Nonparametric confidence interval for conditional quantiles with large-dimensional covariates

Laurent Gardes

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 γ - nonparametric confidence interval for a conditional quantile with a level depending on the sample size. When this level tends 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 the order statistics selected with a nearest neighbor approach by a Beta distribution. We show that its coverage probability converges to the preselected probability γ 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 γ. This is a well known consequence of the data sparsity especially in the tail of the distribution. In a second part, a dimension reduction procedure is proposed in order to select more appropriate nearest neighbors in the right tail of the distribution and in turn to obtain a better coverage probability for extreme conditional quantiles. This procedure is based on the Tail Conditional Independence assumption introduced in (Gardes, *Extremes*, 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 model the level of ozone in the air and

X be the vector gathering the concentration of other pollutants and weather conditions (e.g., Han et al. [17]). The relation between X and Y is commonly studied through the conditional expectation $\mathbb{E}(Y \mid X)$. An alternative way is to analyze conditional quantiles of Y given X. Recall that for all $x \in \mathcal{X} \subset \mathbb{R}^p$, $\mathcal{X}$ being the support of X, and for $\alpha \in (0,1)$, the $(1-\alpha)$-conditional quantile of Y given $X = x$ is $Q(\alpha \mid x) = \inf\{y;\ S(y \mid x) \leq \alpha\}$, where $S(\cdot \mid x) := \mathbb{P}(Y > \cdot \mid X = x)$ is the conditional survival function of Y given $X = x$. Starting from n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of the random vector (X, Y), conditional quantile estimation has been investigated by several authors see for example Stute [27] and Yang [30] for the case of a fixed level α and Gardes [9] and Gardes et al. [14] for an extreme level $\alpha = \alpha_n \to 0$. Instead of a point estimation of $Q(\alpha \mid x_0)$ where x_0 is a given point in $\mathcal{X}$, we are interested here in the construction of a confidence interval for conditional quantiles. More specifically, our goal is to find a random interval $[A_{n,\gamma}(x_0), B_{n,\gamma}(x_0)]$ such that,

$$\lim_{n\to\infty} \mathbb{P}\left\{[A_{n,\gamma}(x_0), B_{n,\gamma}(x_0)] \ni Q(\alpha \mid x_0)\right\} = \gamma, \tag{1}$$

where $\gamma \in (0,1)$ is a preselected probability. Usually, we take $\gamma = 0.9$ or 0.95. To allow us to make inference on the right and left tails of the conditional distribution, we also consider the case where $\alpha = \alpha_n$ depends on the sample size n and tends to 0 or 1 as the sample size increases. In the application on ozone concentration, this can be of high interest since large ozone levels in the air may cause serious effects on public health and on the environment.
The literature on the construction of confidence interval for conditional quantiles is, up to our knowledge, only dedicated to the case where α is a fixed value in $(0,1)$. Several approaches have been considered.

The first one is called direct approach and is discussed for instance in Fan and Liu [7]). Let $q(\cdot \mid x_0)$ be the first derivative of $Q(\cdot \mid x_0)$ and let $0 < \alpha_1 < \alpha_2 < 1$. The construction of the confidence interval is based on the existence of a random process $\widehat{Q}_n(\cdot \mid x_0)$ indexed by $\alpha \in [\alpha_1, \alpha_2]$ for which $c_n(\widehat{Q}_n(\cdot \mid x_0) - Q(\cdot \mid x_0))$ converges to a centered Gaussian process with variance $q^2(\cdot \mid x_0)\sigma^2(\cdot \mid x_0)$ for some positive sequence (c_n). It can be shown that the coverage probability of the interval

$$\left[\widehat{Q}_n\left(\alpha - c_n^{-1}u_{(1+\gamma)/2}\sigma(\alpha \mid x_0) \mid x_0\right), \widehat{Q}_n\left(\alpha + c_n^{-1}u_{(1+\gamma)/2}\sigma(\alpha \mid x_0) \mid x_0\right)\right],$$

converges to γ for any $\alpha \in [\alpha_1, \alpha_2]$ where u_β is the β-quantile of a standard normal distribution. The main drawback of the direct approach is that in most cases, the sequence c_n depends on unknown quantities, such as 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. [23] and Kocherginsky et al. [20]. Unfortunately, these methods are often time consuming.

