



# Nonparametric confidence interval for conditional quantiles with large-dimensional covariates

Laurent Gardes

## ► To cite this version:

Laurent Gardes. Nonparametric confidence interval for conditional quantiles with large-dimensional covariates. 2019. hal-02134882v1

**HAL Id: hal-02134882**

**<https://hal.science/hal-02134882v1>**

Preprint submitted on 20 May 2019 (v1), last revised 15 Oct 2019 (v2)

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

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

# Nonparametric confidence interval for conditional quantiles with large-dimensional covariates

Laurent Gardes

Université de Strasbourg & CNRS, IRMA, UMR 7501,  
7 rue René Descartes, 67084 Strasbourg Cedex, France

## Abstract

The first part of the paper is dedicated to the construction of a  $\gamma$  - nonparametric confidence interval for a conditional quantile of level depending on the sample size. When this level converges to 0 or 1 as the sample size increases, the conditional quantile is said to be extreme and is located in the tail of the conditional distribution. The proposed confidence interval is constructed by approximating the distribution of ordered statistics selected with a nearest neighbor approach by a Beta distribution. We show that its coverage probability converges to the preselected probability  $\gamma$  and its accuracy is illustrated on a simulation study. When the dimension of the covariate increases, the coverage probability of the confidence interval can be very different from  $\gamma$ . This is a well known consequence of the data sparsity especially in the tail of the distribution. In the second part of the paper, a dimension reduction procedure is proposed in order to select more appropriate nearest neighbors and in turn to obtain a better coverage probability. This procedure is based on the Tail Conditional Independence assumption introduced in (Gardes, *Extreme*, pp. 57–95, **18(3)**, 2018).

**Keywords:** Extreme conditional quantiles, confidence interval, dimension reduction.

## 1 Introduction

In a large range of applications, it is necessary to examine the effects of an observable  $\mathbb{R}^p$ -valued random covariate  $X$  on the distribution of a dependent  $\mathbb{R}$ -valued variable  $Y$ . For instance,  $Y$  can modelize the level of ozone in the air and  $X$  the vector gathering the concentration of others pollutants and weather conditions (see *e.g.* Han *et al.* [15]). A first approach consists to focus on the conditional expectation  $\mathbb{E}(Y|X)$ . A more comprehensive way is to analyze the conditional quantile of  $Y$  given  $X$ . Recall that for all  $x \in \mathcal{X} \subset \mathbb{R}^p$ , ( $\mathcal{X}$  being

the support of  $X$ ), the  $(1-\alpha)$ -conditional quantile of  $Y$  given  $X = x$  where  $\alpha \in (0, 1)$  is  $Q(\alpha|x) = \inf\{y; S(y|x) \leq \alpha\}$ , where  $S(\cdot|x) := \mathbb{P}(Y > \cdot | X = x)$  is the conditional survival function of  $Y$  given  $X = x$ . In this paper, we are interested in the construction of a confidence interval of  $Q(\alpha|x_0)$  for a given point  $x_0 \in \mathcal{X}$ . More specifically, based on  $n$  independent copies  $(X_1, Y_1), \dots, (X_n, Y_n)$  of the random vector  $(X, Y)$ , we searching for a random interval  $[A_{n,\gamma}(x_0), B_{n,\gamma}(x_0)]$  such that, as  $n \rightarrow \infty$ ,

$$\mathbb{P}\{[A_{n,\gamma}(x_0), B_{n,\gamma}(x_0)] \ni Q(\alpha|x_0)\} \rightarrow \gamma, \quad (1)$$

where  $\gamma \in (0, 1)$  is a preselected probability (usually,  $\gamma = 0.9$  or  $0.95$ ). To allow us to make inference in the right and left-tails of the conditional distribution, we also consider the case where  $\alpha$  depends on the sample size  $n$  and converges to 0 or 1 as  $n \rightarrow \infty$ . In the application on ozone concentration, this can be of high interest since large level of ozone in the air may cause serious effects on public health and on environment.

The literature on the estimation of confidence interval for quantile is, up to our knowledge, only dedicated to the case where  $\alpha$  is a fixed value in  $(0, 1)$ . Several approaches have been considered. The first one, called direct approach (see *e.g.* Fan and Liu [6]), consists to construct the confidence interval starting from a given estimator  $\hat{Q}_n(\alpha|x_0)$  of  $Q(\alpha|x_0)$ . Under some regularity assumptions on  $Q(\cdot|x_0)$  (in particular the derivative  $q(\cdot|x_0)$  of  $Q(\cdot|x_0)$  must exists), if one can find a positive sequence  $(c_n)$  such that  $c_n(\hat{Q}_n(\alpha|x_0) - Q(\alpha|x_0))$  converges to a centered gaussian distribution with variance  $q^2(\alpha|x_0)$  then, denoting by  $\Phi$  the cumulative distribution function on a centered and normalized gaussian distribution, the coverage probability of the interval

$$\left[ \hat{Q}_n(\alpha - c_n^{-1}\Phi^{-1}((1-\gamma)/2)|x_0), \hat{Q}_n(\alpha + c_n^{-1}\Phi^{-1}((1-\gamma)/2)|x_0) \right],$$

converges to  $\gamma$ . There is a huge literature on the estimation of conditional quantile for  $\alpha$  fixed (see for example Yang [27] and Stute [25]) but also for extreme conditional quantile (*i.e.* when  $\alpha = \alpha_n \rightarrow 0$ , see for instance Gardes [8], Gardes *et al.* [13], *etc.*). The main drawback of the direct approach is that in most of the cases, the sequence  $c_n$  depends on unknown parameters (*e.g.* the probability density function of  $X$ ) that have to be estimated. To avoid the estimation of  $c_n$ , resampling methods have been considered by Parzen *et al.* [21] and Kocherginsky *et al.* [18]. Unfortunately, these methods are often time consuming. A last approach to construct confidence interval of quantile is based on order statistics. The order statistic method has been first introduced in the unconditional case (see Hutson [17]). Assume that  $Y_1, \dots, Y_n$  are independent and identically distributed random variables with common survival function  $S_Y(\cdot)$  and quantile function  $Q_Y(\cdot)$ . Denoting by  $Y_{1,n} \leq \dots \leq Y_{n,n}$  the ordered statistics, if  $S_Y(\cdot)$  is a continuous and strictly increasing function, the probability integral transform ensures that

$$\mathbb{P}(Y_{j,n} < Q_Y(\alpha)) = \mathbb{P}(S_Y(Y_{j,n}) > \alpha) = \mathbb{P}(U_{n-j+1,n} > \alpha),$$

where  $U_{1,n} \leq \dots \leq U_{n,n}$  are the order statistics associated to independent standard uniform random variables. Since  $U_{n-j+1,n}$  is distributed as a Beta distribution of parameters  $n-j+1$  and  $n$ , denoting by  $F_{\text{beta}}(\cdot; a, b)$  the distribution function of a Beta distribution with parameters  $a > 0$  and  $b > 0$  and letting

$$\mathcal{L}_\gamma(m, \alpha) := \max \left\{ j \in \{1, \dots, m\}; F_{\text{beta}}(\alpha; m-j+1, j) \leq \frac{1-\gamma}{2} \right\} \text{ and}$$

$$\mathcal{R}_\gamma(m, \alpha) := \min \left\{ j \in \{1, \dots, m\}; 1 - F_{\text{beta}}(\alpha; m-j+1, j) \leq \frac{1-\gamma}{2} \right\},$$

for  $\gamma \in (0, 1)$ ,  $m \in \mathbb{N} \setminus \{0\}$  and  $\alpha \in (0, 1)$  with the convention  $\max\{\emptyset\} = +\infty$  and  $\min\{\emptyset\} = -\infty$ , one can show that

$$\mathbb{P} \{ [Y_{\mathcal{L}_\gamma(n, \alpha), n}, Y_{\mathcal{R}_\gamma(n, \alpha), n}] \ni Q_Y(\alpha) \} \rightarrow \gamma.$$

This method of construction has been recently adapted by Goldman and Kaplan [14] to the conditional case but always for a fixed quantile level  $\alpha$ .

**The first contribution** of this paper is to adapt the order statistics method to the conditional case by using a nearest neighbors approach. Instead of using the whole sample as in the unconditional case, only the  $k_n$  closest observations to  $x_0$  are used in the order statistics method. The proposed confidence interval can be used if the quantile level  $\alpha$  depends on  $n$  and converges to 0 or 1 as  $n \rightarrow \infty$ . Constructing a confidence interval for extreme conditional quantile ( $\alpha = \alpha_n \rightarrow 0$  or 1) is often more challenging because there are fewer observations available in the tail.

The nearest neighbors strongly depends on the (pseudo)-distance in  $\mathbb{R}^p$  used to select the observations around the point of interest  $x_0$ . The euclidean distance is of course the natural choice but when  $p$  becomes large, some nearest neighbors can be located far away from the point of interest leading to confidence interval with a bad coverage probability. This is the well known curse of dimensionality phenomenon. To overcome this problem, one way is to assume the existence of a function  $g_0 : \mathbb{R}^p \rightarrow \mathbb{R}$  such that the conditional distribution of  $Y$  given  $X$  is equal to the conditional distribution of  $Y$  given  $g_0(X)$ . In other words, it is assumed that  $X$  and  $Y$  are independent conditionally on  $g_0(X)$  (in symbols  $X \perp\!\!\!\perp Y | g_0(X)$ , see for instance Basu and Pereira [2]). The dimension of the covariate is thus reduced since  $X$  can be replaced by  $g_0(X)$ . In this case, it seems preferable to use the pseudo-distance  $d_0$  defined for all  $(x, y) \in \mathbb{R}^p \times \mathbb{R}^p$  by  $d_0(x, y) = |g_0(x) - g_0(y)|$  instead of the euclidean distance in  $\mathbb{R}^p$ . A natural question now is how to find the true function  $g_0$  and therefore the most suitable distance  $d_0$ ? One common approach is to assume that  $g_0$  is linear (*i.e.* for all  $x \in \mathbb{R}^p$ ,  $g_0(x) = b_0^\top x$  where  $b_0 \in \mathbb{R}^p$ ). This corresponds to the single-index model introduced in a regression context for instance by Li [19]. This single-index structure has been

considered by Zhu *et al.* [28] for the estimation of conditional quantile when the level  $\alpha$  is fixed. Finding the distance reduces to finding the direction  $b_0$ . Its estimation has received much attention in the literature (see Li [19] for the classical Sliced Inverse Regression (SIR) method), Cook and Weisberg [4], Samarov [23], Cook and Li [3]).

**The second contribution** of this work is the proposition of a new data-driven procedure to select the appropriate distance  $d_0$  in the nearest neighbors approach in order to construct a confidence interval for an extreme conditional quantile in the right-tail of the distribution (*i.e.* when the quantile level  $\alpha = \alpha_n \rightarrow 0$  as the sample size increases). Condition  $X \perp\!\!\!\perp Y | g_0(X)$  is relaxed by assuming that  $Y$  is tail conditionally independent of  $X$  given  $g_0(X)$ , see Gardes [9]. Basically, under this condition,  $S(y|x_0)$  is equivalent, as  $y \rightarrow \infty$ , to a function depending on  $x_0$  through  $g_0(x_0)$ . Hence, inference on the extreme conditional quantile of  $Y$  given  $X$  can be achieved only by using the information brought by the reduced covariate  $g_0(X)$ . In this paper, we assume that  $g_0$  belongs to a pre-specified parametric family (not necessarily the family of linear functions). Note that assuming that  $Y$  is tail conditionally independent of  $X$  given  $g_0(X)$ , an estimator of  $g_0$  has been proposed by Gardes [9] in the particular case of a linear function. Unfortunately, the estimation procedure is computationally expensive.

The paper is organized as follows. The definition of the confidence interval for conditional extreme quantile is given in Section 2. In particular, we show that the coverage probability of the proposed confidence interval converges to the nominal one. This section corresponds to our first contribution. The second contribution is handled in Section 3 where an estimator of an appropriate distance  $d_0$  is proposed and used for the construction of a confidence interval of extreme conditional quantile. In each section, the methods are illustrated with simulated data. All the proofs are postponed to Section 5.

## 2 Confidence interval estimation

### 2.1 Definition and main result

Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be  $n$  independent copies of a random vector  $(X, Y)$ . In all what follows, we assume that the distribution of  $(X, Y)$  is absolutely continuous with respect to the Lebesgue measure. As mentioned in the introduction, for a given  $x_0 \in \mathcal{X}$  where  $\mathcal{X}$  is the support of  $X$ , our first contribution is to propose a confidence interval of the conditional quantile

$$Q(\alpha|x_0) := \inf\{y; S(y|x_0) \leq \alpha\},$$

where  $S(\cdot|x_0) = \mathbb{P}(Y > \cdot | X = x_0)$ . While much of the literature focuses on a fixed level  $\alpha$  in  $(0, 1)$ , we allow the case where  $\alpha = \alpha_n$  depend on the sample

size  $n$ . We assume that

$$\lim_{n \rightarrow \infty} \alpha_n = c \in [0, 1]. \quad (2)$$

Condition (2) with  $c \in (0, 1)$  corresponds to a classical conditional quantile. For instance, if  $\alpha_n = 1/2$ , the value  $Q(\alpha_n|x_0)$  is the conditional median of  $Y$  given  $X = x_0$ . When  $c \in \{0, 1\}$  in (2), the level is said to be extreme. If  $c = 0$  (resp.  $c = 1$ ), the conditional quantile is located in the right-tail (resp. left-tail) of the conditional distribution of  $Y$  given  $X = x_0$ .

The basic idea to construct a random interval  $[A_{n,\gamma}(x_0), B_{n,\gamma}(x_0)]$  satisfying (1) is to apply the order statistics method to observations close enough to  $x_0$ . The order statistics method to construct confidence interval has been proposed by Hutson [17] in the unconditional case and is described in the introduction. To select the observations, a nearest neighbors approach is considered. More specifically, for some (pseudo)-metric  $d$  on  $\mathbb{R}^p$ , let

$$(X_1^{(d,x_0)}, Y_1^{(d,x_0)}), \dots, (X_n^{(d,x_0)}, Y_n^{(d,x_0)})$$

be the sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  rearranged in order to have  $d(X_1^{(d,x_0)}, x_0) \leq \dots \leq d(X_n^{(d,x_0)}, x_0)$ . For  $k_n \in \{1, \dots, n\}$ , we denote by  $Y_{1,k_n}^{(d,x_0)} \leq \dots \leq Y_{k_n,k_n}^{(d,x_0)}$  the ordered statistics associated to the sample  $Y_1^{(d,x_0)}, \dots, Y_{k_n}^{(d,x_0)}$ . For a preselected probability  $\gamma \in (0, 1)$ , we propose as a confidence interval of  $Q(\alpha_n|x_0)$  the following random interval

$$\text{CI}_{\gamma,\alpha_n}(k_n, d, x_0) := [Y_{\mathcal{L}_\gamma(k_n,\alpha_n),k_n}^{(d,x_0)}, Y_{\mathcal{R}_\gamma(k_n,\alpha_n),k_n}^{(d,x_0)}], \quad (3)$$

where we recall that

$$\begin{aligned} \mathcal{L}_\gamma(k_n, \alpha_n) &:= \max \left\{ j \in \{1, \dots, k_n\}; F_{\text{beta}}(\alpha_n; k_n - j + 1, j) \leq \frac{1 - \gamma}{2} \right\} \\ \mathcal{R}_\gamma(k_n, \alpha_n) &:= \min \left\{ j \in \{1, \dots, k_n\}; \bar{F}_{\text{beta}}(\alpha_n; k_n - j + 1, j) \leq \frac{1 - \gamma}{2} \right\}, \end{aligned}$$

$F_{\text{beta}}(\cdot; a, b)$  being the distribution function of a Beta distribution with parameters  $a$  and  $b$  and  $\bar{F}_{\text{beta}} := 1 - F_{\text{beta}}$ . The confidence interval  $\text{CI}_{\gamma,\alpha_n}(k_n, d, x_0)$  is thus defined as in the unconditional case except that only the  $k_n$  nearest neighbors random variables  $Y_1^{(d,x_0)}, \dots, Y_{k_n}^{(d,x_0)}$  are used.

