_{\lfloor\alpha^{-1}\rfloor}}{\lfloor\alpha^{-1}\rfloor}\right)-b_{\lfloor\alpha^{-1}\rfloor}\right] \to \phi(t). \tag{30}$$

Moreover,

$$Q\left[\left((\lfloor\alpha^{-1}\rfloor+1)^{-1}(1+tc_{\lfloor\alpha^{-1}\rfloor}\right)\right] = Q\left\{\lfloor\alpha^{-1}\rfloor^{-1}\left[1+tc_{\lfloor\alpha^{-1}\rfloor}\xi_t(\lfloor\alpha^{-1}\rfloor)\right]\right\},$$

where for all $m \in \mathbb{N}$,

$$\xi_t(m) := \frac{m}{1+m}\left(1-\frac{1}{tmc_m}\right).$$

Since $mc_m \to \infty$, we have $\xi_t(m) \to 1$ as $m \to \infty$. Dini's theorem together with (28) entail that

$$\frac{1}{a_{\lfloor \alpha^{-1} \rfloor}} \left[Q\left(\frac{1 + tc_{\lfloor \alpha^{-1} \rfloor}}{\lfloor \alpha^{-1} \rfloor + 1} \right) - b_{\lfloor \alpha^{-1} \rfloor} \right] \to \phi(t). \tag{31}$$

Hence, by collecting (29), (30) and (31) we obtain

$$\frac{Q[\alpha + tg(\alpha)] - b(\alpha)}{a(\alpha^{-1})} \to \phi(t), \tag{32}$$

with $g(\alpha) = \alpha c_{\lfloor \alpha^{-1} \rfloor}$, $b(\alpha) = b_{\lfloor \alpha^{-1} \rfloor}$ and $a(\alpha^{-1}) = a_{\lfloor \alpha^{-1} \rfloor}$. Using twice, convergence (32) yields

$$\frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(\alpha^{-1})} \to \phi(t) - \phi(0) = \phi(t).$$

□

Proof of Proposition 3 It suffices to remark that

$$\frac{Q[\alpha + tg(\alpha)] - Q(\alpha)}{a(\alpha^{-1})} = \frac{Q'(\alpha)g(\alpha)}{a(\alpha^{-1})} \int_0^t \frac{Q'[\alpha + sg(\alpha)]}{Q'(\alpha)} \mathrm{d}s.$$

The local uniform convergence (11) concludes the proof. □

Proof of Proposition 4 From Proposition 2, the TFO condition entails that $nc_n S(a_n t + b_n) - c_n \to \phi^{-1}(t)$ as $n \to \infty$ with $c_n = 1/[ng(1/n)]$, $a_n = a(n)$ and $b_n = Q(1/n)$. First assume that $\alpha/g(\alpha) \to c$ as $\alpha \to 0$ with $c > 0$. We have that $c_n \to c > 0$ as $n \to \infty$ and thus $nS(a_n t + b_n) \to 1 + \phi^{-1}(t)/c$. In particular, we have that $S(a_n t + b_n) \to 0$ and thus that, letting $F := 1 - S$, $-nS(a_n t + b_n) \sim \ln F^n(a_n t + b_n)$ as $n \to \infty$. Hence,

$$\lim_{n \to \infty} F^n(a_n t + b_n) = G(t) = \exp\left[-\left(1 + \frac{\phi^{-1}(t)}{c} \right) \right],$$

showing that condition (12) is satisfied. Now, let us consider the case $c = 0$. From Proposition 2, we have $nc_n S(a_n t + b_n) \to \phi^{-1}(t)$. Let $m_n := nc_n = 1/g(1/n) =: \tilde{g}(n)$. Since $g(\alpha) \to 0$ as $\alpha \to 0$, $m_n \to \infty$ as $n \to \infty$. Since g is a continuous and increasing function, we have that $\tilde{g}^{-1}(m) \to \infty$ as $m \to \infty$. Letting $\tilde{a}_m := a_{\tilde{g}^{-1}(m)}$ and $\tilde{b}_m := b_{\tilde{g}^{-1}(m)}$, we obtain the convergence