A last approach to construct confidence interval for conditional quantiles is based on order statistics. The order statistic method has been first introduced

in the unconditional case, see e.g., Thompson [28], Hutson [19] and David and Nagaraja [5]. Let us briefly recall the construction procedure. Assume that $Y_1, \ldots, Y_n$ are independent and identically distributed random variables with common survival function $S_Y(\cdot)$ and quantile function $Q_Y(\cdot)$. Denoting by $Y_{1,n} \leq \ldots \leq Y_{n,n}$ the order statistics, if $S_Y(\cdot)$ is a continuous and strictly decreasing 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 $U_{1,n} \leq \ldots \leq U_{n,n}$ are the order statistics associated to independent standard uniform random variables. Denote by $F_{\text{beta}}(\cdot; a, b)$ the distribution function of a Beta distribution with parameters $a > 0$ and $b > 0$ and let

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

for $\gamma \in (0,1)$, $m \in \mathbb{N} \setminus \{0\}$ and $\alpha \in (0,1)$ with the convention $\max\{\emptyset\} = +\infty$ and $\min\{\emptyset\} = -\infty$. Since $U_{n-j+1,n}$ is distributed as a Beta distribution of parameters $n - j + 1$ and n, one can show that

$$\lim_{n\to\infty} \mathbb{P}\left\{ [Y_{\mathcal{L}_\gamma(n,\alpha),n}, Y_{\mathcal{R}_\gamma(n,\alpha),n}] \ni Q_Y(\alpha) \right\} = \gamma.$$

This method of construction has been recently adapted by Goldman and Kaplan [16] to the conditional case but always for a fixed quantile level α.

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 for extreme conditional quantiles i.e., when the quantile level depends on n and tends to 0 or 1 as the sample size increases. The construction of confidence intervals for extreme conditional quantiles is more challenging because there are fewer observations available in the tail.

The nearest neighbors method 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 intervals with bad coverage probabilities. This is the well known curse of dimensionality phenomenon. To overcome this problem, one way is to assume the existence of a function $g_0 : \mathbb{R}^p \to \mathbb{R}$ 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 $X \perp\!\!\!\perp Y \mid g_0(X)$, see Basu and Pereira [1]. 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., that $g_0(x) = b_0^\top x$ for all $x \in \mathbb{R}^p$, where b_0 is a given vector in $\mathbb{R}^p$. This corresponds to the single-index model introduced in a regression context for instance by Li [21]. This single-index structure has been considered by Zhu et al. [31] for the estimation of conditional quantiles when the level α is fixed. Finding the distance reduces to finding the direction b_0. Its estimation has received much attention in the literature; see Li [21] for the classical Sliced Inverse Regression (SIR) method, Cook and Weisberg [3], Samarov [25] and Cook and Li [2]).

Our second contribution is the proposition of a new data driven procedure to find an appropriate distance to use in the nearest neighbors selection process. This distance is then used in the nearest neighbors order statistics approach for the construction of confidence intervals for conditional quantiles with extreme levels $\alpha = \alpha_n \to 0$. To reach this goal, we start with the dimension reduction assumption introduced in Gardes [10]. Roughly speaking, for some function $g_0 : \mathbb{R}^p \to \mathbb{R}$, we suppose that $S(y \mid x_0)$ is equivalent, as y goes to infinity, to a function depending on x_0 only through $g_0(x_0)$. Hence, inference on extreme conditional quantiles of Y given X can be achieved only by using the information brought by the reduced covariate $g_0(X)$ and a good way to measure the closeness of the data to x_0 is to use the pseudo-distance defined for $(x, y) \in \mathbb{R}^{2p}$ by $d_0(x, y) = |g_0(x) - g_0(y)|$. This distance is estimated by assuming that g_0 belongs to a parametric family. Note an estimator of g_0 has already been proposed by Gardes [10] 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 quantiles 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 the appropriate distance d_0 is proposed and used for the construction of a confidence interval for extreme conditional quantiles. In each section, the methods are illustrated with simulated data. An application to Chicago air pollution data set is proposed in Section 4. Section 5 concludes. All the proofs are postponed to Section 6.