It remains to prove that the coverage probability of this interval converges to  $\gamma$  as  $n \rightarrow \infty$ . The accuracy of the confidence interval  $\text{CI}_{\gamma,\alpha_n}(k_n, d, x_0)$  depends on the smoothness of the function  $x \rightarrow S[Q(\alpha|x_0)|x]$ . For  $\alpha \in (0, 1)$  and  $\zeta > 0$ , we introduce the quantity

$$\omega(\alpha, \zeta) := \sup_{d(x,x_0) \leq \zeta} \left( \frac{S[Q(\alpha|x_0)|x]}{\alpha} - 1 \right)^2.$$

which is the largest deviation of the ratio  $S[Q(\alpha|x_0)|x]/S[Q(\alpha|x_0)|x_0]$  from 1 when  $x$  belongs to the ball of center  $x_0$  and radius  $\zeta$ . Note that this quantity is classically considered when dealing with conditional distribution (see for instance Daouia *et al.* [1]). In the following result, the conditions required to ensure that the coverage probability of (3) converges to  $\gamma$  are established.

**Theorem 1** *Let  $\gamma \in (0, 1)$  and  $x_0$  in the support of  $X$ . Assume that  $k_n \rightarrow \infty$  and let  $h_n > 0$  such that  $\mathbb{P}(d(X_{k_n}^{(d, x_0)}, x_0) \leq h_n) = 1$ . For a sequence of level  $\alpha_n \in (0, 1)$  satisfying (2), if  $S(\cdot|x_0)$  is continuous and strictly increasing*

$$\delta_n^2 := \frac{\ln^2(k_n)}{k_n \alpha_n (1 - \alpha_n)} \rightarrow 0, \quad (4)$$

and if

$$\eta_n^2 := \frac{k_n \alpha_n}{1 - \alpha_n} \omega(\alpha_n, h_n) \rightarrow 0, \quad (5)$$

then,  $\mathbb{P}[\text{CI}_{\gamma, \alpha_n}(k_n, d, x_0) \ni Q(\alpha_n|x_0)] = \gamma + \mathcal{O}(\delta_n) + \mathcal{O}(\eta_n) \rightarrow \gamma$ .

First note that the confidence interval given by (3) is not the only possible confidence interval of  $Q(\alpha_n|x_0)$  with asymptotic coverage probability  $\gamma$ . One can also consider the one-sided confidence intervals

$$\begin{aligned} \text{CI}_{\gamma, \alpha_n}^{(L)}(k_n, d, x_0) &:= [Y_{\mathcal{L}_{2\gamma-1}(k_n, \alpha_n), k_n}^{(d, x_0)}, \infty) \\ \text{and } \text{CI}_{\gamma, \alpha_n}^{(R)}(k_n, d, x_0) &:= (-\infty, Y_{\mathcal{R}_{2\gamma-1}(k_n, \alpha_n), k_n}^{(d, x_0)}], \end{aligned}$$

Obviously, under the conditions of Theorem 1, the coverage probabilities of these two intervals converge to  $\gamma$ .

The proof of Theorem 1 is based on the decomposition

$$\mathbb{P}[Y_{\mathcal{L}_\gamma(k_n, \alpha_n), k_n}^{(x_0)} > Q(\alpha_n|x_0)] = (1 - \gamma)/2 + B_{1,n}(\mathcal{L}_\gamma(k_n, \alpha_n)) + B_{2,n},$$

and on a similar decomposition for  $\mathbb{P}[Y_{\mathcal{R}_\gamma(k_n, \alpha_n), k_n}^{(x_0)} \leq Q(\alpha_n|x_0)]$ . This decomposition highlights two terms of error: the term  $B_{1,n}(\mathcal{L}_\gamma(k_n, \alpha_n))$  given for  $j \in \{1, \dots, k_n\}$  by

$$B_{1,n}(j) := \mathbb{P}[Y_{j, k_n}^{(x_0)} > Q(\alpha_n|x_0)] - F_{\text{beta}}(\alpha_n; k_n - j + 1, j)$$

and the quantity

$$B_{2,n} := F_{\text{beta}}(\alpha_n; k_n - \mathcal{L}_\gamma(k_n, \alpha_n) + 1, \mathcal{L}_\gamma(k_n, \alpha_n)) - (1 - \gamma)/2.$$

The first term of error is a consequence of the approximation of the distribution  $S(Y_{j, k_n}^{(x_0)}|x_0)$  by a Beta distribution. We show in the proof of Theorem 1 that

$$\max_{j=1, \dots, k_n} |B_{1,n}(j)| = \mathcal{O}(\eta_n).$$

Condition (5) ensures that  $B_{1,n}(\mathcal{L}_\gamma(k_n, \alpha_n))$  converges to 0. Note that this condition entails that  $k_n$  should be chosen not too large. In the unconditional case (*i.e.* if  $X$  and  $Y$  are independent) then  $\eta_n = B_{1,n}(j) = 0$  for all  $j$  and one can take  $k_n = n$ . Remark also that in the unconditional case, the accuracy of the confidence interval does not depend on the underlying distribution.

The second term of error is related to the behavior of the distribution function of a beta distribution. In Lemma 2, it is established that  $B_{2,n} = \mathcal{O}(\delta_n)$  and

thus  $B_{2,n} \rightarrow 0$  under condition (4). If  $c = 0$ , the rate of convergence of  $\alpha_n$  to 0 is limited by (4) (namely,  $\alpha_n \gg \ln^2(k_n)/k_n$ ). Similarly, when  $c = 1$ , one can construct an asymptotic confidence interval only if  $1 - \alpha_n \gg \ln^2(k_n)/k_n$ . It also appears that, as expected, the rate of convergence of the coverage probability can be very slow for an extreme conditional quantile.

In the next result, a sequence  $h_n$  such that  $\mathbb{P}(d(X_{k_n}^{(d,x_0)}, x_0) \leq h_n) = 1$  is proposed when  $d$  is the euclidean distance given for  $(x, y) \in \mathbb{R}^p \times \mathbb{R}^p$  by  $d_e(x, y) = [(x - y)^\top (x - y)]^{1/2}$

**Proposition 1** *Assume that the distribution of  $X$  admits  $f_X$  as a probability density function. If  $k_n/(\ln \ln n) \rightarrow \infty$  and  $n/k_n \rightarrow \infty$  then, for*

$$h_n = \left( \frac{2}{f_X(x_0)} \frac{k_n}{n} \right)^{1/p},$$

*one has  $\mathbb{P}(d_e(X_{k_n}^{(d_e,x_0)}, x_0) \leq h_n) = 1$  for  $n$  large enough.*

It thus appears that for a given value of  $k_n$ , the radius  $h_n$  increases with the dimension  $p$ . As a consequence, when  $p$  becomes large, some of the  $k_n$ -nearest neighbors can be located far away from the point of interest and the confidence interval can performs very badly. This phenomenon is well known as the “curse of dimensionality”. In Section 3, a procedure to overcome this difficulty is proposed.

## 2.2 Numerical illustration

Let us take a look at the finite sample performance of the estimated confidence interval introduced in the previous section. Using the observations of a sample  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$  driven from a random pair  $(X, Y)$ , our objective is the construction of a confidence interval with asymptotic coverage probability  $\gamma$  of the conditional quantile  $Q(\alpha|x_0)$  associated to the conditional cumulative distribution function  $S(\cdot|x_0)$ . In the estimation procedure, the nearest neighbors are selected by using the classical euclidean distance  $d_e$  in  $\mathbb{R}^p$ .

Let  $\xi : \mathbb{R} \rightarrow (0, \infty)$  defined by  $\xi(z) := 5z^2/36 + 1/4$  and let  $g_0 : \mathbb{R}^p \rightarrow \mathbb{R}$ . Two models for the distribution of  $(X, Y)$  are considered:

- **Model 1:** Conditional Burr distribution. For  $y > 0$ ,

$$S(y|X) = \left( 1 + y^{c(X)} \right)^{-1/\tau(X)},$$

where  $c$  and  $\tau$  are positive functions defined for all  $x \in \mathbb{R}^p$  by  $c(x) = \|x\|_1$  and  $\tau(x) = c(x)\xi(g_0(x))$ .

- **Model 2:** Conditional Weibull distribution. For  $y > 0$ ,

$$S(y|X) := \exp \left( -y^{1/\xi(g_0(X))} \right).$$



In all what follows, the  $p$  components of the random vector  $X$  are independent and uniformly distributed on  $[-5, 5]$  and the point of interest  $x_0$  is the vector with all its components equal to 1. In **Model 1**, the conditional distribution of  $Y$  given that  $X = x$  is heavy-tailed since for all  $t > 0$  and  $x \in \mathcal{X}$ ,

$$\lim_{y \rightarrow \infty} \frac{S(ty|x)}{S(y|x)} = t^{-\xi(g_0(x))}.$$

The conditional extreme index is the function  $\xi \circ g_0$ . The conditional distribution of  $Y$  given that  $X = x$  in **Model 2** is a conditional Weibull type distribution (see for instance [12, 11]) and  $\xi(g_0(x))$  is referred to as the conditional Weibull-tail index.

Three conditional quantile levels are considered:  $\alpha = \alpha_{1,n} = 1 - 8 \ln(n)/n$ ,  $\alpha_n = \alpha_2 = 1/2$  and  $\alpha_n = \alpha_{3,n} = 8 \ln(n)/n$ . In these three situations, condition (2) holds with respectively  $c = 1$ ,  $c = 1/2$  and  $c = 0$ . The first case corresponds to an extreme quantile located in the left-tail of the conditional distribution. The quantile level  $\alpha_2$  corresponds to the conditional median and  $\alpha_{3,n}$  to an extreme quantile in the right-tail. To evaluate the performance of the confidence interval, we compute its coverage probability  $\mathbb{P}[\text{CI}_{\gamma, \alpha_n}(k_n, d_e, x_0) \ni Q(\alpha_n|x_0)]$ . This probability is approximated numerically by a Monte-Carlo procedure. More specifically,  $N = 2\,000$  independent samples of size  $n = 1000$  were generated. For given values of  $k_n \in \{1, \dots, n\}$  and  $\gamma \in (0, 1)$ , the confidence interval obtained with the  $r$ -th replication is denoted  $\text{CI}_{\gamma, \alpha_n}^{(r)}(k_n, d_e, x_0)$ . The coverage probability is then approximated by

$$p_{\gamma, \alpha_n}(k_n, d_e, x_0) := \frac{1}{N} \sum_{r=1}^N \mathbb{I}_{\text{CI}_{\gamma, \alpha_n}^{(r)}(k_n, d_e, x_0)}(Q(\alpha_n|x_0)).$$

This value is expected to be close to the prespecified probability  $\gamma$ .

**Selection of the number of nearest neighbors** — We first take a look at the influence of  $k_n$  (the number of nearest neighbors) on the coverage probability. In Figure 1, the values of  $p_{\gamma, \alpha_n}(k_n, d_e, x_0)$  (with  $\alpha_n \in \{\alpha_{n,1}, \alpha_2, \alpha_{n,3}\}$ ) are represented as a function of  $k_n \in \{10, \dots, 200\}$  for **Model 1** with  $g_0 \in \{g_0^{(j)}, j = 1, 2, 3\}$  (see Table 1 for the definition of the functions  $g_0^{(j)}$ ). It appears that when the quantile level is extreme, only few values of  $k_n$  provide a reasonable coverage probability. It is thus relevant to propose a data driven procedure to select the value of  $k_n$ . The selected number of nearest neighbors depends on: the quantile level  $\alpha_n$ , the point of interest  $x_0 \in \mathbb{R}^p$ , the nominal coverage probability  $\gamma$  and the distance  $d$  used to collect the nearest neighbors. First, let

$$C(k) := \frac{1}{2} \left( Y_{\mathcal{L}_{\gamma}(k, \alpha_n), k}^{(d, x_0)} + Y_{\mathcal{R}_{\gamma}(k, \alpha_n), k}^{(d, x_0)} \right)$$

be the random variable corresponding to the center of the confidence interval  $\text{CI}_{\gamma, \alpha_n}(k_n, d, x_0)$ . The basic idea to select a convenient number of nearest neighbors is to take  $k$  is a stability region of the finite sequence

$\{C(n_0), \dots, C(n_1)\}$  where  $1 \leq n_0 < n_1 \leq n$ . More precisely, we are searching for the value

$$\tilde{k}_n^{(sel)} := \arg \min_{i \in \{n_0, \dots, n_1\}} \text{Var}(C(i)).$$

Of course, the variance of  $C(i)$  (and consequently the number  $\tilde{k}_n^{(sel)}$ ) is unknown in practice. We propose the following method to obtain an estimator of  $\tilde{k}_n^{(sel)}$ . Let  $a \in (0, 1)$  and denote by  $\lfloor \cdot \rfloor$  the floor function. For  $i \in \{n_0, \dots, n_1\}$ , the variance of  $C(i)$  is estimated by the local estimator

$$\hat{C}_n(i) := \frac{1}{\lfloor na \rfloor} \sum_{j \in \mathcal{V}(i)} \left( C(j) - \frac{1}{\lfloor na \rfloor} \sum_{\ell \in \mathcal{V}(i)} C(\ell) \right)^2,$$

where  $\mathcal{V}(i) \subset \{n_0, \dots, n_1\}$  is the set of the  $\lfloor na \rfloor$  nearest neighbors of  $i$ . Finally, for a given  $\eta \geq 0$ , we propose to take the following number of nearest neighbors:

$$\hat{k}_n^{(sel)} := \min\{i \in \{n_0, \dots, n_1\}; \text{Var}(\hat{C}_n(i)) \leq \eta\}, \quad (6)$$

with the convention  $\min\{\emptyset\} = n_0$ . Note that when  $\eta = 0$ ,  $\hat{k}_n^{(sel)}$  is the argument of the minimum of the sequence  $\{\hat{C}_n(n_0), \dots, \hat{C}_n(n_1)\}$ . The role of  $\eta$  is to obtain a value of  $\hat{k}_n^{(sel)}$  less sensitive to the fluctuations of the sequence  $\{\hat{C}_n(n_0), \dots, \hat{C}_n(n_1)\}$ . To sum up, the setting parameters required to compute (6) are: the integers  $n_0$  and  $n_1$  delimiting the searching region for the value of  $k_n$ , the value of  $a$  to compute the local estimator of the variance and the value of  $\eta$ . In all what follows, these parameters are fixed to  $n_0 = \lfloor 0.05n/p \rfloor$ ,  $n_1 = 200$ ,  $a = 0.006$  and  $\eta$  to the first quartile of the sequence  $\{\hat{C}_n(n_0), \dots, \hat{C}_n(n_1)\}$ .

In Figure 1, one can check that for the conditional median ( $\alpha_n = \alpha_2 = 1/2$ ) the coverage probability obtained with the selected value of  $k_n$  is close to the best attainable coverage probability. The choice of  $k_n$  is much more difficult for the extreme quantiles of level  $\alpha_{n,1}$  and  $\alpha_{n,3}$ .

**Obtained results** — The values of  $\tilde{p}_{\gamma, \alpha_n}(d_e, x_0) := p_{\gamma, \alpha_n}(\hat{k}_n^{(sel)}, d_e, x_0)$ , are gathered in Table 2 for **Model 1** and Table 3 for **Model 2** (see also Table 1 for the definitions of the functions  $g_0^{(j)}$ ).

For the conditional median ( $\alpha_n = 1/2$ ), the coverage probability is quite close to  $1 - \gamma$  and the accuracy of the confidence interval is not affected by the dimension  $p$  of the covariate. For a right-tail extreme quantile ( $\alpha_n = 8 \ln(n)/n$ ), the coverage probability is close to the nominal one when  $p = 1$ , but the precision of the confidence interval is strongly deteriorated when  $p$  increases (for instance, for  $p = 4$  and  $\gamma = 0.9$ , the coverage probability of  $\text{CI}_{\gamma, \alpha_n}(\hat{k}_n^{(sel)}, d_e, x_0)$  is equal to 0.6975). As discussed before, this is an expected consequence of the data sparsity around  $x_0$  when  $p$  increases. Finally, for a left-tail extreme quantile ( $\alpha_n = 1 - 8 \ln(n)/n$ ), the accuracy mostly depends on the function  $g_0$ . For instance, the expected number of observations in the left-tail of the

conditional distribution of  $Y$  given  $X = x_0$  is larger when  $g_0 = g_0^{(3)}$  than when  $g_0 = g_0^{(2)}$ . This can explain the good coverage probability obtained for  $g_0 = g_0^{(3)}$  (and the bad one for  $g_0 = g_0^{(2)}$ ).