$$\lim_{m \to \infty} mS(\tilde{a}_m t + \tilde{b}_m) = \phi^{-1}(t).$$

The end of the proof is similar to the one in the case $c > 0$. □

Proof of Proposition 5 Let us show that (16) implies (15), the converse being similar. Let $g(\alpha) = \alpha^2 b(\alpha^{-1})$. Since $g(\alpha)/\alpha \to 0$ as $\alpha \to 0$, one has for all $t \in \mathbb{R}$

$$\Delta(\alpha, t) := \frac{\alpha}{g(\alpha)} \left[\left(1 + t \frac{g(\alpha)}{\alpha} \right)^{-1} - 1 \right] \to -t,$$

as $\alpha \to 0$. Hence,

$$\frac{Q[\alpha + tg(\alpha)]}{Q(\alpha)} = \frac{U[\alpha^{-1} + b(\alpha^{-1})\Delta(\alpha, t)]}{U(\alpha^{-1})}.$$

From Dini's theorem, convergence (16) is locally uniform leading to (15). □

7.2 Proof of Theorem 1

As explained in Sect. 4.2, the asymptotic normality of the conditional survival estimator is established in two steps: a) prove the asymptotic normality of the variance term and b) show that the bias term is negligible. These two steps are based on technical results given below, and whose proofs are postponed to the supplementary material.

The first step is a direct consequence of the following lemma.

Lemma 1 *Let $\{V_{n,1}, V_{n,2}, \ldots, V_{n,n}\}$ be a triangular array of independent copies of a centered random variable V_n. Assume that $\mathbb{E}(V_n^2) = 1$ and $\mathbb{E}(|V_n|^3) < \infty$. Let $\mathcal{T}_n := \{T_{n,i},\ 1 \le i \le n\}$ be a triangular array of positive random variables independent of the $V_{n,i}$ and such that $T_{n,1}^2 + \ldots + T_{n,n}^2 = 1$.*
For $T_n := \max\{T_{n,i},\ 1 \le i \le n\}$, if $\mathbb{E}(|V_n|^3)T_n \xrightarrow{a.s.} 0$ then

$$\sum_{i=1}^{n} T_{n,i} V_{n,i} \xrightarrow{d} \mathcal{N}(0, 1).$$

We can now establish the asymptotic normality of the variance term. Let $\sigma_n^2(x_0) := S(y_n(x_0)|x_0)[1 - S(y_n(x_0)|x_0)]$ and recall that $m_n(x_0)$ is a sequence such that $n_{x_0}/m_n(x_0) \xrightarrow{a.s.} 1$ and that $v_n^2(x_0) = m_n(x_0)/S(y_n(x_0)|x_0)$.

Proposition 7 *For $x_0 \in \mathbb{R}^p$, let $y_n(x_0)$ be a sequence converging to the right endpoint $y^*(x_0)$ of the conditional distribution of Y given $\{X = x_0\}$. If condition (21) holds then $v_n(x_0)\left(\widehat{S}_n^{x_0}(y_n(x_0)) - S(y_n(x_0)|x_0)\right) \xrightarrow{d} \mathcal{N}(0, 1)$.*

The second step of the proof is treated in the following result.

Proposition 8 *Let $x_0 \in \mathbb{R}^p$ and $y_n(x_0)$ be a sequence converging to the right endpoint $y^*(x_0)$ of the conditional distribution of Y given $\{X = x_0\}$. If condition (22) holds then $v_n(x_0)\left(\widehat{S}_n(y_n(x_0)|x_0) - \widehat{S}_n^{x_0}(y_n(x_0))\right) \xrightarrow{\mathbb{P}} 0$.*