2 Confidence interval construction

2.1 Definition and main result

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be n independent copies of a random vector (X, Y). It is assumed throughout the paper 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 for the conditional quantile

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

where $S(\cdot \mid x_0) = \mathbb{P}(Y > \cdot \mid X = x_0)$. In this paper, we assume that the quantile level $\alpha = \alpha_n$ depends on the sample size n. More specifically,

$$\lim_{n\to\infty} \alpha_n = c \in [0, 1]. \tag{2}$$

Condition (2) with $c \in (0, 1)$ corresponds to a classical conditional quantile level. This is the situation most frequently encountered in the literature. For instance, if $\alpha_n = 1/2$, the value $Q(\alpha_n \mid x_0)$ 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. In the unconditional case, the order statistics method to construct confidence interval 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

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

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

$$\mathrm{CI}_{\gamma,\alpha_n}(k_n, d, x_0) := \left[Y^{(d,x_0)}_{\mathcal{L}_\gamma(k_n,\alpha_n),k_n}, Y^{(d,x_0)}_{\mathcal{R}_\gamma(k_n,\alpha_n),k_n} \right], \tag{3}$$

where we recall that

$$\mathcal{L}_\gamma(k_n, \alpha_n) := \max\left\{ j \in \{1, \ldots, k_n\};\ F_{\mathrm{beta}}(\alpha_n; k_n - j + 1, j) \leq \frac{1-\gamma}{2} \right\}$$

$$\mathcal{R}_\gamma(k_n, \alpha_n) := \min\left\{ j \in \{1, \ldots, k_n\};\ F_{\mathrm{beta}}(\alpha_n; k_n - j + 1, j) \geq \frac{1+\gamma}{2} \right\}.$$

The confidence interval $\mathrm{CI}_{\gamma,\alpha_n}(k_n, d, x_0)$ is defined as in the unconditional case except that only the k_n nearest neighbors random variables $Y_1^{(d,x_0)}, \ldots, Y_{k_n}^{(d,x_0)}$ are used.

It remains to prove that the coverage probability of this interval tends to γ as the sample size increases. The accuracy of the confidence interval $\mathrm{CI}_{\gamma,\alpha_n}(k_n, d, x_0)$ depends on the smoothness of the function $x \to S[Q(\alpha \mid x_0) \mid x]$. For $\alpha \in (0,1)$ and $\zeta > 0$, we introduce the quantity

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

which is the largest deviation of the ratio $S[Q(\alpha \mid x_0) \mid x]/S[Q(\alpha \mid x_0) \mid x_0]$ from 1 when x belongs to the ball of center x_0 and radius ζ. Note that this quantity is classically considered when dealing with conditional distribution, see for instance Daouia et al. [4]. In the following result, the conditions required for the convergence of the coverage probability of (3) to γ are established.

Theorem 1 *Let $\gamma \in (0,1)$ and $x_0 \in \mathcal{X}$. Assume that $k_n \to \infty$ and let h_n 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 $S(\cdot \mid x_0)$ is continuous and strictly decreasing,*

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

and if

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

then, $\mathbb{P}\left[\mathrm{CI}_{\gamma,\alpha_n}(k_n, d, x_0) \ni Q(\alpha_n \mid x_0)\right] = \gamma + \mathcal{O}(\delta_n) + \mathcal{O}(\eta_n) \to \gamma$.

Note that, under the conditions of Theorem 1, other confidence intervals with asymptotic coverage probability γ can be proposed as for instance the one-sided confidence intervals

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

The proof of Theorem 1 is based on the decomposition

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

and on a similar one for $\mathbb{P}[Y_{\mathcal{R}_\gamma(k_n,\alpha_n),k_n}^{(d,x_0)} \leq Q(\alpha_n \mid x_0)]$. This decomposition highlights two terms of error: $B_{1,n}(\mathcal{L}_\gamma(k_n, \alpha_n))$ where for $j \in \{1, \ldots, k_n\}$ $B_{1,n}(j) := \mathbb{P}[Y_{j,k_n}^{(d,x_0)} > Q(\alpha_n \mid x_0)] - F_{\mathrm{beta}}(\alpha_n; k_n - j + 1, j)$ and $B_{2,n} := F_{\mathrm{beta}}(\alpha_n; k_n - \mathcal{L}_\gamma(k_n, \alpha_n) + 1, \mathcal{L}_\gamma(k_n, \alpha_n)) - (1-\gamma)/2$. The first one is due to

the approximation of the distribution of the random variable $S(Y_{j,k_n}^{(d,x_0)} \mid x_0)$ by a Beta distribution. We show in the proof of Theorem 1 that