### 3 Dimension reduction and confidence interval of large conditional quantile

As seen in the previous section, a large value of the covariate dimension ( $p \geq 4$ ) can seriously deteriorate the coverage probability of the confidence interval of an extreme quantile located in the right-tail of the conditional distribution. In this section, a dimension reduction procedure is proposed to obtain a better probability coverage of extreme quantile with  $\alpha_n \rightarrow 0$ . Note that the procedure described below can be easily adapted to the situation where  $\alpha_n \rightarrow 1$ .

#### 3.1 Tail conditional independence

Without any further assumptions, the classical euclidean distance is the natural distance to use in order to select the nearest neighbors. Unfortunately, due to the data sparsity when  $p$  is large, this distance selects observations that can be located far away from the point of interest  $x_0$ . In the literature devoted to dimension reduction, it is commonly assumed that there exists a function  $g_0 : \mathbb{R}^p \rightarrow \mathbb{R}$  such that the conditional distribution of  $Y$  given  $X$  is equal to the conditional distribution of  $Y$  given  $g_0(X)$ . The dimension of the covariate is thus reduced since  $X$  can be replaced by  $g_0(X)$ . In this case, to select the nearest neighbors, it seems preferable to use the pseudo-distance  $d_0$  defined for all  $(x, y) \in \mathbb{R}^p \times \mathbb{R}^p$  by  $d_0(x, y) := |g_0(x) - g_0(y)|$  instead of the euclidean distance in  $\mathbb{R}^p$ . Since we are only interested in the right-tail of the conditional distribution, we assume hereafter that  $Y$  is tail conditionally independent of  $X$  given  $g_0(X)$  as defined in Gardes [8]. More specifically, we assume that

- (TCI) the right endpoint of the conditional distribution of  $Y$  given  $X = x$  is infinite for all  $x \in \mathcal{X}$  and that there exists a function  $\varphi_y : \mathbb{R} \rightarrow \mathbb{R}$  depending on  $y$  and such that, as  $y \rightarrow \infty$ ,

$$\frac{\mathbb{P}[Y > y|X]}{\varphi_y(g_0(X))} \xrightarrow{a.s.u.} 1. \quad (7)$$

The notation  $\xrightarrow{a.s.u.}$  stands for the almost surely uniform convergence<sup>1</sup> (see for instance Lukács [20] or Rambaud [22, Proposition 1]). Roughly speaking under (7), inference on the extreme conditional quantile of  $Y$  given  $X$  can be achieved only by using the information brought by the reduced covariate  $g_0(X)$ . The appropriate distance to select the nearest neighbors is thus

---

<sup>1</sup>A stochastic process  $(Z_y, y \in \mathbb{R})$  converges almost surely uniformly to 1 as  $y \rightarrow \infty$  (in symbol  $Z_y \xrightarrow{a.s.u.} 1$ ) if for all  $\varepsilon > 0$ , there exists  $A$  such that for all  $y > A$ ,  $\mathbb{P}[|Z_y - 1| \leq \varepsilon] = 1$ .

the distance  $d_0$ .

Note that if there exist  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\phi \neq Id$  and  $\tilde{g}_0 : \mathbb{R}^p \rightarrow \mathbb{R}$  such that  $g_0 = \phi \circ \tilde{g}_0$  then if  $g_0$  satisfies (7) same holds for the function  $\tilde{g}_0$ . To ensure that  $g_0$  is the only function satisfying (7), we must assume that  $g_0 \in \mathcal{G}$  where  $\mathcal{G}$  is a set of functions satisfying the following property:

- ( $\mathcal{P}$ ) for all  $g : \mathbb{R}^p \rightarrow \mathbb{R} \in \mathcal{G}$ , there are no functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  (with  $\phi \neq Id$ ) and  $\tilde{g} : \mathbb{R}^p \rightarrow \mathbb{R} \in \mathcal{G}$  such that  $g = \phi \circ \tilde{g}$ .

Let  $u_p = (1, \dots, 1)^\top \in \mathbb{R}^p$ . A classical set satisfying ( $\mathcal{P}$ ) is the set of linear functions given by

$$\mathcal{G}_L := \left\{ g : \mathbb{R}^p \rightarrow \mathbb{R}; g(x) = b^\top x; b \in \Theta_p \right\}, \quad (8)$$

with  $\Theta_p := \{b \in \mathbb{R}^p \text{ with } b^\top b = 1 \text{ and } b^\top u_p > 0\}$ . Note that this set is the one considered in Gardes [9]. One can also consider set on non-linear functions (see Section 3.3 for an example). Finding an appropriate distance for the selection of nearest neighbors (in an extreme value framework) amounts to finding the function  $g_0$  satisfying (7). In Gardes [9], a method has been proposed in the particular case of a linear function  $g_0$ . Unfortunately, the procedure is computationally expensive and can be used only for a linear function  $g_0 \in \mathcal{G}_L$ .

### 3.2 Selection of the distance

In this section, we propose a new data-driven procedure to select an appropriate distance  $d_0$  for the selection of the nearest neighbors. As explained above, we thus have to estimate the function  $g_0 \in \mathcal{G}$  satisfying (7). For  $H \in \mathbb{N} \setminus \{0\}$  and  $g \in \mathcal{G}$ , let  $\mathcal{S}_{1,g}, \dots, \mathcal{S}_{H,g}$  be non-overlapping intervals covering the support of  $g(X)$ . We denote by  $\{(X_{(i)}, Y_{n-i,n}), i = 0, \dots, n-1\}$  the random vectors  $\{(X_i, Y_i), i = 1, \dots, n\}$  rearranged in order to have  $Y_{1,n} \leq \dots \leq Y_{n,n}$ . For a given  $\beta_n \in (0, 1)$  with  $\beta_n \rightarrow 0$  as  $n \rightarrow \infty$ , let

$$\hat{E}_{h,g}(\beta_n) := \frac{1}{M_{h,g}^2} \sum_{i,j=1}^{\lfloor n\beta_n \rfloor - 1} |g(X_{(i)}) - g(X_{(j)})| \mathbb{I}_{\mathcal{S}_{h,g}^2}((g(X_{(i)}), g(X_{(j)}))),$$

with

$$M_{h,g} := \sum_{i=1}^{\lfloor n\beta_n \rfloor} \mathbb{I}_{\mathcal{S}_{h,g}}(g(X_{(i)})).$$

The statistic  $\hat{E}_{h,g}(\beta_n)$  is the average distance between the observations  $g(X_i)$  with  $(X_i, Y_i) \in \mathcal{S}_{h,g} \times [Y_{n-\lfloor n\beta_n \rfloor, n}, \infty)$ . If condition (7) holds for  $g_0 \in \mathcal{G}$  then, for  $y$  large enough and given that  $Y > y$ , the random variable  $g_0(X)$  is more likely to be observed around the local maximum points of the function  $\varphi_y$ . This is illustrated in Figure 2 where  $n$  independent random vectors  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  are generated from **Model 1** with  $g_0 = g_0^{(2)}$  (see Table 1). Recall that **Model 1** satisfies condition (7) with  $\varphi_y(z) = y^{-1/\xi(z)}$ .

Note that this function admits two local maxima at the points  $-3\sqrt{5}$  and  $3\sqrt{5}$ . In the left panel of Figure 2, the histogram shows clearly that the observations  $\{g_0(X_i), Y_i > y\}$  are located around the two maximum points of the function  $\varphi_y$ . In the right panel, we can check that this is no longer the case replacing the function  $g_0$  by another function (here  $g_1(x) = (1, \dots, 1)^\top x / \sqrt{4}$ ). Therefore, for  $n$  large enough, the number of sets  $\mathcal{S}_{h,g_0} \times [Y_{n-\lfloor n\beta_n \rfloor, n}, \infty)$  containing at least one observation is expected to be smaller than the number of local maximum points and for these sets,  $\hat{E}_{h,g_0}(\beta_n)$  must be close to 0. We thus propose to estimate the function  $g_0 \in \mathcal{G}$  satisfying (7) by

$$\hat{g}_0 := \arg \min_{g \in \mathcal{G}} \left\{ \sum_{h \in J_{H,g}} \hat{E}_{h,g}(\beta_n) + \lambda \text{card}(J_{H,g}) \right\}, \quad (9)$$

where  $\lambda > 0$ ,  $J_{H,g} := \{h \in \{1, \dots, H\}; M_{h,g} > 0\}$  and  $\text{card}(J_{H,g})$  is the number of elements in  $J_{H,g}$ . For  $g \in \mathcal{G}$  and  $h \in \{1, \dots, H\}$ , we take  $\mathcal{S}_{h,g} := [\hat{\xi}_{h-1,g}, \hat{\xi}_{h,g}]$  where  $\hat{\xi}_{h,g}$  is the sample quantile with corresponding probability  $h/H$  of the sample  $\{g(X_i), i = 1, \dots, n\}$ . The setting parameters of our procedure of estimation are: the sequence  $\beta_n$ , the number  $H$  of intervals and finally the penalty term  $\lambda$ . In the numerical illustration below, these parameters are set to  $\beta_n = 5/(3\sqrt{n})$ ,  $\lambda = 1$  and  $H = 20$ .

To conclude this section, assuming that (7) holds, we propose a more theoretical justification of the estimator given by (9). Before that, let us introduce the following additional condition on the function  $\varphi_y$  involved in (7). In what follows, we denote by  $\mathcal{X}_0$  the support of  $g_0(X)$  where  $g_0$  satisfies condition (7).

- (H1) The function  $\varphi_y$  is continuous. Furthermore, there exist  $J \in \mathbb{N} \setminus \{0\}$ , a collection  $I_1, \dots, I_J$  of non-overlapping intervals covering  $\mathcal{X}_0$  and  $y_0 \in \mathbb{R}$ , such that for all  $y \geq y_0$  and  $j \in \{1, \dots, J\}$ , the function  $\varphi_y$  admits on  $I_j$  a unique local maximum point  $z_j^*$  such that  $z_j^*$  is an interior point of  $I_j$ .

This condition entails that for  $y$  large enough, the function  $\varphi_y$  admits a finite number of local maximum points. Assuming that (H1) holds, the following condition on the non-overlapping intervals  $\mathcal{S}_{1,g}, \dots, \mathcal{S}_{H,g}$  is required.

- (H2) For all  $g \in \mathcal{G}$ , there exists  $H_{0,g} \in \mathbb{N} \setminus \{0\}$  such that for all  $H \geq H_{0,g}$  and  $j \in \{1, \dots, J\}$ ,

$$z_j^* \in \bigcup_{h=1}^H \mathring{\mathcal{S}}_{h,g}$$

where  $\mathring{I}$  is the interior of the interval  $I$ .

Hence, under (H1) and (H2), the  $J$  local maximum points of  $\varphi_y$  belong to the interior of an interval  $\mathcal{S}_{h,g}$ . A supplementary condition on the distribution of  $(X, Y)$  is also required. Let  $\mathcal{C} \subset \mathcal{X}$  such that  $\mathbb{P}(X \in \mathcal{C}) > 0$ . For all  $y$  and  $t \in (0, 1)$ , let  $p_y(\cdot | \mathcal{C})$  be the survival function of  $S(y|X)$  given  $X \in \mathcal{C}$

$(p_y(t|\mathcal{C}) := \mathbb{P}[S(y|X) > t | X \in \mathcal{C}])$ . The associated quantile function is denoted by  $q_y(\cdot|\mathcal{C})$  ( $q_y(\cdot|\mathcal{C}) := \inf\{t; p_y(t|\mathcal{C}) \leq \cdot\}$ ).

(H3) For all  $(\eta, d) \in (0, 1)^2$ ,

$$\lim_{y \rightarrow \infty} \frac{q_y(\eta|\mathcal{C})}{q_y(d\eta|\mathcal{C})} = 0.$$

This technical condition is satisfied (under some regularity condition, see Lemma 5) by conditional heavy-tailed distributions defined for all  $x \in \mathcal{X}$  by  $S(y|x) := y^{-1/\gamma(x)}\mathcal{L}(y|x)$ , where  $\gamma$  is a positive function and for all  $x \in \mathcal{X}$ ,  $\mathcal{L}(\cdot|x)$  is a slowly varying function. Condition (H3) entails that the observations of  $g_0(X)$  given that  $Y > y$  and  $g_0(X) \in \mathcal{C}$  are located on a small probability set (see Lemma 4). We are now in position to provide a theoretical justification of (9).

**Proposition 2** *Assume that there exists a function  $g_0$  satisfying condition (7) and such that  $g_0(X)$  admits a density function  $f_0$ . If there exists  $m > 0$  such that  $f_0(x) \geq m$  for all  $x \in \mathcal{X}_0$ , if condition (H3) holds for all  $\mathcal{C} = g_0^{-1}(I)$  where  $I \subset \mathcal{X}_0$  is an interval then, for  $y$  large enough and all  $v > 0$ , one has under (H1) and (H2) that, introducing the set  $\mathcal{J}_{H,y}(v) := \{h \in \{1, \dots, H\}; \mathbb{P}(g_0(X) \in \mathcal{S}_{h,g_0} | Y > y) > v\}$ ,*

$$\lim_{y \rightarrow \infty} \sum_{h \in \mathcal{J}_{H,y}(v)} E_{g_0}(y; \mathcal{S}_{h,g_0}) = 0 \text{ and } \text{card}(\mathcal{J}_{H,y}(v)) \leq J, \quad (10)$$

where, for an independent copie  $(X^*, Y^*)$  of  $(X, Y)$ ,  $E_{g_0}(y; \mathcal{S}_{h,g_0})$  is equal to

$$\mathbb{E} [|g_0(X) - g_0(X^*)| | \{(g_0(X), g_0(X^*)) \in \mathcal{S}_{h,g_0}^2\} \cap \{\min(Y, Y^*) > y\}].$$

Denoting by  $Q_Y(\cdot) := \inf\{y; \mathbb{P}(Y > y) \leq \cdot\}$  the quantile function of the distribution of  $Y$ , one can see that the statistic  $\hat{E}_{h,g}(\beta_n)$  is the empirical counterpart of the expectation  $E_g(Q_Y(\beta_n); \mathcal{S}_{h,g})$  where  $Q_Y(\beta_n)$  has been replaced by the order statistic  $Y_{n-\lfloor n\beta_n \rfloor, n}$ . Moreover,  $M_{h,g}/\lfloor n\beta_n \rfloor$  is an estimator of the probability  $\mathbb{P}(g(X) \in \mathcal{S}_{h,g} | Y > Q_Y(\beta_n))$ . Since  $Q_Y(\beta_n) \rightarrow \infty$  as  $n \rightarrow \infty$ , (10) can be seen as a theoretical justification of (9).

### 3.3 Numerical illustrations

Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be  $n$  independent copies of a random vector  $(X, Y)$ , where  $X$  is a  $\mathbb{R}^p$ -valued random variable with  $p > 1$  and  $Y$  is a  $\mathbb{R}$ -valued random variable. The random vector  $(X, Y)$  is distributed according to **Model 1** or **Model 2** (see Section 2.2). Here we focus on extreme quantiles in the right-tail and we take  $\alpha_n = \alpha_{n,3} = 8 \ln(n)/n$  for the quantile level.

It can be shown that **Model 1** satisfies condition (TCI) with  $\varphi_y(z) = y^{-1/\xi(z)}$ . Of course, condition (TCI) also holds for **Model 2** with  $\varphi_y(z) = \exp(-y^{-1/\xi(z)})$ . In what follows, the function  $g_0$  involved in **Model 1** and **Model 2** is taken in the set  $\{g_0^{(j)}, j = 2, \dots, 5\}$  (see Table 1). Note

that the function  $g_0^{(5)}$  defined for all  $x \in \mathbb{R}^p$  by  $g_0^{(5)}(x) = |(b_0^{(5)})^\top x^2|^{1/2}$  (where the components of  $x^2$  are the square of the components of  $x$ ) is a non-linear function of  $x$ . It belongs to the set

$$\mathcal{G}_{NL,r} := \left\{ g : \mathbb{R}^p \rightarrow \mathbb{R}^p; g(x) = |b^\top x^r|^{1/r}; b \in \Theta_r \right\},$$

with  $r \in \mathbb{N} \setminus \{0\}$  and  $\Theta_p := \{b \in \mathbb{R}^p \text{ with } b^\top b = 1 \text{ and } b^\top u_p > 0\}$ . Instead of the classical euclidean distance, one could (should) use for each  $j \in \{2, \dots, 5\}$  the (unknown) distance  $d_0^{(j)}$  (with  $d_0^{(j)}(x, y) = |g_0^{(j)}(x) - g_0^{(j)}(y)|$ ) to select the nearest neighbors. Using this distance amounts to replace the  $p$ -dimensional covariate  $X$  by the real-valued variable  $g_0^{(j)}(X)$ . It is thus expected to obtain better coverage probabilities than in Section 2.2.