Theorem 1 is thus proved by gathering Propositions 7 and 8. □

7.3 Proof of Theorem 2

The proof follows the lines described in Sect. 2. Let us introduce the sequences $t_n^{-1}(x_0) := -v_n(x_0)g(\alpha_n)$ and $\sigma_n^{-1}(x_0) = a(\alpha_n^{-1})t_n(x_0)$. It is easy to check that for all $z \in \mathbb{R}$,

$$\mathbb{P}\left\{\sigma_n(x_0)[\widehat{Q}_n(\alpha_n|x_0) - Q(\alpha_n|x_0)] \leq z\right\} = \mathbb{P}\left\{Z_n(x_0) \leq z_n(x_0)\right\},$$

where $y_n(x_0) := Q(\alpha_n|x_0) + \sigma_n^{-1}(x_0)z$, $z_n(x_0) = v_n(x_0)[\alpha_n - S(y_n(x_0)|x_0)]$ and $Z_n(x_0) := v_n(x_0)[\widehat{S}_n(y_n(x_0)|x_0) - S(y_n(x_0)|x_0)]$. From Proposition 1, condition **(H)** entails that for all $t_0 \in I$,

$$\lim_{(t,y)\to(t_0,y^*(x_0))} \frac{\Psi(y)}{t}\left(\frac{S[y+td(y)|x_0]}{S(y|x_0)} - 1\right) = \lim_{t\to t_0} \frac{\phi^{-1}(t)}{t}. \tag{33}$$

Since $y_n(x_0) = Q(\alpha_n|x_0) + a(\alpha_n^{-1})t_n(x_0)z = Q(\alpha_n|x_0) + d(Q(\alpha_n|x_0))t_n(x_0)z$ with $t_n(x_0) \to 0$ as $n \to \infty$, (33) entails that as $n \to \infty$

$$z_n(x_0) \sim -zv_n(x_0)t_n(x_0)g(\alpha_n) = z. \tag{34}$$

Now, to prove that $Z_n(x_0) \xrightarrow{d} \mathcal{N}(0,1)$, it suffices to show that conditions (21) and (22) hold for $y_n(x_0)$. From (34),

$$1 - \frac{S[y_n(x_0)|x_0]}{\alpha_n} \sim z\alpha_n^{-1}v_n^{-1}(x_0) = z(\alpha_n m_n(x_0))^{-1/2} \to 0, \tag{35}$$

as $n \to \infty$ and thus $S[y_n(x_0)|x_0] \sim \alpha_n$. This entails that condition

$$v_n(x_0) \max_{1\leq i\leq n} \mathcal{W}_{n,i}(x_0) \xrightarrow{a.s.} 0$$

is equivalent to condition (21) with $y_n(x_0)$. It remains to prove condition (22). From (35), there exists $\xi > 0$ such that for n large enough, $S(y_n(x_0)|x_0) \in [(1-\xi)\alpha_n, (1+\xi)\alpha_n]$. Hence, for n large enough,

$$\sum_{i=1}^n \mathcal{W}_{n,i}(x_0)\left|\frac{S(y_n(x_0)|X_i)}{S(y_n(x_0)|x_0)} - 1\right| \leq \sup_{|\beta/\alpha_n - 1|\leq\xi} \sum_{i=1}^n \mathcal{W}_{n,i}(x_0)\left|\frac{S[Q(\beta|x_0)|X_i]}{\beta} - 1\right|,$$

and the proof is complete. □

7.4 Proof of Corollaries 1, 2 and 3

We first recall a useful result dealing with the almost sure convergence of the statistic

$$\hat{f}_n(x) := \frac{1}{nh_n^p}\sum_{i=1}^n K\left(\frac{X_i - x_0}{h_n}\right),$$

which is the kernel estimator of the density f of the random value X. The following result can be found for instance in Dony and Einmahl (2009, Corollary 2.1).