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

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 $B_{2,n} \to 0$ under condition (4). If $c = 0$, the rate of convergence of α_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 $1 - \alpha_n \gg \ln^2(k_n)/k_n$. Note that condition (5) is more restrictive when $\alpha_n \to 1$ than when $\alpha_n \to c \in [0, 1)$. As shown in the simulation studies, the construction of confidence intervals in the left tail can thus be more difficult than in the right tail. It also appears that, as expected, the rate of convergence of the coverage probability can be very slow for extreme conditional quantiles.
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 $k_n/(\ln \ln n) \to \infty$ and $n/k_n \to \infty$ 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 perform 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 Illustration on simulated data

Let us take a look at the finite sample performance of the confidence interval introduced in the previous section. Using the observations of a sample $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ drawn from a random pair (X, Y), our objective is

to construct a γ-confidence interval for the conditional quantile $Q(\alpha_n \mid x_0)$. In the estimation procedure, the nearest neighbors are selected with the classical Euclidean distance d_e. Two models for the distribution of (X, Y) are considered:

– **Model 1**: The p components of the random vector X are independent and uniformly distributed on $[-5, 5]$. The conditional survival function of Y given X is given for $y > 0$ by

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

where c and τ are positive functions defined for all $x \in \mathbb{R}^p$ by $c(x) = \|x\|_1$ and $\tau(x) = c(x)\xi(g_0(x))$. The function $\xi : \mathbb{R} \to (0, \infty)$ is defined by $\xi(z) := 5z^2/36 + 1/4$. Note that **Model 1** is not well defined when $X = 0$ since in this case $S(y \mid 0) = 1$ for all y.

In this model, the conditional distribution of Y given that $X = x$ is a Burr distribution. Such a distribution is said to be heavy-tailed since for all $t > 0$ and $x \in \mathcal{X}$,

$$\lim_{y \to \infty} \frac{S(ty \mid x)}{S(y \mid x)} = t^{-1/\xi(g_0(x))}.$$

The function $\xi \circ g_0$ is referred to as the conditional extreme value index. It controls the right tail heaviness of the conditional distribution. This model is investigated with different values for the dimension p of the covariate and different functions $g_0 : \mathbb{R}^p \to \mathbb{R}$. More specifically, 4 settings are considered:

(S_1) $p = 1$ with for $x \in \mathbb{R}$, $g_0(x) = x$,

(S_2) $p = 2$ with for $x \in \mathbb{R}^2$, $g_0(x) = (1, 2)^\top x/\sqrt{5}$,

(S_3) $p = 4$ with for $x \in \mathbb{R}^4$, $g_0(x) = (0, 1, 2, 0)^\top x/\sqrt{5}$,

(S_4) $p = 8$ with for $x \in \mathbb{R}^8$, $g_0(x) = (0, 1, 2, 0, 0, 0, 1, 1)^\top x/\sqrt{7}$.

– **Model 2**: The p components of the random vector X are independent and uniformly distributed on $[-5, 5]$. The conditional survival function of Y given X is given for $y > 0$ by

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

where $\xi : \mathbb{R} \to (0, \infty)$ is defined by $\xi(z) := 5z^2/36 + 1/4$.

The conditional distribution of Y given that $X = x$ in **Model 2** is a conditional Weibull type distribution, see for instance Gardes and Girard [12] or Gardes et al. [13] and $\xi(g_0(x))$ is referred to as the conditional Weibull-tail index. As the conditional extreme value index, $\xi(g_0(x))$ controls the tail heaviness of the conditional distribution. For p and g_0, we consider the 4 settings (S_1) to (S_4).

To evaluate the performance of the confidence interval, we compute its coverage probability $\mathbb{P}[\mathrm{CI}_{\gamma,\alpha_n}(k_n, d_e, x_0) \ni Q(\alpha_n \mid x_0)]$. This probability is approximated numerically by a Monte-Carlo procedure. More specifically, $N = 2\ 000$ independent samples of size n were generated. For given values of $k_n \in \{1, \ldots, n\}$ and $\gamma \in (0, 1)$, the confidence interval obtained with the r-th replication is denoted $\mathrm{CI}^{(r)}_{\gamma,\alpha_n}(k_n, d_e, x_0)$. The coverage probability is then approximated by

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

This value is expected to be close to the preselected probability γ.