The first step consists in the estimation of the function  $g_0^{(j)}$  (or equivalently of the distance  $d_0^{(j)}$ ) by (9). For  $j \in \{2, 3, 4\}$ , the minimization is achieved over the set  $\mathcal{G} = \mathcal{G}_L$  and for  $j = 5$ , over the set  $\mathcal{G} = \mathcal{G}_{NL,2}$ . The optimization problem (9) is solved by using a coordinate search method (see Hooke and Jeeves [16] and Appendix B for more details). The estimation of  $g_0^{(j)}$  reduces to the estimation of the vector  $b_0^{(j)}$ . Denoting by  $\hat{b}_{n,0}^{(j)}$  this estimator, the accuracy of the estimation procedure is measured by the average and the standard deviation of the error

$$\delta_n^{(j)} := (\hat{b}_{n,0}^{(j)} - b_0^{(j)})^\top (\hat{b}_{n,0}^{(j)} - b_0^{(j)}).$$

The corresponding estimator of  $g_0^{(j)}$  is denoted  $\hat{g}_{n,0}^{(j)}$ . For each  $j \in \{1, \dots, 5\}$ , the average and the standard deviation of  $\delta_n^{(j)}$  are estimated by a Monte-Carlo procedure by replicating  $N = 2\,000$  times the sample  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ . Next, the confidence interval defined in (3) is computed by selecting the nearest neighbors with the estimated distance  $\tilde{d}_0^{(j)}(x, y) := |\hat{g}_{n,0}^{(j)}(x) - \hat{g}_{n,0}^{(j)}(y)|$  instead of the euclidean distance. For a given number  $k$  of nearest neighbors, the obtained coverage probability is denoted by  $p_{\gamma, \alpha_n}(k, \tilde{d}_0^{(j)}, x_0)$ . When  $k$  is chosen by the procedure described in Section 2.2, the coverage probability is denoted  $\tilde{p}_{\gamma, \alpha_n}(\tilde{d}_0^{(j)}, x_0) := p_{\gamma, \alpha_n}(\hat{k}_n^{(sel)}, \tilde{d}_0^{(j)}, x_0)$ . This last probability is compared to  $\tilde{p}_{\gamma, \alpha_n}(d_0^{(j)}, x_0)$  which is the coverage probability obtained by using the unknown distance  $d_0^{(j)}(x, y) := |g_0^{(j)}(x) - g_0^{(j)}(y)|$  in the selection of the nearest neighbors.

The values of  $\tilde{p}_{\gamma, \alpha_n}(\tilde{d}_0^{(j)}, x_0)$ ,  $\tilde{p}_{\gamma, \alpha_n}(d_0^{(j)}, x_0)$  and  $\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$  are given in Tables 4 and 5. For linear functions (*i.e.*  $g_0 \in \mathcal{G}_L$ ), replacing the euclidean distance by the estimated distance  $\tilde{d}_0^{(j)}$  leads to a significant improvement in the coverage probability. Note that the estimation of the function  $g_0^{(5)} \in \mathcal{G}_{NL}$  is more challenging (especially in **Model 1**) but the obtained coverage probability remains better than the one obtained with the euclidean distance. In Figure 3, the coverage probabilities  $p_{\gamma, \alpha_n}(k, d_e, x_0)$ ,  $p_{\gamma, \alpha_n}(k, \tilde{d}_0^{(j)}, x_0)$  and  $p_{\gamma, \alpha_n}(k, d_0^{(j)}, x_0)$  are represented as a function of the number  $k$  of nearest neighbors. It appears that the choice of  $k$  is really less crucial when one use the

estimated distance  $\hat{d}_0^{(j)}$ . We can check again that the selection of  $k_n$  by our procedure provides a confidence interval with a coverage probability close to  $\gamma$ .

## 4 Chicago air pollution dataset

The results obtained in this paper are illustrated on the Chicago air pollution dataset. This dataset, available on the R package `NMMAPS Data Lite`, gathers the daily concentrations of different pollutants (ozone ( $O_3$ ), particular matter with diameter smaller than 10 microns or 25 microns ( $PM_{10}$  or  $PM_{25}$ ), sulphur dioxide ( $SO_2$ ), nitrogen dioxide ( $NO_2$ ), carbon monoxide ( $CO$ ), etc.) and some meteorology and mortality variables. The data were collected in Chicago from 1987 to 2000 during  $n = 4841$  days. This dataset has been studied by several authors in a dimension reduction context (see for instance Scrucca [24] and Xia [26]) and, in an extreme value context, by Gardes [9].

We are interested in the conditional distribution of  $Y$  given  $X = x_0$  where  $Y$  corresponds to the (centered and normalized) concentration of  $O_3$  (in parts per billion) and  $X$  is the covariate vector of dimension  $p = 4$  corresponding to the daily maximum concentrations of  $PM_{10}$ ,  $SO_2$ ,  $NO_2$  and  $CO$ . As in Gardes [9], we assume that condition (7) holds with  $g_0 \in \mathcal{G}_L$ . As a first step, we use (9) to estimate the direction  $b_0$  such that  $g_0(x) = b_0^\top x$  for  $x \in \mathbb{R}^p$ . The obtained estimated vector is  $\hat{b}_{n,0} := (0.198, -0.155, 0.963, 0.093)^\top$ . Note that this estimated vector is quite close to the direction estimated in [9] and given by  $\tilde{b}_{n,0} := (0.175, -0.036, 0.962, -0.207)^\top$  (we have  $(\hat{b}_{n,0} - \tilde{b}_{n,0})^\top (\hat{b}_{n,0} - \tilde{b}_{n,0}) = 0.105$ ). As noted by Scrucca [24] or Gardes [9], the covariate  $NO_2$  seems to bring the most important information on large values of ozone concentration. Our goal is now to construct a confidence interval with prespecified coverage probability  $\gamma = 0.9$  of the conditional quantile  $Q(\alpha_n|x_0)$ . Two possible situations for  $x_0$  are considered:

**Situation 1** –  $x_0 = (x_0^{PM_{10}}(0.5), x_0^{SO_2}(0.5), x_0^{NO_2}(0.5), x_0^{CO}(0.5))^\top$  where for  $\tau \in (0, 1)$ ,  $x_0^{PM_{10}}(\tau)$ ,  $x_0^{SO_2}(\tau)$ ,  $x_0^{NO_2}(\tau)$  and  $x_0^{CO}(\tau)$  are the sample quantile of order  $1 - \tau$  of the (centered and normalized) daily maximum values of  $PM_{10}$ ,  $SO_2$ ,  $NO_2$  and  $CO$ . This value of  $x_0$  is quite close to a situation observed in Chicago during the period 1987-2000 with moderate values of the four primary pollutants.

**Situation 2** –  $x_0 = (x_0^{PM_{10}}(0.5), x_0^{SO_2}(0.25), x_0^{NO_2}(0.05), x_0^{CO}(0.05))^\top$  corresponding to large values for  $NO_2$  and  $CO$ .

For the quantile level  $\alpha_n$ , we assume that  $\alpha_n \in [8 \ln(n)/n, 64 \ln(n)/n]$ . The obtained confidence interval (with  $\gamma = 0.9$ ) are represented on Figure 4 as a function of  $\alpha_n$ . As already noted in Gardes [9] or Han *et al.* [15], very important ozone concentration is more likely to be observed when concentrations of  $NO_2$  and  $CO$  are important. As expected, one can also check that the length of the confidence interval increases when the quantile level  $\alpha_n$  decreases.



## 5 Proofs

### 5.1 Preliminaries results

In this section we give two useful results on Beta distribution. The probability density function of a Beta distribution with parameters  $a$  and  $b$  is given by

$$f_{\text{beta}}(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbb{I}_{[0,1]}(x),$$

where  $\Gamma$  is the gamma function.

**Lemma 1** *For all  $m \in \mathbb{N} \setminus \{0\}$  and  $\alpha \in (0, 1)$ ,*

$$\arg \max_{j \in \{1, \dots, m\}} f_{\text{beta}}(\alpha; m-j+1, j) = m - \lfloor m\alpha \rfloor.$$

*Furthermore, if  $m_n \in \mathbb{N} \setminus \{0\}$  and  $\alpha_n \in (0, 1)$  are sequences such that  $m_n \rightarrow \infty$  and  $m_n(\alpha_n \wedge (1 - \alpha_n)) \rightarrow \infty$  as  $n \rightarrow \infty$ , then for all sequence  $\varepsilon_n$  such that*

$$\varepsilon_n^2 = o\left(\frac{1 - \alpha_n}{m_n \alpha_n}\right),$$

*and for  $\alpha_{n,\tau} := \alpha_n(1 + \tau\varepsilon_n)$ , there exist  $0 < c_1 < c_2$  such that for  $n$  large enough,*

$$\left(\frac{\alpha_n(1 - \alpha_n)}{m_n}\right)^{1/2} \max_{j \in \{1, \dots, m_n\}} \sup_{\tau \in [-1, 1]} f_{\text{beta}}(\alpha_{n,\tau}; m_n - j + 1, j) \in [c_1, c_2].$$

**Proof** – For  $m \in \mathbb{N} \setminus \{0\}$  and  $\alpha \in (0, 1)$ , let

$$a_j := f_{\text{beta}}(\alpha; m-j+1, j) = \frac{m!}{(j-1)!(m-j)!} \alpha^{m-j} (1-\alpha)^{j-1}.$$

It is easy to check that for all  $j \in \{1, \dots, m-1\}$ ,

$$\frac{a_{j+1}}{a_j} = \frac{m-j}{j} \frac{1-\alpha}{\alpha}.$$

Hence,  $a_{j+1}/a_j \geq 1$  if and only if  $j \leq m(1-\alpha)$ , proving the first part of the Lemma. To prove the second part, we start with

$$\begin{aligned} & \max_{j \in \{1, \dots, m_n\}} \sup_{\tau \in [-1, 1]} f_{\text{beta}}(\alpha_{n,\tau}; m_n - j + 1, j) \\ &= \sup_{\tau \in [-1, 1]} f_{\text{beta}}(\alpha_{n,\tau}; \lfloor m_n \alpha_{n,\tau} \rfloor + 1, m_n - \lfloor m_n \alpha_{n,\tau} \rfloor). \end{aligned}$$

In order to study the factor

$$\frac{m_n!}{(\lfloor m_n \alpha_{n,\tau} \rfloor)!(m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1)!},$$

appearing in the expression of  $f_{\text{beta}}(\alpha_{n,\tau}; \lfloor m_n \alpha_{n,\tau} \rfloor + 1, m_n - \lfloor m_n \alpha_{n,\tau} \rfloor)$ , we use the Stirling's bounds given for all  $r \in \mathbb{N} \setminus \{0\}$  by  $\sqrt{2\pi} r^{r+1/2} e^{-r} \leq r! \leq$

$r^{r+1/2}e^{1-r}$  (see for instance [7, Paragraph 2.9]).

First, taking  $r = m_n$  leads to

$$\sqrt{2\pi}m_n^{1/2+m_n}e^{-m_n} \leq m_n! \leq m_n^{1/2+m_n}e^{1-m_n}. \quad (11)$$

Next, using the Stirling's bounds with  $r = \lfloor m_n\alpha_{n,\tau} \rfloor$  yields to  $\sqrt{2\pi}s_n \leq (\lfloor m_n\alpha_{n,\tau} \rfloor)! \leq e \times s_n$  with

$$s_n := \frac{(m_n\alpha_{n,\tau})^{\lfloor m_n\alpha_{n,\tau} \rfloor + 1/2}}{e^{\lfloor m_n\alpha_{n,\tau} \rfloor}} \left( \frac{\lfloor m_n\alpha_{n,\tau} \rfloor}{m_n\alpha_{n,\tau}} \right)^{1/2 + \lfloor m_n\alpha_{n,\tau} \rfloor}.$$

It is easy to check that for all  $\tau \in [-1, 1]$ ,

$$1 - \frac{1}{m_n\alpha_n(1 - \varepsilon_n)} \leq \frac{\lfloor m_n\alpha_{n,\tau} \rfloor}{m_n\alpha_{n,\tau}} \leq 1,$$

and  $\lfloor m_n\alpha_{n,\tau} \rfloor \leq m_n\alpha_n(1 + \varepsilon_n)$ . As a consequence, one has for all  $\tau \in [-1, 1]$  and for  $n$  large enough that,

$$\frac{1}{2e} \leq \left( 1 - \frac{1}{m_n\alpha_n(1 - \varepsilon_n)} \right)^{1/2 + m_n\alpha_n(1 + \varepsilon_n)} \leq \left( \frac{\lfloor m_n\alpha_{n,\tau} \rfloor}{m_n\alpha_{n,\tau}} \right)^{1/2 + \lfloor m_n\alpha_{n,\tau} \rfloor} \leq 1.$$

Note that the first inequality is due to the fact that

$$\left( 1 - \frac{1}{m_n\alpha_n(1 - \varepsilon_n)} \right)^{1/2 + m_n\alpha_n(1 + \varepsilon_n)} \rightarrow e^{-1},$$

since by assumption  $m_n\alpha_n \rightarrow \infty$  and  $\varepsilon_n \rightarrow 0$ . We finally get that for  $n$  large enough and all  $\tau \in [-1, 1]$ ,

$$\sqrt{\frac{\pi}{2}} \frac{(m_n\alpha_{n,\tau})^{1/2 + \lfloor m_n\alpha_{n,\tau} \rfloor}}{e^{\lfloor m_n\alpha_{n,\tau} \rfloor + 1}} \leq (\lfloor m_n\alpha_{n,\tau} \rfloor)! \leq \frac{(m_n\alpha_{n,\tau})^{1/2 + \lfloor m_n\alpha_{n,\tau} \rfloor}}{e^{\lfloor m_n\alpha_{n,\tau} \rfloor - 1}}. \quad (12)$$

Finally, the Stirling's bounds applied to  $r = m_n - \lfloor m_n\alpha_{n,\tau} \rfloor - 1$  leads to  $\sqrt{2\pi}t_n \leq (m_n - \lfloor m_n\alpha_{n,\tau} \rfloor - 1)! \leq e \times t_n$  with

$$t_n := \frac{(m_n(1 - \alpha_{n,\tau}))^{m_n - \lfloor m_n\alpha_{n,\tau} \rfloor - 1/2}}{e^{m_n - \lfloor m_n\alpha_{n,\tau} \rfloor - 1}} \left( \frac{m_n - \lfloor m_n\alpha_{n,\tau} \rfloor - 1}{m_n(1 - \alpha_{n,\tau})} \right)^{m_n - \lfloor m_n\alpha_{n,\tau} \rfloor - 1/2}.$$

Remark that for all  $\tau \in [-1, 1]$ ,

$$1 - \frac{1}{m_n(1 - \alpha_n - \alpha_n\varepsilon_n)} \leq \frac{m_n - \lfloor m_n\alpha_{n,\tau} \rfloor - 1}{m_n(1 - \alpha_{n,\tau})} \leq 1.$$

Furthermore, since by assumption

$$\left( \frac{\alpha_n\varepsilon_n}{1 - \alpha_n} \right)^2 = o\left( \frac{\alpha_n}{m_n(1 - \alpha_n)} \right) = o(1),$$

one has for  $n$  large enough that

$$1 - \alpha_n - \alpha_n\varepsilon_n = (1 - \alpha_n) \left( 1 - \frac{\alpha_n\varepsilon_n}{1 - \alpha_n} \right) \geq \frac{1 - \alpha_n}{2}.$$

As a consequence, we get

$$1 - \frac{1}{2m_n(1 - \alpha_n)} \leq \frac{m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1}{m_n(1 - \alpha_{n,\tau})} \leq 1.$$