Lemma 2 *Let $x \in \mathbb{R}^p$ such that f is continuous at x and $f(x) > 0$. If the kernel K is a bounded density with support included in the unit ball $\mathcal{U}_p$ of $\mathbb{R}^p$ and if $\mathcal{K} := \{K(\gamma(t - \cdot)),\ \gamma > 0,\ t \in \mathbb{R}^p\}$, is a pointwise measurable Vapnik–Chervonenkis (VC) type class of functions from $\mathbb{R}^p$ to $\mathbb{R}$ then for a sequence $h_n \to 0$ such that $nh_n^p / \log\log n \to \infty$, we have that $\hat{f}_n(x) \xrightarrow{a.s.} f(x)$.*

Conditions on the family $\mathcal{K}$ of functions are not easy to check in practice. Nevertheless, the measurability condition on $\mathcal{K}$ is satisfied whenever K is right-continuous (see Einmahl and Mason 2005) or K is an indicator function on a cell of $\mathbb{R}^p$ (see van der Vaart and Wellner 1996, Example 2.3.4). Concerning the VC condition, it is satisfied for kernel function K such that $K(x) = L(\|x\|)$ where L is of bounded variation (see Giné and Nickl 2015, Exercice 3.6.13). For the sake of simplicity, we have preferred to replace in Lemma 2 all the conditions involving the kernel function by the stronger (but simpler to check) condition **(K)**.

Corollaries 1, 2 and 3 are direct consequences of Theorem 2 and of the three following lemmas establishing the asymptotic distribution of the corresponding conditional survival function estimators, whose proofs are given in the supplementary material.

Lemma 3 *Let $x_0 \in \mathbb{R}^p$ such that f is continuous at x_0 and $f(x_0) > 0$ and let K be a kernel satisfying* **(K)**. *For sequences $h_n \to 0$ and $y_n(x_0) \uparrow y^*(x_0)$ such that $nh_n^p[S(y_n(x_0)|x_0) \wedge (\log\log n)^{-1}] \to \infty$ and*

$$\sup_{\|x-x_0\|\le h_n} \left| \frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1 \right|^2 = o\left(\frac{1}{nh_n^p S(y_n(x_0)|x_0)} \right),$$

one has

$$\left(nh_n^p S(y_n(x_0)|x_0)\right)^{1/2} \left(\frac{\widehat{S}_n^{\mathrm{NW}}(y_n(x_0)|x_0)}{S(y_n(x_0)|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\|K\|_2^2}{f(x_0)}\right).$$

Lemma 4 *Let $x_0 \in \mathbb{R}^p$. For sequences k_n and $y_n(x_0)$ such that, as $n \to \infty$, $y_n(x_0) \uparrow y^*(x_0)$, $k_n S(y_n(x_0)|x_0) \to \infty$ and*

$$\sup_{\|x-x_0\|\le D_{(k_n)}(x_0)} \left| \frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1 \right|^2 = o\left(\frac{1}{k_n S(y_n(x_0)|x_0)} \right),$$

with $D_{(k_n)}(x_0) = \|X_{r(k_n)} - x_0\|$, one has

$$(k_n S(y_n(x_0)|x_0))^{1/2} \left(\frac{\widehat{S}_n^{\mathrm{KNN}}(y_n(x_0)|x_0)}{S(y_n(x_0)|x_0)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{(\ell+1)^2}{2\ell+1}\right).$$

Lemma 5 *Let $x_0 \in \mathbb{R}^p$ such that f is continuous at x_0 and $f(x_0) > 0$. Let h_n, k_n and $y_n(x_0) \uparrow y^*(x_0)$ be sequences such that $nh_n^p/\log\log n \to \infty$, $\ell_n S(y_n(x_0)|x_0) \to \infty$ with $\ell_n := (nh_n^p \wedge k_n)$ and*

$$\sup_{\|x-x_0\|\le(h_n\vee D_{(k_n)}(x_0))}\left|\frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)}-1\right|^2 = o\left(\frac{1}{\ell_n S(y_n(x_0)|x_0)}\right).$$

If there exists $\kappa \in [0,\infty]$ such that $k_n/(nh_n^p) \to \kappa$ then

$$(\ell_n S(y_n(x_0)|x_0))^{1/2}\left(\frac{\widehat{S}_n^{\mathrm{LC}}(y_n(x_0)|x_0)}{S(y_n(x_0)|x_0)}-1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{C^2(\kappa)}{2^p f(x_0)}\right),$$

where $C^2(\kappa) := (1\wedge\kappa^{-1})\left[\kappa\tau^2 + 2^p f(x_0)(1-\tau)^2 + 2\tau(1-\tau)\,(\kappa \wedge 2^p f(x_0))\right]$.