Selection of the number of nearest neighbors − We first take a look at the influence of the number of nearest neighbors k_n. In Figure 1, for a sample of size $n = 1000$, the values of the coverage probabilities are represented as a function of $k_n \in \{10, \ldots, 200\}$ for **Model 1** with the settings (S_1), (S_2) and (S_3) for g_0 and p. Three different values for the conditional quantile level are considered: $\alpha = 1-8\ln(n)/n \approx 0.9447$, $\alpha = 1/2$ and $\alpha = 1-\alpha_{1,n} \approx 0.0553$. The point of interest x_0 is the vector with all its components equal to 1.
It appears that when the quantile level is close to 0 or 1, 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 conditional quantile level α_n, the point of interest $x_0 \in \mathbb{R}^p$, the nominal coverage probability γ and the distance d used to collect the nearest neighbors. First, let

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

be the random variable corresponding to the center of the confidence interval $\mathrm{CI}_{\gamma,\alpha_n}(k_n, d, x_0)$. The basic idea to select a convenient number of nearest neighbors is to take k is a stability region of the finite sequence $\{C(n_0), \ldots, C(n_1)\}$ where $1 \le n_0 < n_1 \le n$. More precisely, we are searching for the value

$$\tilde{k}^{(sel)}_n := \underset{i\in\{n_0,\ldots,n_1\}}{\arg\min}\ \mathrm{Var}(C(i)).$$

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

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

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

$$\widehat{k}^{(sel)}_n := \min\{i \in \{n_0, \ldots, n_1\};\ \widehat{\mathrm{Var}}_n(C(i)) \le \eta\}, \tag{6}$$

with the convention $\min\{\emptyset\} = n_0$. Note that when $\eta = 0$, $\widehat{k}_n^{(sel)}$ is the argument of the minimum of the sequence $\{\widehat{\mathrm{Var}}_n(n_0), \ldots, \widehat{\mathrm{Var}}_n(n_1)\}$. The role of η is to obtain a value of $\widehat{k}_n^{(sel)}$ less sensitive to the fluctuations of the sequence $\{\widehat{\mathrm{Var}}_n(n_0), \ldots, \widehat{\mathrm{Var}}_n(n_1)\}$. To sum up, the setting parameters required to compute (6) are the integers n_0 and n_1 delimiting the possible values for k_n, the value of a to compute the variance local estimator and the value of η. Throughout the simulation study, these parameters are fixed to $n_0 = \lfloor 0.05n/p \rfloor$, $n_1 = 200$, $a = 0.006$ and η to the first quartile of the sequence $\{\widehat{\mathrm{Var}}_n(n_0), \ldots, \widehat{\mathrm{Var}}_n(n_1)\}$.

In Figure 1, one can check that for the conditional median ($\alpha_n = 1/2$) 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 conditional quantiles of level close to 0 or 1. Note that for the settings (S_1) and (S_2), the coverage probabilities are clearly better in the right tail than in the left one. This fact can partially be explained by the condition (5) in Theorem 1 since the factor $k_n\alpha_n/(1-\alpha_n)$ approaches 0 faster when α_n goes to 0 than when α_n goes to 1. However, for p and g_0 as in (S_3), the coverage probabilities are better in the left tail. Indeed, in this situation, the quantity $\omega(\alpha_n, h_n)$ is very close to 0 when α_n is close to 1, counteracting the bad effect of the factor $k_n\alpha_n/(1-\alpha_n)$.

Influence of the sample size – To illustrate the influence of the sample size on the confidence intervals, we generate samples from **Model 1** with different sample sizes $n \in \{100, 200, \ldots, 2000\}$. The number of nearest neighbors are given by (6). In Figure 2, the values of the coverage probabilities are represented as a function of n. Three conditional quantiles levels are considered: $\alpha_n = 1 - [n^{-3/10}\ln(n)]^3/14$ (left tail), $\alpha_n = 1/2$ (conditional median) and $\alpha_n = [n^{-3/10}\ln(n)]^3/14$ (right tail). The point of interest is the vector of ones. Concerning the choice of p and g_0 in **Model 1**, the settings (S_1) and (S_3) are investigated.