Since

$$\left(1 - \frac{1}{2m_n(1 - \alpha_n)}\right)^{m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1/2} \rightarrow e^{-1/2},$$

we obtain the inequality

$$\frac{1}{2e^{1/2}} \leq \left(\frac{m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1}{m_n(1 - \alpha_{n,\tau})}\right)^{m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1/2} \leq 1$$

leading to

$$\begin{aligned} & \sqrt{\frac{\pi}{2}} \frac{(m_n(1 - \alpha_{n,\tau}))^{m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1/2}}{e^{m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1/2}} \leq (m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1)! \\ & \leq \frac{(m_n(1 - \alpha_{n,\tau}))^{m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 1/2}}{e^{m_n - \lfloor m_n \alpha_{n,\tau} \rfloor - 2}}. \end{aligned} \quad (13)$$

Gathering (11), (12) and (13) yields to

$$\begin{aligned} \frac{\sqrt{2\pi}}{e^3} \left(\frac{m_n}{\alpha_{n,\tau}(1 - \alpha_{n,\tau})}\right)^{1/2} & \leq f_{\text{beta}}(\alpha_{n,\tau}; \lfloor m_n \alpha_{n,\tau} \rfloor + 1, m_n - \lfloor m_n \alpha_{n,\tau} \rfloor) \\ & \leq \frac{2}{\pi} e^{3/2} \times \left(\frac{m_n}{\alpha_{n,\tau}(1 - \alpha_{n,\tau})}\right)^{1/2}. \end{aligned}$$

Finally, since for all  $\tau \in [-1, 1]$ ,  $|\alpha_{n,\tau}/\alpha_n - 1| \leq \varepsilon_n \rightarrow 0$  and  $|(1 - \alpha_{n,\tau})/(1 - \alpha_n) - 1| \leq \alpha_n \varepsilon_n / (1 - \alpha_n) \rightarrow 0$ ,  $\alpha_n(1 - \alpha_n)/2 \leq \alpha_{n,\tau}(1 - \alpha_{n,\tau}) \leq 2\alpha_n(1 - \alpha_n)$  and the proof is complete by letting  $c_1 := \sqrt{\pi/2}e^{-1}$  and  $c_2 := 4e$ .  $\blacksquare$

**Lemma 2** *Let  $m_n$  and  $\alpha_n \in (0, 1)$  be two sequences such that  $m_n \rightarrow \infty$  and*

$$\delta_n^2 := \frac{\ln^2(m_n)}{m_n \alpha_n (1 - \alpha_n)} \rightarrow 0, \quad (14)$$

*as  $n \rightarrow \infty$ . For all  $\gamma \in (0, 1)$ , one has  $1 \leq \mathcal{L}_\gamma(m_n, \alpha_n) \leq \mathcal{R}_\gamma(m_n, \alpha_n) \leq m_n$ . Furthermore,*

$$F_{\text{beta}}(\alpha_n; m_n - \mathcal{L}_\gamma(m_n, \alpha_n) + 1, \mathcal{L}_\gamma(m_n, \alpha_n)) = \frac{1 - \gamma}{2} + \mathcal{O}(\delta_n) \quad (15)$$

$$\text{and } F_{\text{beta}}(\alpha_n; m_n - \mathcal{R}_\gamma(m_n, \alpha_n) + 1, \mathcal{R}_\gamma(m_n, \alpha_n)) = 1 - \frac{1 - \gamma}{2} + \mathcal{O}(\delta_n). \quad (16)$$

**Proof** – Remark that since  $m_n \rightarrow \infty$ , condition (14) entails that  $m_n(\alpha_n \wedge (1 - \alpha_n)) \rightarrow \infty$ . Hence, since the function  $j \rightarrow F_{\text{beta}}(\alpha; m_n - j + 1, j)$  is increasing for all  $\alpha \in (0, 1)$

$$\max_{j=1, \dots, m_n} F_{\text{beta}}(\alpha_n; n - j + 1, j) = F_{\text{beta}}(\alpha_n; 1, m_n) = 1 - (1 - \alpha_n)^{m_n} \rightarrow 1,$$

as  $n \rightarrow \infty$  and, using the inequality  $\ln(x) \leq x - 1$  that holds for all  $x \in [0, 1]$

$$\begin{aligned} \min_{j=1, \dots, m_n} F_{\text{beta}}(\alpha_n; n - j + 1, j) &= F_{\text{beta}}(\alpha_n; m_n, 1) = \alpha_n^{m_n} \\ &= \exp[m_n \ln(\alpha_n)] \leq \exp[-m_n(1 - \alpha_n)] \rightarrow 0. \end{aligned}$$

Hence, for  $n$  large enough,

$$\left\{ j \in \{1, \dots, m_n\}; F_{\text{beta}}(\alpha_n; m_n - j + 1, j) \leq \frac{1 - \gamma}{2} \right\} \neq \emptyset,$$

and

$$\left\{ j \in \{1, \dots, m_n\}; F_{\text{beta}}(\alpha_n; m_n - j + 1, j) \geq 1 - \frac{1 - \gamma}{2} \right\} \neq \emptyset.$$

This conclude the first part of the proof.

We now prove (15). The proof of (16) is similar and is thus omitted. The definition of  $\mathcal{L}_\gamma(m_n, \alpha_n)$  ensures that

$$0 \leq \frac{1 - \gamma}{2} - F_{\text{beta}}(\alpha_n; m_n - \mathcal{L}_\gamma(m_n, \alpha_n) + 1, \mathcal{L}_\gamma(m_n, \alpha_n)) \leq D_n(\mathcal{L}_\gamma(m_n, \alpha_n))$$

where  $D_n(m_n) := 1 - F_{\text{beta}}(\alpha_n; 1, m_n)$  and for  $j = 1, \dots, m_n - 1$ ,

$$D_n(j) := F_{\text{beta}}(\alpha_n; m_n - j, j + 1) - F_{\text{beta}}(\alpha_n; m_n - j + 1, j).$$

Hence to prove (15) it suffices to show that

$$\max_{j=1, \dots, m_n} D_n(j) = \mathcal{O} \left( \frac{\ln(m_n)}{[m_n \alpha_n (1 - \alpha_n)]^{1/2}} \right). \quad (17)$$

First,  $D_n(m_n) = (1 - \alpha_n)^{m_n}$ . Using the inequality  $(1 - u)^\xi \leq \exp(-\xi u)$  that holds for all  $u \in (0, 1)$  and  $\xi > 0$  and the fact that  $1 - \alpha_n \in (0, 1)$ , we get that

$$\frac{D_n(m_n)[m_n \alpha_n (1 - \alpha_n)]^{1/2}}{\ln(m_n)} \leq \frac{\exp(-m_n \alpha_n)(m_n \alpha_n)^{1/2}}{\ln(m_n)} \rightarrow 0,$$

as  $n \rightarrow \infty$ . We thus have shown that

$$D_n(m_n) = o \left( \frac{\ln(m_n)}{[m_n \alpha_n (1 - \alpha_n)]^{1/2}} \right). \quad (18)$$

Now, let  $U_1, \dots, U_{m_n}$  be  $m_n$  independent standard uniform random variables and let  $U_{1, m_n} \leq \dots \leq U_{m_n, m_n}$  be the corresponding order statistics. It is well known that for all  $j \in \{1, \dots, m_n\}$ , the order statistic  $U_{j, m_n}$  follows a beta distribution with parameters  $j$  and  $m_n - j + 1$ . Hence, for all  $j = 2, \dots, m_n$

$$\begin{aligned} D_n(m_n - j + 1) &= \mathbb{P}[U_{j-1, m_n} \leq \alpha_n] - \mathbb{P}[U_{j, m_n} \leq \alpha_n] \\ &\leq \mathbb{P} \left[ U_{j, m_n} \leq \alpha_n + \max_{j=2, \dots, m_n} (U_{j, m_n} - U_{j-1, m_n}) \right] \\ &\quad - \mathbb{P}[U_{j, m_n} \leq \alpha_n]. \end{aligned}$$

Let

$$A_n := \left\{ \max_{j=2,\dots,m_n} (U_{j,m_n} - U_{j-1,m_n}) \leq 2 \frac{\ln(m_n)}{m_n} \right\}$$

and  $\bar{A}_n := \left\{ \max_{j=2,\dots,m_n} (U_{j,m_n} - U_{j-1,m_n}) > 2 \frac{\ln(m_n)}{m_n} \right\}.$

It is easy to check that

$$\begin{aligned} D_n(m_n - j + 1) &\leq \mathbb{P} \left[ \left\{ U_{j,m_n} \leq \alpha_n + \max_{j=2,\dots,m_n} (U_{j,m_n} - U_{j-1,m_n}) \right\} \cap A_n \right] \\ &= \mathbb{P}[U_{j,m_n} \leq \alpha_n] + \mathbb{P}(\bar{A}_n) \\ &\leq D_n^{(1)}(m_n - j + 1) + \mathbb{P}(\bar{A}_n), \end{aligned} \quad (19)$$

with

$$D_n^{(1)}(m_n - j + 1) := \mathbb{P} \left[ U_{j,m_n} \leq \alpha_n + 2 \frac{\ln(m_n)}{m_n} \right] - \mathbb{P}[U_{j,m_n} \leq \alpha_n].$$

Using the mean value theorem, for all  $j = 2, \dots, m_n$ , there exists  $\theta_{n,j} \in (0, 1)$  such that

$$D_n^{(1)}(m_n - j + 1) = 2 \frac{\ln(m_n)}{m_n} f_{\text{beta}} \left( \alpha_n + 2\theta_{n,j} \frac{\ln(m_n)}{m_n}; j, m_n - j + 1 \right).$$

Under (14), the second part of Lemma 1 entails that

$$\max_{j=2,\dots,m_n} D_n^{(1)}(m_n - j + 1) = \mathcal{O} \left( \frac{\ln(m_n)}{[m_n \alpha_n (1 - \alpha_n)]^{1/2}} \right). \quad (20)$$

It remains now to deal with the probability  $\mathbb{P}(\bar{A}_n)$ . Let  $E_1, \dots, E_{m_n+1}$  be independent standard exponential random variables. From Rényi's representation theorem,

$$\mathbb{P}(\bar{A}_n) = \mathbb{P} \left( \max_{j=2,\dots,m_n} \frac{E_j}{E_1 + \dots + E_{m_n+1}} > 2 \frac{\ln(m_n)}{m_n} \right).$$

Let  $T_{m_n+1} := (E_1 + \dots + E_{m_n+1}) / (m_n + 1)$ . From the law of large numbers,  $T_{m_n+1} \xrightarrow{a.s.} 1$  and thus, for all  $\eta \in (0, 1/4)$ , there exists  $N_\eta \in \mathbb{N} \setminus \{0\}$  such that for all  $n \geq N_\eta$ ,  $\mathbb{P}(T_{m_n+1} > 1 - \eta) = 1$ . As a consequence, for  $n \geq N_\eta$

$$\begin{aligned} \mathbb{P}(\bar{A}_n) &= \mathbb{P} \left( \left\{ E_{m_n-1,m_n-1} > 2 \frac{(m_n+1) \ln(m_n)}{m_n} T_{m_n+1} \right\} \cap \{T_{m_n+1} > 1 - \eta\} \right) \\ &\leq \mathbb{P} \left( E_{m_n-1,m_n-1} > 2(1 - \eta) \frac{(m_n+1) \ln(m_n)}{m_n} \right) \sim m_n^{2\eta-1}. \end{aligned}$$

Since for  $\eta \in (0, 1/4)$ ,

$$m_n^{2\eta-1} = o \left( \frac{\ln(m_n)}{[m_n \alpha_n (1 - \alpha_n)]^{1/2}} \right),$$

we have shown that

$$\mathbb{P}(\bar{A}_n) = \mathcal{O} \left( \frac{\ln(m_n)}{[m_n \alpha_n (1 - \alpha_n)]^{1/2}} \right). \quad (21)$$

By gathering (18), (19), (20) and (21) we get (17) and the proof is complete. ■

For  $i = 1, \dots, n$ , let  $V_i := S(Y_i|X_i)$  and  $V_i^{(x_0)} := S(Y_i^{(x_0)}|X_i^{(x_0)})$ .

**Lemma 3** *i) The random variables  $V_1, \dots, V_n$  are independent standard uniform random variables. Furthermore, they are independent from  $X_1, \dots, X_n$ .  
ii) The random variables  $V_1^{(x_0)}, \dots, V_n^{(x_0)}$  are independent standard uniform random variables.*

**Proof** – i) Since the random pairs  $\{(X_i, Y_i), i = 1, \dots, n\}$  are independent copies of  $(X, Y)$ , the random variables  $V_1, \dots, V_n$  are  $n$  independent copies of  $V = S(Y|X)$ . Now, for all  $t \in [0, 1]$ , denoting by  $f_X$  the probability density function of  $X$ ,

$$\mathbb{P}(V \leq t) = \int \mathbb{P}[S(Y|x) \leq t|X = x]f_X(x)dx = \int S[Q(t|x)|x]f_X(x)dx = t,$$

and thus  $V$  is a standard uniform random variable. To prove that the random variables  $V_1, \dots, V_n$  are independent from  $X_1, \dots, X_n$ , it suffices to prove that  $X$  and  $V$  are independent. Let  $A \in \mathcal{B}(\mathbb{R}^p)$  and  $t \in [0, 1]$ ,

$$\begin{aligned} \mathbb{P}[\{V \leq t\} \cap \{X \in A\}] &= \int \mathbb{P}[\{Y \geq Q(t|x)\} \cap \{x \in A\}|X = x]f_X(x)dx \\ &= t \int \mathbb{I}_A(x)f_X(x)dx = t\mathbb{P}[X \in A], \end{aligned}$$

proving the independence.

ii) Let  $(t_1, \dots, t_n) \in [0, 1]^n$ . Let  $\Sigma_n$  be the set of the permutations of  $\{1, \dots, n\}$ . One has

$$\begin{aligned} &\mathbb{P}[\{V_1^{(x_0)} \leq t_1\} \cap \dots \cap \{V_n^{(x_0)} \leq t_n\}] \\ &= \sum_{\sigma \in \Sigma_n} \mathbb{P}[\{V_{\sigma(1)} \leq t_1, \dots, V_{\sigma(n)} \leq t_n\} \cap \{\|X_{\sigma(1)} - x_0\| \leq \dots \leq \|X_{\sigma(n)} - x_0\|\}] \end{aligned}$$

From i), since the standard uniform random variables  $V_1, \dots, V_n$  are independent from  $X_1, \dots, X_n$ ,

$$\begin{aligned} &\mathbb{P}[\{V_1^{(x_0)} \leq t_1\} \cap \dots \cap \{V_n^{(x_0)} \leq t_n\}] \\ &= \sum_{\sigma \in \Sigma_n} \mathbb{P}[V_{\sigma(1)} \leq t_1, \dots, V_{\sigma(n)} \leq t_n] \mathbb{P}[\|X_{\sigma(1)} - x_0\| \leq \dots \leq \|X_{\sigma(n)} - x_0\|] \\ &= t_1 \dots t_n \sum_{\sigma \in \Sigma_n} \mathbb{P}[\|X_{\sigma(1)} - x_0\| \leq \dots \leq \|X_{\sigma(n)} - x_0\|] = t_1 \dots t_n, \end{aligned}$$

and the proof is complete. ■

The next lemma is a technical result used in the proof of Proposition 2.

**Lemma 4** *Assume that there exists a function  $g_0$  satisfying condition (7). For a given interval  $I_0 \subset \mathbb{R}$  such that  $\mathbb{P}(X \in g_0^{-1}(I_0)) > 0$ , assume that condition (H2) holds for  $\mathcal{C}_0 := g_0^{-1}(I_0)$  then, for all  $\varepsilon \in (0, 1)$ ,*

$$\begin{aligned} &\mathbb{P}[g_0(X) \in B_{y,\varepsilon}|X \in \mathcal{C}_0] \leq \varepsilon \\ &\text{and } \lim_{y \rightarrow \infty} \mathbb{P}[g_0(X) \in B_{y,\varepsilon}|\{X \in \mathcal{C}_0\} \cap \{Y > y\}] = 1, \end{aligned}$$

where  $B_{y,\varepsilon} := \{z \in I_0; \varphi_y(z) \geq q_y(\varepsilon/2|\mathcal{C}_0)\}$ .