□

7.5 Proof of Proposition 6

Proposition 6 is a consequence of the following lemma.

Lemma 6 *Let $(X, Y, Z)^\top$ be a random vector for which (X, Y) and Z are independent. Let g be a measurable function such that $g(X, Y, Z)$ is integrable. One has $\mathbb{E}[g(X, Y, Z)] = \mathbb{E}[\Psi(X, Z)]$, where $\Psi(x, z) := \mathbb{E}[g(x, Y, z)|X = x]$.*

Proof of Proposition 6 First remark that the assumption on the weights entails that the $W_{n,i}(x_0)$ are identically distributed. Furthermore, since the $W_{n,i}(x_0)$ sum to 1, it is clear that $\mathbb{E}[W_{n,1}(x_0)] = \ldots = \mathbb{E}[W_{n,n}(x_0)] = 1/n$. It thus remains to show that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \mathcal{W}_{n,i}(x_0)\mathbb{I}_{\{Y_i>\widehat{Q}_{n,-i}(\alpha_n|X_i,\lambda)\}}\right] = \mathbb{E}\left[\mathcal{W}_{n,1}(x_0)S[\widehat{Q}_{n,-1}(\alpha_n|X_1,\lambda)|X_1]\right].$$

We apply Lemma 6 with $X = X_1$, $Y = Y_1$, $Z = \mathbb{X}_{-1}$ and $g(t, y, u) = \varphi(t, u)\mathbb{I}_{\{y>\phi(\alpha_n,t,u)\}}$ where the function ϕ is such that

$$\widehat{Q}_{n,-1}(\alpha_n|X_1,\lambda) = \phi(\alpha_n, X_1, \mathbb{X}_{-1}).$$

The conclusion is straightforward since, with the notation of Lemma 6, $\Psi(t, u) = \varphi(t, u)S(\phi(\alpha_n, t, u)|t)$. □

Acknowledgements The authors would like to thank the reviewers, the associate editor and editor for their helpful comments and suggestions that led to substantial improvement of the paper.

References

Berlinet, A., Gannoun, A., Matzner-Lober, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles. *Statistics—A Journal of Theoretical and Applied Statistics*, *35*(2), 139–169.

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

de Haan, L., Ferreira, A. (2006). *Extreme value theory*: *An introduction*., Springer series in operations research and financial engineering New York: Springer.

Dony, J., Einmahl, U. (2009). Uniform in bandwidth consistency of kernel regression estimators at fixed point. In *High dimensional probability V*: *The Luminy volume* (pp. 308–325). Beachwood, OH: Institute of Mathematical Statistics.

Dziewonski, A. M., Chou, T. A., Woodhouse, J. H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research*, *86*, 2825–2852.

Ekström, G., Nettles, M., Dziewonski, A. M. (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, *200–201*, 1–9.

Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling extremal events*., For insurance and finance Berlin: Springer.

Einmahl, U., Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, *33*(3), 1380–1403.

Fisher, R. A., Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190.

Fraga Alves, I., de Haan, L., Neves, C. (2009). A test procedure for detecting super-heavy tails. *Journal of Statistical Planning and Inference*, *139*, 213–227.

Giné, E., Nickl, R. (2015). *Mathematical foundations of infinite-dimensional statistical models*. New York: Cambridge University Press.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, *44*(3), 423–453.

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

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

He, X., Ng, P. (1999). Quantile splines with several covariates. *Journal of Statistical Planning and Inference*, *75*, 343–352.

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Koenker, R., Ng, P., Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, *81*, 673–680.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, *9*, 141–142.

Racine, J. S., Li, K. (2017). Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach. *Journal of Econometrics*, *201*, 72–94.

van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical processes with applications to statistics*., Springer series in statistics Berlin: Springer.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya A*, *26*, 359–372.

Weisstein, E. W. (2003). *CRC concise encyclopedia of mathematics*2nd ed. Boca Raton: Chapman and Hall.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.