**Proof of Lemma 4** – To not overload the equations, we write in the rest of the proof  $p_y(\cdot|\mathcal{C}_0) =: p_{y,0}(\cdot)$ . The corresponding quantile function is denoted  $q_{y,0}(\cdot) = \inf\{t; p_{y,0}(t) \leq \cdot\}$ . Note also that  $\{X \in \mathcal{C}_0\} = \{g_0(X) \in I_0\}$ . Let us prove first that  $\mathbb{P}(g_0(X) \in B_{y,\varepsilon} | g_0(X) \in I_0) \leq \varepsilon$ . We start with

$$\begin{aligned} & \mathbb{P}(\{g_0(X) \in B_{y,\varepsilon}\} \cap \{g_0(X) \in I_0\}) \\ &= \mathbb{P}\left(\left\{S(y|X) \geq q_{y,0}(\varepsilon/2) \frac{S(y|X)}{\varphi_y(g_0(X))}\right\} \cap \{X \in \mathcal{C}_0\}\right). \end{aligned}$$

From (7), for all  $\delta > 0$ , there exists  $y_0$  such that for all  $y \geq y_0$ ,

$$1 - \delta \leq \frac{S(y|X)}{\varphi_y(g_0(X))} \leq 1 + \delta, \quad (22)$$

almost surely. Hence, for  $y \geq y_0$ ,

$$\begin{aligned} & \mathbb{P}(\{S(y|X) \geq (1 + \delta)q_{y,0}(\varepsilon/2)\} \cap \{X \in \mathcal{C}_0\}) \\ & \leq \mathbb{P}(\{g_0(X) \in B_{y,\varepsilon}\} \cap \{g_0(X) \in I_0\}) \\ & \leq \mathbb{P}(\{S(y|X) \geq (1 - \delta)q_{y,0}(\varepsilon/2)\} \cap \{X \in \mathcal{C}_0\}). \end{aligned}$$

Since  $p_{y,0}$  is the survival function of  $S(y|X)$  given  $X \in \mathcal{C}_0$ , we have for  $y \geq y_0$ ,

$$p_{y,0}((1 + \delta)q_{y,0}(\varepsilon/2)) \leq \mathbb{P}[g_0(X) \in B_{y,\varepsilon} | g_0(X) \in I_0] \leq p_{y,0}((1 - \delta)q_{y,0}(\varepsilon/2)).$$

Now, since (H2) holds with  $\mathcal{C}_0$ , there exists  $y_1$  such that for  $y \geq y_1$ ,

$$\frac{q_{y,0}(\varepsilon)}{q_{y,0}(\varepsilon/2)} < 1 - \delta \text{ and } \frac{q_{y,0}(\varepsilon/4)}{q_{y,0}(\varepsilon/2)} > 1 + \delta.$$

Hence,  $q_{y,0}(\varepsilon) < (1 - \delta)q_{y,0}(\varepsilon/2)$  and by applying the non-increasing and right-continuous function  $p_{y,0}$ , one has that  $\varepsilon \geq p_{y,0}((1 + \delta)q_{y,0}(\varepsilon/2))$ . In the same way,  $\varepsilon/4 \leq p_{y,0}((1 - \delta)q_{y,0}(\varepsilon/2))$ . As a consequence, for  $y \geq \max(y_0, y_1)$ ,

$$\varepsilon/4 \leq \mathbb{P}[g_0(X) \in B_{y,\varepsilon} | g_0(X) \in I_0] \leq \varepsilon,$$

proving the first part of the lemma. Now, let us prove that  $\pi_y(\varepsilon) := \mathbb{P}(g_0(X) \in B_{y,\varepsilon} | \{g_0(X) \in I_0\} \cap \{Y > y\})$  converges to 1 as  $y \rightarrow \infty$ . It suffices to prove that  $\pi_y(\varepsilon)/\bar{\pi}_y(\varepsilon) \rightarrow \infty$  as  $y \rightarrow \infty$  where  $\bar{\pi}_y(\varepsilon) = 1 - \pi_y(\varepsilon)$ . First, denoting by  $B_{y,\varepsilon}^C = I_0 \setminus B_{y,\varepsilon}$  the complement of the set  $B_{y,\varepsilon}$  in  $I_0$ ,

$$\bar{\pi}_y(\varepsilon) = \frac{1}{\mathbb{P}(\{Y > y\} \cap \{g_0(X) \in I_0\})} \int_{\mathcal{X}} \mathbb{I}_{B_{y,\varepsilon}^C}(g_0(x)) S(y|x) f_X(x) dx.$$

Using (22) and the fact that  $\varphi_y(g_0(x)) \leq q_{y,0}(\varepsilon/2)$  for  $g(x) \in B_{y,\varepsilon}^C$ , one has for  $y \geq \max(y_0, y_1)$  that

$$\bar{\pi}_y(\varepsilon) \leq \frac{(1 + \delta)(1 - \varepsilon/4)}{\mathbb{P}(Y > y | g_0(X) \in I_0)} q_{y,0}(\varepsilon/2).$$

Next using similar arguments and the fact that  $B_{y,\varepsilon/2} \subset B_{y,\varepsilon}$ ,

$$\begin{aligned} \pi_y(\varepsilon) & \geq \frac{1}{\mathbb{P}(\{Y > y\} \cap \{g_0(X) \in I_0\})} \int_{\mathcal{X}} \mathbb{I}_{B_{y,\varepsilon/2}}(g(x)) S(y|x) f_X(x) dx \\ & \geq \frac{(1 - \delta)\varepsilon/8}{\mathbb{P}(Y > y | g_0(X) \in I_0)} q_{y,0}(\varepsilon/4). \end{aligned}$$

The proof is then complete by using condition (H2). ■

In the next result, we prove that the technical condition (H2) is satisfied for conditional heavy-tailed distributions.

**Lemma 5** *Let us consider the random vector  $(X, Y)$  such that for  $y > 0$  and  $x \in \mathcal{X} \subset \mathbb{R}^p$ ,  $S(y|x) = y^{-1/\gamma(x)} \mathcal{L}(y, x)$ , where  $\gamma$  is a positive function defined on  $\mathcal{X}$  and for all  $x \in \mathcal{X}$ ,  $\mathcal{L}(\cdot|x)$  is a slowly varying function. Let  $\mathcal{C} \subset \mathcal{X}$  with  $\mathbb{P}(X \in \mathcal{C}) > 0$ . If the cumulative distribution of  $\gamma(X)$  given  $X \in \mathcal{C}$  is continuous and if*

$$\lim_{y \rightarrow \infty} \sup_{x \in \mathcal{C}} \frac{\ln \mathcal{L}(y, x)}{\ln y} = 0 \quad (23)$$

*then condition (H2) holds.*

**Proof of Lemma 5** – For all  $y$  and  $t \in (0, 1)$ , let us introduce the set  $A_y(t) := \{x \in \mathcal{X}; S(y|x) > t\}$ . One has

$$A_y(t) = \left\{ x \in \mathcal{X}; \gamma(x) > \left( -\frac{\ln t}{\ln y} + \frac{\ln \mathcal{L}(y, x)}{\ln y} \right)^{-1} \right\}.$$

Condition (23) entails that for all  $\delta > 0$ , there exists  $y_\delta$  such that for all  $y > y_\delta$ ,  $A_y^-(t) \subset A_y(t) \subset A_y^+(t)$  with

$$A_y^\pm(t) := \left\{ x \in \mathcal{X}; \gamma(x) \geq \left( -\frac{\ln t}{\ln y} \pm \delta \right)^{-1} \right\}.$$

Hence, denoting by  $G$  the survival function of  $\gamma(X)$  given that  $X \in \mathcal{C}$ , one has for all  $y$  and  $t \in (0, 1)$

$$G \left[ \left( -\frac{\ln t}{\ln y} - \delta \right)^{-1} \right] \leq p_y(t|\mathcal{C}) \leq G \left[ \left( -\frac{\ln t}{\ln y} + \delta \right)^{-1} \right]. \quad (24)$$

Let  $G^\leftarrow$  be the generalized inverse of  $G$ . For  $(\eta, d) \in (0, 1)^2$ , replacing  $t$  by  $y^{-1/G^\leftarrow(\eta)-\delta}$  in the first inequality leads to  $p_y(y^{-1/G^\leftarrow(d\eta)-\delta}|\mathcal{C}) \geq d\eta$ . Applying the function  $q_y(\cdot|\mathcal{C})$  (the inverse of  $p_y(\cdot|\mathcal{C})$ ) conducts us to the inequality  $y^{-1/\bar{G}^\leftarrow(d\eta)-\delta} \leq q_y(d\eta|\mathcal{C})$ . Similarly, using the second inequality in (24), one has for  $\eta \in (0, 1)$  that  $y^{-1/\bar{G}^\leftarrow(\eta)+\delta} \geq q_y(\eta|\mathcal{C})$ . Gathering these inequalities yields

$$\frac{q_y(d\eta|\mathcal{C})}{q_y(\eta|\mathcal{C})} \leq y^{1/G^\leftarrow(d\eta)-1/G^\leftarrow(\eta)+2\delta}.$$

This inequality is true for all  $\delta > 0$ . Since  $G$  is continuous, one can take

$$0 < \delta < \frac{1}{2} \left( \frac{1}{G^\leftarrow(d\eta)} - \frac{1}{G^\leftarrow(\eta)} \right),$$

to conclude the proof. ■



## 5.2 Proofs of main results

**Proof of Theorem 1** – Using the notations introduced in Lemma 3, we start with

$$\begin{aligned}\mathbb{P}[Y_{j,k_n}^{(x_0)} \leq Q(\alpha_n|x_0)] &= \mathbb{P}\left[\sum_{i=1}^{k_n} \mathbb{I}_{(Q(\alpha_n|x_0), \infty)}(Y_i^{(x_0)}) \leq k_n - j\right] \\ &= \mathbb{P}\left[\sum_{i=1}^{k_n} \mathbb{I}_{(-\infty, S[Q(\alpha_n|x_0)|X_i^{(x_0)}])}(V_i^{(x_0)}) \leq k_n - j\right]\end{aligned}$$

Let  $\varepsilon_n := \omega^{1/2}(\alpha_n, h_n; x_0)$ . Since for all  $i = 1, \dots, k_n$ ,

$$\alpha_n(1 - \varepsilon_n) \leq S[Q(\alpha_n|x_0)|X_i^{(x_0)}] \leq \alpha_n(1 + \varepsilon_n),$$

one has that

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{k_n} \mathbb{I}_{(-\infty, \alpha_n(1+\varepsilon_n))}(V_i^{(x_0)}) \leq k_n - j\right] &\leq \mathbb{P}[Y_{j,k_n}^{(x_0)} \leq Q(\alpha_n|x_0)] \\ &\leq \mathbb{P}\left[\sum_{i=1}^{k_n} \mathbb{I}_{(-\infty, \alpha_n(1-\varepsilon_n))}(V_i^{(x_0)}) \leq k_n - j\right].\end{aligned}$$

Remarking that from Lemma 3, ii)

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{k_n} \mathbb{I}_{(-\infty, \alpha_n(1\pm\varepsilon_n))}(V_i^{(x_0)}) \leq k_n - j\right] &= \mathbb{P}\left[V_{k_n-j+1,k_n}^{(x_0)} > \alpha_n(1\pm\varepsilon_n)\right] \\ &= \overline{F}_{\text{beta}}(\alpha_n(1\pm\varepsilon_n); k_n - j + 1, j),\end{aligned}$$

where for all  $a > 0$  and  $b > 0$ ,  $\overline{F}_{\text{beta}}(\cdot; a, b) = 1 - F_{\text{beta}}(\cdot; a, b)$ , one has

$$\begin{aligned}\overline{F}_{\text{beta}}(\alpha_n(1 + \varepsilon_n); k_n - j + 1, j) &\leq \mathbb{P}[Y_{j,k_n}^{(x_0)} \leq Q(\alpha_n|x_0)] \\ &\leq \overline{F}_{\text{beta}}(\alpha_n(1 - \varepsilon_n); k_n - j + 1, j).\end{aligned}$$

Using the mean value theorem, for all  $j = 1, \dots, k_n$ , there exists  $\tau_{n,j}^{(+)} \in (0, 1)$  and  $\tau_{n,j}^{(-)} \in (0, 1)$  such that

$$R_n(\tau_{n,j}^{(+)}; x_0) \leq \mathbb{P}[Y_{j,k_n}^{(x_0)} \leq Q(\alpha_n|x_0)] - \overline{F}_{\text{beta}}(\alpha_n; k_n - j + 1, j) \leq R_n(\tau_{n,j}^{(-)}; x_0),$$

where  $R_n(\tau_{n,j}^{(\pm)}; x_0) := \mp \alpha_n \varepsilon_n f_{\text{beta}}(\alpha_n(1 \pm \tau_{n,j}^{(\pm)} \varepsilon_n; k_n - j + 1, j))$ . Hence, Lemma 1 leads to

$$\mathbb{P}[Y_{j,k_n}^{(x_0)} \leq Q(\alpha_n|x_0)] = \overline{F}_{\text{beta}}(\alpha_n; k_n - j + 1, j) + \mathcal{O}\left(\varepsilon_n \left(\frac{k_n \alpha_n}{1 - \alpha_n}\right)^{1/2}\right),$$

uniformly on  $j = 1, \dots, k_n$ . We conclude the proof by applying Lemma 2 with  $m_n = k_n$ . ■

**Proof of Proposition 1** – Let

$$\sum_{i=1}^n \mathbb{I}_{(-\infty, h_n)}(d_e(X_i, x_0))$$

be the number of covariates in the ball of center  $x_0$  and radius  $h_n = (2k_n/[nf_X(x_0)])^{1/p}$ . To prove Proposition 1, it suffices to show that for  $n$  large enough,  $\mathbb{P}[N_n \geq k_n] = 1$ . From [5, Corollary 2.1] (see also [13, Lemma 2]), since  $nh_n^p/[\ln \ln n] \rightarrow \infty$ , one has  $N_n/(nh_n^p) \xrightarrow{a.s.} f_X(x_0)$ . Hence, for  $n$  large enough,

$$\mathbb{P}\left[\frac{N_n}{nh_n^p} > \frac{f_X(x_0)}{2}\right] = 1.$$

The end of the proof is straightforward.  $\blacksquare$

**Proof of Proposition 2** – Let  $j \in \{1, \dots, J\}$  where  $J$  is defined in condition (H1). The first step of the proof consists in showing that

$$\lim_{y \rightarrow \infty} E_{g_0}(y; I_j) = 0. \quad (25)$$

Let us introduce the following measurable sets:  $\mathcal{A}_y := \{Y > y\}$ ;  $\mathcal{A}_y^* := \{Y^* > y\}$ ;  $\mathcal{B}_j := \{g_0(X) \in I_j\}$  and  $\mathcal{B}_j^* := \{g_0(X^*) \in I_j\}$ , where  $(X^*, Y^*)$  is an independent copie of  $(X, Y)$ . For all  $\varepsilon > 0$ , let  $B_{y,\varepsilon} = \{z \in I_j; \varphi_y(z) \geq q_y(\varepsilon/2|\mathcal{C}_j)\}$  where  $\mathcal{C}_j := g_0^{-1}(I_j)$ . Finally, let  $\mathcal{B}_{j,\varepsilon} := \{g_0(X) \in B_{y,\varepsilon}\}$ . Before proving (25), let us give some results on the previous defined sets. From Lemma 4,

$$\mathbb{P}[\mathcal{B}_{j,\varepsilon} | \mathcal{B}_j] \leq \varepsilon, \quad (26)$$

and

$$\lim_{y \rightarrow \infty} \mathbb{P}[\mathcal{B}_{j,\varepsilon} | \mathcal{B}_j \cap \mathcal{A}_y] = 1. \quad (27)$$

Since on  $I_j$ ,  $\varphi_y$  admits a unique maximum point  $z_j^*$  in the interior of  $I_j$ ,  $B_{y,\varepsilon}$  is an interval included in  $I_j$  and containing  $z_j^*$ . Since  $f_0(x) \geq m$  for all  $x \in \mathcal{X}_0$ , conditions (26) conducts to

$$m \times l(B_{y,\varepsilon}) \leq \int_{B_{y,\varepsilon}} f_0(x) dx \leq \varepsilon \mathbb{P}(\mathcal{B}_j).$$

As a consequence,

$$\mathbb{P}\left[|g_0(X) - z_j^*| \leq \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \mid \mathcal{B}_{j,\varepsilon}\right] = 1. \quad (28)$$

We are now in position to prove (25). For  $y \in \mathbb{R}$ ,

$$E_{g_0}(y; I_j) = \frac{\mathbb{E}[|g_0(X) - g_0(X^*)| \mathbb{I}_{\mathcal{B}_j \cap \mathcal{B}_j^* \cap \mathcal{A}_y \cap \mathcal{A}_y^*}]}{[\mathbb{P}(\mathcal{A}_y \cap \mathcal{B}_j)]^2}.$$

Remarking that  $|g_0(X) - g_0(X^*)| = g_0(X) + g_0(X^*) - 2 \min(g_0(X), g_0(X^*))$ , one has

$$E_{g_0}(y; I_j) = \frac{2}{[\mathbb{P}(\mathcal{A}_y \cap \mathcal{B}_j)]^2} [T_{1,y} - T_{2,y}], \quad (29)$$

where  $T_{1,y} := \mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y) \mathbb{E}[g_0(X) \mathbb{I}_{\mathcal{B}_j \cap \mathcal{A}_y}]$  and  
 $T_{2,y} := \mathbb{E}[\min(g_0(X), g_0(X^*)) \mathbb{I}_{\mathcal{B}_j \cap \mathcal{B}_j^* \cap \mathcal{A}_y \cap \mathcal{A}_y^*}]$ .  
Let us first focus on the term  $T_{1,y}$ . We start with

$$T_{1,y} = \mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y) \left[ \mathbb{E}(g_0(X) \mathbb{I}_{\mathcal{B}_{j,\in} \cap \mathcal{A}_y}) + \mathbb{E}(g_0(X) \mathbb{I}_{\mathcal{B}_{j,\notin} \cap \mathcal{A}_y}) \right],$$

where  $\mathcal{B}_{j,\notin} = I_j \setminus \mathcal{B}_{j,\in}$ . From (28), and since  $\mathcal{B}_{j,\in} \subset \mathcal{B}_j$ ,

$$\begin{aligned} \mathbb{E}(g_0(X) \mathbb{I}_{\mathcal{B}_{j,\in} \cap \mathcal{A}_y}) &\leq \left( z_j^* + \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \right) \mathbb{P}(\mathcal{B}_{j,\in} \cap \mathcal{A}_y) \\ &\leq \mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y) \left( z_j^* \mathbb{P}(\mathcal{B}_{j,\in} | \mathcal{B}_j \cap \mathcal{A}_y) + \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \right). \end{aligned}$$

From (27), for all  $\varepsilon > 0$ , there exists  $y_{1,\varepsilon} \in \mathbb{R}$  such that for all  $y > y_{1,\varepsilon}$ ,  $1 - \varepsilon \leq \mathbb{P}(\mathcal{B}_{j,\in} | \mathcal{B}_j \cap \mathcal{A}_y) \leq 1 + \varepsilon$ . Furthermore, since  $B_{y,\varepsilon}$  is a closed interval, there exists  $c_j > 0$  such that  $|z_j^*| \leq c_j$  and thus,  $z_j^* \mathbb{P}(\mathcal{B}_{j,\in} | \mathcal{B}_j \cap \mathcal{A}_y) \leq z_j^* + \varepsilon c_j$ . Hence,

$$\mathbb{E}(g_0(X) \mathbb{I}_{\mathcal{B}_{j,\in} \cap \mathcal{A}_y}) \leq \mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y) \left( z_j^* + \varepsilon c_j + \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \right). \quad (30)$$

Moreover, for all  $y > y_{1,\varepsilon}$

$$\begin{aligned} \mathbb{E}(g_0(X) \mathbb{I}_{\mathcal{B}_{j,\notin} \cap \mathcal{A}_y}) &\leq c_j \mathbb{P}(\mathcal{B}_{j,\notin} \cap \mathcal{A}_y) = c_j \mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y) \mathbb{P}(\mathcal{B}_{j,\notin} | \mathcal{B}_j \cap \mathcal{A}_y) \\ &\leq c_j \varepsilon \mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y) \end{aligned} \quad (31)$$

Gathering (30) and (31) yield to

$$T_{1,y} \leq [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2 \left[ z_j^* + \varepsilon \left( 2c_j + \frac{\mathbb{P}(\mathcal{B}_j)}{m} \right) \right] \quad (32)$$

for all  $y > y_{1,\varepsilon}$ . Let us now focus on the term  $T_{2,y}$ . We start with the decomposition  $T_{2,y} = T_{2,y}^{(1)} + 2T_{2,y}^{(2)} + T_{2,y}^{(3)}$  where

$$T_{2,y}^{(1)} := \mathbb{E} \left[ \min(g_0(X), g_0(X^*)) \mathbb{I}_{\mathcal{B}_{j,\in} \cap \mathcal{B}_{j,\in}^* \cap \mathcal{A}_y \cap \mathcal{A}_y^*} \right],$$

$$T_{2,y}^{(2)} := \mathbb{E} \left[ \min(g_0(X), g_0(X^*)) \mathbb{I}_{\mathcal{B}_{j,\in} \cap \mathcal{B}_{j,\notin}^* \cap \mathcal{A}_y \cap \mathcal{A}_y^*} \right],$$

and

$$T_{2,y}^{(3)} := \mathbb{E} \left[ \min(g_0(X), g_0(X^*)) \mathbb{I}_{\mathcal{B}_{j,\notin} \cap \mathcal{B}_{j,\notin}^* \cap \mathcal{A}_y \cap \mathcal{A}_y^*} \right].$$

First, from (28) and since  $\mathcal{B}_{j,\in} \subset \mathcal{B}_j$ ,

$$\begin{aligned} T_{2,y}^{(1)} &\geq \left( z_j^* - \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \right) [\mathbb{P}(\mathcal{B}_{j,\in} \cap \mathcal{A}_y)]^2 \\ &\geq [\mathbb{P}(\mathcal{B}_{j,\in} \cap \mathcal{A}_y)]^2 \left( z_j^* \mathbb{P}(\mathcal{B}_{j,\in} | \mathcal{B}_j \cap \mathcal{A}_y) - \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \right). \end{aligned}$$

Using the same arguments than those leading to (30), we obtain

$$T_{2,y}^{(1)} \geq [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2 \left( z_j^* - \varepsilon c_j - \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \right). \quad (33)$$

Now, using (26), one has for  $y > y_{1,\varepsilon}$ ,

$$\begin{aligned} T_{2,y}^{(2)} &\geq -c_j \mathbb{P}(\mathcal{B}_{j,\in} \cap \mathcal{A}_y) \mathbb{P}(\mathcal{B}_{j,\notin}^* \cap \mathcal{A}_y^*) \\ &= -c_j [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2 \mathbb{P}(\mathcal{B}_{j,\in} | \mathcal{B}_j \cap \mathcal{A}_y) \mathbb{P}(\mathcal{B}_{j,\notin} | \mathcal{B}_j \cap \mathcal{A}_y) \\ &\geq c_j \varepsilon [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2. \end{aligned} \quad (34)$$

Finally, from (28), one has for  $y > y_{1,\varepsilon}$

$$T_{2,y}^{(3)} \geq -c_j [\mathbb{P}(\mathcal{B}_{j,\in} \cap \mathcal{A}_y)]^2 \geq -c_j \varepsilon^2 [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2. \quad (35)$$

Collecting (33), (34) and (35) yield to

$$T_{2,y} \geq [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2 \left( z_j^* - \varepsilon \frac{\mathbb{P}(\mathcal{B}_j)}{m} + \varepsilon c_j + c_j \varepsilon^2 \right). \quad (36)$$

for all  $\varepsilon > 0$  and  $y > y_{1,\varepsilon}$ . Gathering (29), (32) and (36) conduct to

$$E_{g_0}(y; I_j) \leq 2\varepsilon \left( c_j + 2 \frac{\mathbb{P}(\mathcal{B}_j)}{m} - c_j \varepsilon \right),$$

proving (25) since  $\varepsilon$  can be chosen arbitrarily small.

Now, conditions (H1) and (H2) entail that there exist  $h_1^*, \dots, h_J^*$  such that for all  $j \in \{1, \dots, J\}$ ,  $z_j^* \in \mathcal{S}_{h_j^*, g_0} \subset I_j$ . Furthermore, taking

$$\varepsilon \leq \min_{j \in \{1, \dots, J\}} \mathbb{P}(g_0(X) \in \mathcal{S}_{h_j^*, g_0}),$$

conditions (26) and (27) entail that for all  $v > 0$ , there exists  $y_2 \in \mathbb{R}$  such that for all  $y \geq y_2$  and  $h \notin \{h_1^*, \dots, h_J^*\}$ ,

$$\mathbb{P}(g_0(X) \in \mathcal{S}_{h, g_0} | \mathcal{A}_y) \leq v,$$

showing that  $\text{card}(\mathcal{J}_{H,y}(v)) \leq J$ . Finally, mimicking the proof of (25), it is easy to check that for all  $h \in \{h_1^*, \dots, h_J^*\} \supset \mathcal{J}_{H,y}(v)$ ,  $E_{g_0}(y; \mathcal{S}_{h, g_0}) \rightarrow 0$  as  $y \rightarrow \infty$ , concluding the proof.  $\blacksquare$

## References

- [1] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extreme quantile regression, *Bernoulli*, **19**(5B), 2557–2589.
- [2] Basu, D. and Pereira, C.A.B. (1983). Conditional independence in statistics, *Sankhyā*, **45**(3), Series A, 324–337.
- [3] Cook, R.D. and Li, B. (2002). Dimension Reduction for Conditional Mean in Regression, *The Annals of Statistics*, **30**, 455–474.
- [4] Cook, R.D. and Weisberg, S. (1991). Discussion of "Sliced Inverse Regression for dimension reduction", *Journal of the American Statistical Association*, **86**, 328–332.

- [5] Dony, J. and Einmahl, U. (2009). *Uniform in bandwidth consistency of kernel regression estimators at fixed point*, High Dimensional Probability V: The Luminy Volume, 308–325, Institute of Mathematical Statistics, Beachwood, Ohio, USA.
- [6] Fan, Y. and Liu, R. (2016). A direct approach to inference in nonparametric and semiparametric quantile models, *Journal of Econometrics*, **191**(1), 196–216.
- [7] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Volume 1, third edition, New York, Wiley.
- [8] Gardes, L. (2015). A general estimator for the extreme value index: applications to conditional and heteroscedastic extremes, *Extremes*, **18**(3), 479–510.
- [9] Gardes, L. (2018). Tail Dimension Reduction for extreme quantile estimation, *Extremes*, **21**(1), 57–95.
- [10] Gardes, L. and Girard, S. (2008). Estimation of the Weibull tail-coefficient with linear combination of upper order statistics, *Journal of Statistical Planning and Inference*, **138**(5), 1416–1427.
- [11] Gardes, L. and Girard, S. (2016). On the estimation of the functional Weibull tail-coefficient, *Journal of Multivariate Analysis*, **146**, 29–45.
- [12] Gardes, L., Girard, S. and Guillou, A. (2011). Weibull tail-distributions revisited: a new look at some tail estimators, *Journal of Statistical Planning and Inference*, **141**, 429–444.
- [13] Gardes, L., Guillou, A. and Roman, C. (2019). Estimation of extreme conditional quantiles under a general tail first order condition, *Annals of the Institute of Statistical Mathematics*, To appear.
- [14] Goldman, M. and Kaplan, D.M. (2018). Non-parametric inference on (conditional) quantile differences and interquartile ranges using L-statistics, *The Econometrics Journal*, **21**, 136–169.
- [15] Han, S., Bian, H., Feng, Y., Liu, A., Li, ., Zeng, F. and Zhang, X. (2011). Analysis of the relationship between O<sub>3</sub>, NO and NO<sub>2</sub> in Tianjin, China, *Aerosol and Air Quality Research*, **11**, 128–139.
- [16] Hooke, R. and Jeeves, T.A. (1961). Direct search solution of numerical and statistical problems, *Journal of the ACM*, **8**(2), 212–229.
- [17] Hutson, A.D. (1999). Calculating nonparametric confidence intervals for quantiles using fractional order statistics, *Journal of applied Statistics*, **26**(3), 343–353.
- [18] Kocherginsky, M., He, X. and Mu, Y. (2005). Practical confidence intervals for regression quantiles, *Journal of Computational and Graphical Statistics*, **14**(1), 41–55.

- [19] Li, K. C. (1991). Sliced Inverse Regression for Dimension Reduction, *Journal of the American Statistical Association*, **86**, 316–342.
- [20] Lukács, E. (1975). *Stochastic convergence*, 2nd edition, Academic Press, New York.
- [21] Parzen, M.I., Wei, L.J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions, *Biometrika*, **81(2)**, 341–350.
- [22] Rambaud S.C. (2011). A note on almost sure uniform and complete convergence of a sequence of random variables, *Stochastics: An International Journal of Probability and Stochastic Processes*, **83(03)**, 215–221.
- [23] Samarov, A.M. (1993). Exploring Regression Structure Using Nonparametric Functional Estimation, *Journal of the American Statistical Association*, **88**, 836–847.
- [24] Scrucca, L. (2011). Model-based SIR for dimension reduction, *Computational Statistics and Data Analysis*, **55**, 3010–3026.
- [25] Stute, W. (1984). Asymptotic normality of nearest neighbor regression function estimates, *The Annals of Statistics*, **12**, 917–926.
- [26] Xia, Y. (2009). Model checking in regression via dimension reduction, *Biometrika*, **96(1)**, 133–148.
- [27] Yang, S.S. (1981). Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function, *Journal of the American Statistical Association*, **76**, 658–662.
- [28] Zhu, L., Huang, M. and Li, R. (2012). Semiparametric quantile regression with high-dimensional covariate, *Statistica Sinica*, **22(4)**, 1379–1401.

## Appendix A: Tables and figures

Table 1: Expressions of the functions  $g_0^{(j)}$  used in the numerical illustrations. Recall that  $x \in \mathbb{R}^p$  and that the components of  $x^2$  are the square of the components of  $x$ .

$j$	$g_0^{(j)}(x)$	$p$	$b_0^{(j)}$	Set $\mathcal{G}$
1	$(b_0^{(1)})^\top x$	1	1	$\mathcal{G}_L$
2	$(b_0^{(2)})^\top x$	2	$(1, 2)/\sqrt{5}$	$\mathcal{G}_L$
3	$(b_0^{(3)})^\top x$	4	$(0, 1, 2, 0)/\sqrt{5}$	$\mathcal{G}_L$
4	$(b_0^{(4)})^\top x$	8	$(0, 1, 2, 0, 0, 0, 1, 1)/\sqrt{7}$	$\mathcal{G}_L$
5	$ (b_0^{(5)})^\top x^{(2)} ^{1/2}$	4	$(1, 0, 0, 1)/\sqrt{2}$	$\mathcal{G}_{NL,2}$

Table 2: Values of the coverage probability for **Model 1**.

$p$	$\gamma$	$g_0(x)$	$\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$ $\alpha_{n,1}$	$\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$ $\alpha_2$	$\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$ $\alpha_{n,3}$
1	0.9	$g_0^{(1)}(x)$	0.3130	0.8770	0.9125
1	0.95	$g_0^{(1)}(x)$	0.330	0.9320	0.9410
2	0.9	$g_0^{(2)}(x)$	0.0900	0.8700	0.8040
2	0.95	$g_0^{(2)}(x)$	0.1265	0.9385	0.8530
4	0.9	$g_0^{(3)}(x)$	0.8675	0.7865	0.6265
4	0.95	$g_0^{(3)}(x)$	0.9110	0.8670	0.6975
8	0.9	$g_0^{(4)}(x)$	0.7045	0.9115	0.5215
8	0.95	$g_0^{(4)}(x)$	0.7330	0.9550	0.6095

Table 3: Values of the coverage probability for **Model 2**.

$p$	$\gamma$	$g_0(x)$	$\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$ $\alpha_{n,1}$	$\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$ $\alpha_2$	$\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$ $\alpha_{n,3}$
1	0.9	$g_0^{(1)}(x)$	0.9300	0.9270	0.9350
1	0.95	$g_0^{(1)}(x)$	0.9525	0.9640	0.9630
2	0.9	$g_0^{(2)}(x)$	0.7140	0.9280	0.8310
2	0.95	$g_0^{(2)}(x)$	0.7985	0.9630	0.8930
4	0.9	$g_0^{(3)}(x)$	0.2885	0.9245	0.6430
4	0.95	$g_0^{(3)}(x)$	0.3565	0.9640	0.7200
8	0.9	$g_0^{(4)}(x)$	0.3070	0.8805	0.6230
8	0.95	$g_0^{(4)}(x)$	0.3785	0.9400	0.7190

Table 4: Reduction of dimension for **Model 1** with  $\alpha_n = \alpha_{n,3} = 8 \ln(n)/n$ .

$p$	$g_0^{(j)}(x)$	$\mathbb{E}(\delta_n^{(j)})$ ( $\text{sd}(\delta_n^{(j)})$ )	$\gamma$	$\tilde{p}_{\gamma, \alpha_n}(\hat{d}_0^{(j)}, x_0)$	$\tilde{p}_{\gamma, \alpha_n}(d_0^{(j)}, x_0)$	$\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$
2	$g_0^{(2)}(x)$	0.0163	0.9	0.8805	0.9255	0.8040
		(0.0233)	0.95	0.9210	0.9550	0.8530
4	$g_0^{(3)}(x)$	0.0265	0.9	0.8665	0.9235	0.6225
		(0.0235)	0.95	0.9060	0.9510	0.6975
4	$g_0^{(5)}(x)$	0.1047	0.9	0.7965	0.8910	0.2470
		(0.0905)	0.95	0.8370	0.9065	0.2990
8	$g_0^{(4)}(x)$	0.0595	0.9	0.8895	0.9345	0.5215
		(0.0412)	0.95	0.9260	0.9615	0.6095

Table 5: Reduction of dimension for **Model 2** with  $\alpha_n = \alpha_{n,3} = 8 \ln(n)/n$ .

$p$	$g_0^{(j)}(x)$	$\mathbb{E}(\delta_n^{(j)})$ ( $\text{sd}(\delta_n^{(j)})$ )	$\gamma$	$\tilde{p}_{\gamma, \alpha_n}(\hat{d}_0^{(j)}, x_0)$	$\tilde{p}_{\gamma, \alpha_n}(d_0^{(j)}, x_0)$	$\tilde{p}_{\gamma, \alpha_n}(d_e, x_0)$
2	$g_0^{(2)}(x)$	0.0113	0.9	0.9205	0.9275	0.8310
		(0.0161)	0.95	0.9445	0.9530	0.8930
4	$g_0^{(3)}(x)$	0.0213	0.9	0.8910	0.9280	0.6430
		(0.0189)	0.95	0.9310	0.9515	0.7200
4	$g_0^{(5)}(x)$	0.0869	0.9	0.8550	0.8820	0.4940
		(0.0793)	0.95	0.8840	0.9030	0.5775
8	$g_0^{(4)}(x)$	0.0763	0.9	0.9025	0.9385	0.6230
		(0.1918)	0.95	0.9270	0.9615	0.7190



Figure 1: For **Model 1** with  $\gamma = 0.1$ , values of  $p_{\gamma, \alpha_n}(k_n, d_e, x_0)$  as a function of  $k_n$  with  $\alpha_n = 1 - 8 \ln(n)/n$  (left panels),  $\alpha_n = 1/2$  (middle panels) and  $\alpha_n = 8 \ln(n)/n$  (right panels). Top panels:  $g_0 = g_0^{(1)}$  ( $p = 1$ ), center panels:  $g_0 = g_0^{(2)}$  ( $p = 2$ ), bottom panels:  $g_0 = g_0^{(3)}$  ( $p = 4$ ). The horizontal full line is the prespecified coverage probability ( $\gamma = 0.9$ ) and the dashed horizontal line represents the coverage probability obtained with the selected value  $\widehat{k}_n^{(sel)}$ .

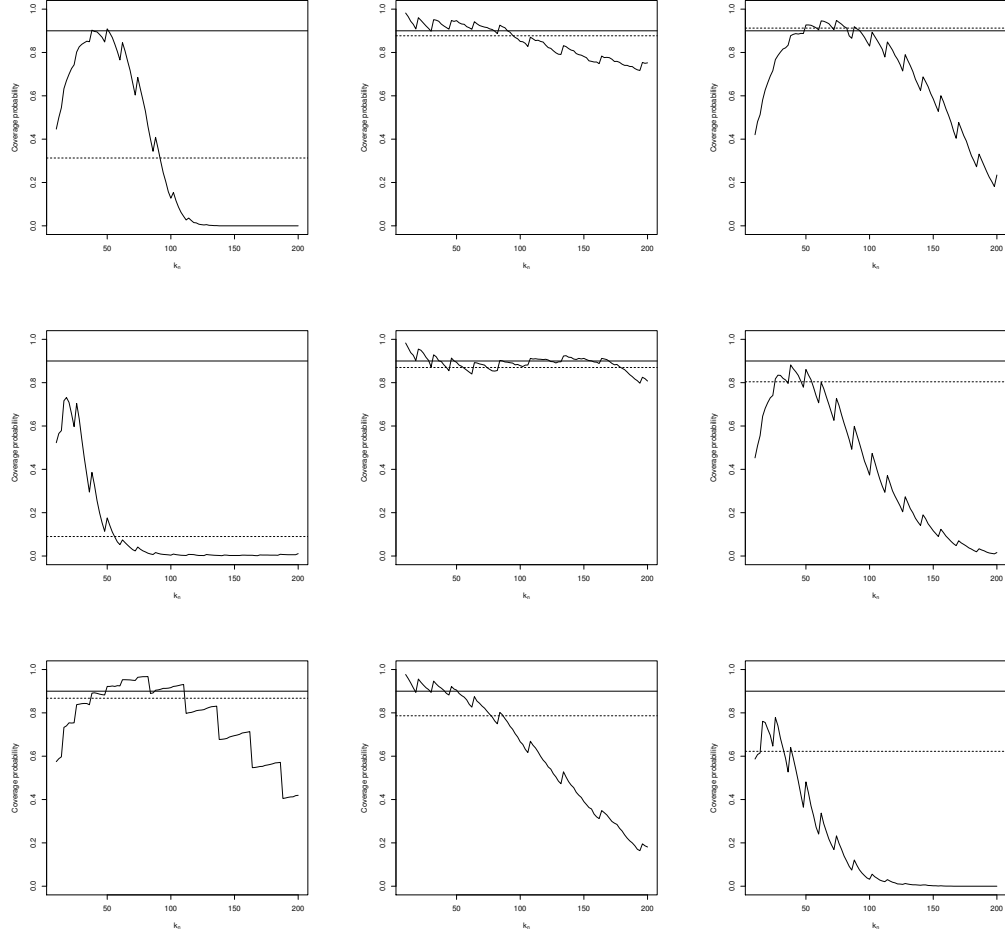


Figure 2: Histogram of the observations of  $g_0(X)$  (left panel) and  $g_1(X)$  (right panel) given that  $Y > y$  with  $y = 10\,000$ .

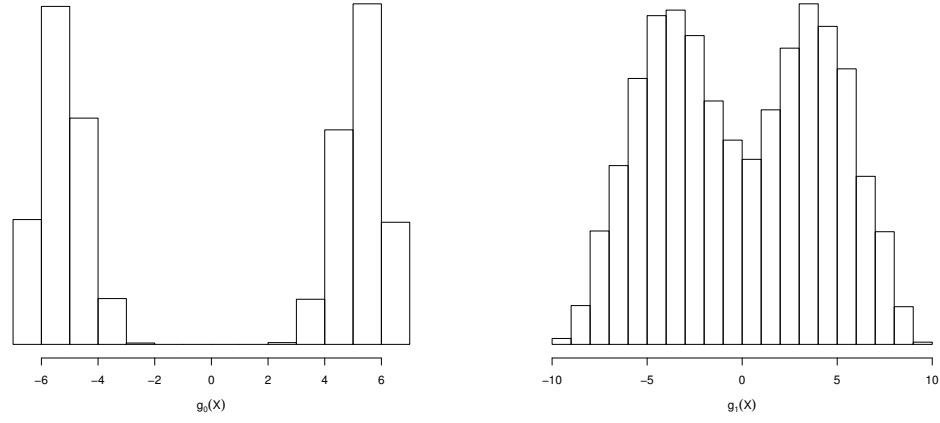


Figure 3: For **Model 1** with  $\gamma = 0.1$ , values of the coverage probabilities as a function of  $k_n$  with  $\alpha_n = 8 \ln(n)/n$ . Full line:  $p_{\gamma, \alpha_n}(k_n, d_e, x_0)$ , dashed line:  $p_{\gamma, \alpha_n}(k_n, \hat{d}_0^{(j)}, x_0)$  and dotted line  $p_{\gamma, \alpha_n}(k_n, d_0^{(j)}, x_0)$ . Top panels:  $g_0 = g_0^{(2)}$  ( $p = 2$ ) and  $g_0 = g_0^{(3)}$  ( $p = 4$ ), bottom panels:  $g_0 = g_0^{(4)}$  ( $p = 8$ ) and  $g_0 = g_0^{(5)}$  ( $p = 4$ ). The horizontal full line is the prespecified coverage probability ( $\gamma = 0.9$ ) and the dashed horizontal line represents the coverage probability obtained with the selected value  $\hat{k}_n^{(sel)}$ .

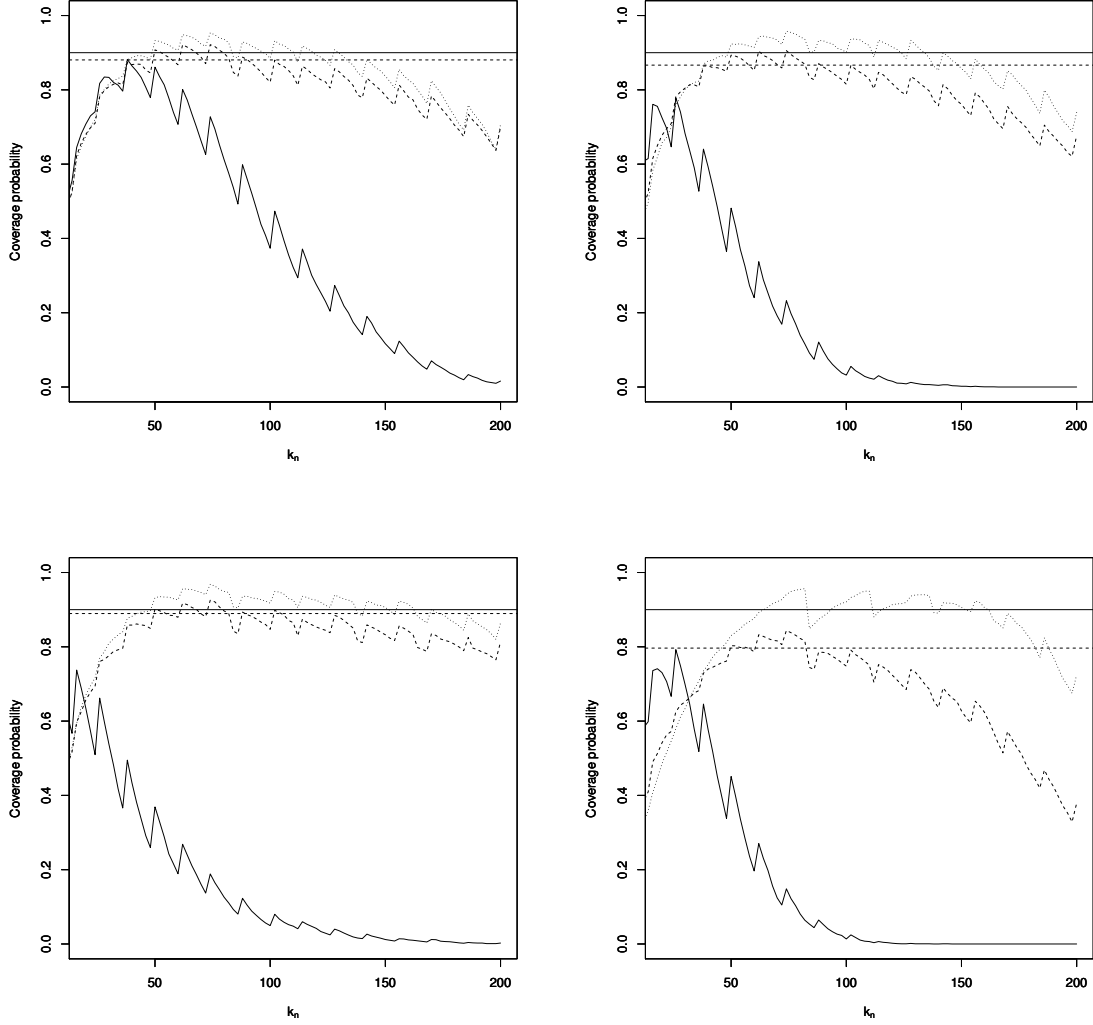
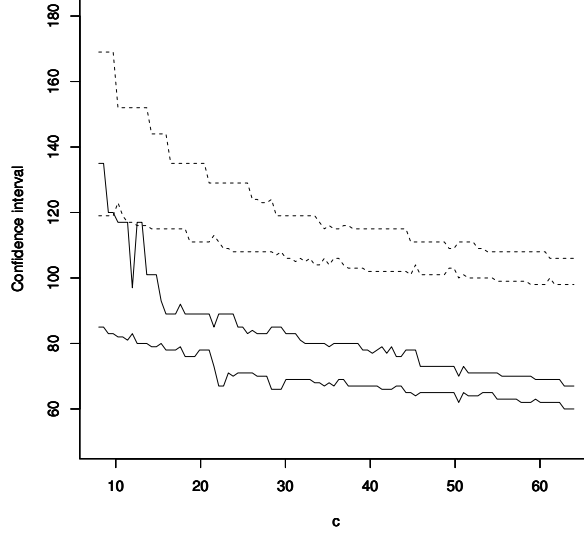


Figure 4: Confidence intervals of  $Q(c \ln(n)/n | x_0)$  with prespecified coverage probability of  $\gamma = 0.9$  as a function of  $c \in [8, 64]$ : full lines, situation 1, dashed lines, situation 2.



## Appendix B: Coordinate search method

We present here the coordinate search algorithm to solve the minimization problem:

$$\min_{x \in \mathbb{R}^p} \Phi(x),$$

where  $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}$  can be any complicated function. Let  $D := [I_p, -I_p]$  be the  $p \times 2p$  matrix where  $I_p$  is the  $p \times p$  identity matrix. For  $i \in \{1, \dots, 2p\}$ , we denote by  $D_i$  the  $i$ th column of  $D$ .

**Initialization** – Let  $x_0^* \in \mathbb{R}^p$  be an initial guess of the solution. The setting parameters of the algorithm are:  $\alpha_0 > 0$ ,  $0 < \alpha_{tol} < \alpha_0$  and  $\zeta \in (0, 1)$ . Let  $k \in \mathbb{N}$ .

**Step  $k$**  – If  $\alpha_k \leq \alpha_{tol}$  then STOP. Else,

- if

$$\Phi(x_k^*) \leq \min_{i=1, \dots, 2p} \Phi(x_k^* + \alpha_k D_i)$$

then  $x_{k+1}^* = x_k^*$  and  $\alpha_{k+1} = \zeta \alpha_k$ . Go to Step  $k + 1$ .

- if

$$\Phi(x_k^*) > \min_{i=1, \dots, 2p} \Phi(x_k^* + \alpha_k D_i)$$

then  $\alpha_{k+1} = \zeta^{-1}\alpha_k$  and

$$x_{k+1}^* = \arg \min \{ \Phi(x); x \in \{x_k^* + \alpha_k D_1, \dots, x_k^* + \alpha_k D_{2p}\} \}.$$

Go to Step  $k + 1$ .

This algorithm is used to solve (9). Recall that in Section 3.3, the set of function  $\mathcal{G}$  is a set of parametric functions with parameter  $b \in \Theta_p \subset \mathbb{R}^p$ . For any  $\tilde{b} \in \mathbb{R}^p$  let

$$b := \frac{u^\top \tilde{b}}{|u^\top \tilde{b}|(\tilde{b}^\top \tilde{b})^{1/2}} \tilde{b} \in \Theta_p,$$

be the corresponding vector in  $\Theta_p$ . Denoting by  $g_b$  a function in  $\mathcal{G}$ , the solution of (9) is obtained by applying the coordinate search method to the function  $\Phi$  defined for all  $\tilde{b} \in \mathbb{R}^p$  by

$$\Phi(\tilde{b}) := \sum_{J_{H,g_b}} \widehat{E}_{g_b}(\beta_n; \mathcal{S}_{h,g_b}) + \lambda \text{card}(J_{H,g_b}).$$

The setting parameters of the algorithm are fixed to  $\alpha_0 = 5$ ,  $\alpha_{tol} = 0.05$  and  $\zeta = 1/2$ .