When $p = 1$, the coverage probability for the conditional median ($\alpha_n = 1/2$) is correct for any value of the sample size between 100 and 2000. For a conditional quantile in the right tail, i.e., for a level close to 0, the coverage probability converges to the preselected value γ. This is no longer the case when the level is close to 1. This phenomenon can be explained by the difficulty to choose a correct number k_n of nearest neighbors, see Figure 1 and the corresponding discussion.

When $p = 4$ and $\alpha_n = [n^{-3/10}\ln(n)]^3/14 \to 0$, the coverage probability does not converge to γ. This is not a surprising fact in view of the data sparsity in the right tail of the distribution. For the conditional median, it seems that the coverage probability getting worse when n increases. This can perhaps be explained by a bad choice of k_n, see Figure 1. Finally, a better behavior is observed for a conditional quantile close to 1 and a large value of n.

Influence of the point of interest x_0 − We generate a sample of size $n = 1000$ from **Model 1** and we construct confidence intervals for the conditional quantile $Q(\alpha \mid x_0)$ with $x_0 = (t, \ldots, t)^\top \in \mathbb{R}^p$, $t \in \{-45/9, -35/9, \ldots, 45/9\}$ and $\alpha \in \{1-\beta; 1/2; \beta\}$, $\beta = [n^{-3/10}\ln(n)]^3/14 \approx 0.047$. For p and g_0, we consider the two settings (S_1) and (S_3). In Figure 3, the values of the coverage probabilities are represented as a function of x_0.
It appears that the coverage probability deteriorates when x_0 get closer to the boundary of the support, i.e., when t get closer to -5 or 5. This boundary effect is a classical source of bias for local estimators as for instance the density kernel estimator. The coverage probability is also poor when t is close to 0. Indeed, in this case, x_0 is close to 0 and, as mentioned before, **Model 1** is not defined when the covariate X is equal to 0.

Influence of the covariate dimension − Our goal here is to assess the finite sample performance of the confidence interval for different values of the covariate dimension. The point of interest x_0 is the vector with all its components equal to 1 and the sample size is $n = 1000$. Three different levels for the conditional quantile $Q(\alpha_n \mid x_0)$ are considered: $\alpha_1 = 1 - 8\ln(n)/n \approx 0.945$ (left tail), $\alpha_2 = 1/2$ (conditional median) and $\alpha_3 = 8\ln(n)/n \approx 0.055$ (right tail). The values of the coverage probabilities are gathered in Table 1 for **Model 1** and Table 2 for **Model 2**.
For the conditional median, the coverage probability is quite close to γ and the accuracy of the confidence interval is not affected by the dimension p of the covariate. For a right tail extreme conditional quantile, 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. As discussed before, this is an expected consequence of the data sparsity around x_0 when p increases. Finally, for a left tail extreme conditional quantile, the accuracy mostly depends on the function g_0. As mentioned before, a bad performance in the left tail can be explained by the factor $k_n\alpha_n/(1-\alpha_n)$ in condition (5). However, for some functions g_0, the quantity $\omega(\alpha_n, h_n)$ is very close to 0 leading to good coverage probabilities.

3 Selection of the nearest neighbors for large-dimensional covariates

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. The obtained confidence intervals can then perform very badly in particular for conditional quantiles in the tail of the distribution. This phenomenon has been illustrated in the

previous section. We propose below a data driven procedure to choose a more convenient distance for the selection of the nearest neighbors located in the right tail of the distribution. Our data driven procedure is based on a tail dimension reduction model presented in the next section. The method described below is devoted to the right tail but it can be easily adapted to the left tail.

3.1 Dimension reduction model

In the literature dedicated to dimension reduction, it is commonly assumed that there exists a function $g_0 : \mathbb{R}^p \to \mathbb{R}$ such that $X \perp\!\!\!\perp Y \mid g_0(X)$ or equivalently 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 $g_0(X)$ without loss of information. 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 $\mathbb{R}^p$. Recall that our goal is to select nearest neighbors located in the right tail of the conditional distribution. The classical condition $X \perp\!\!\!\perp Y \mid g_0(X)$ is thus relaxed by assuming that Y is tail conditionally independent of X given $g_0(X)$, see Gardes [9]. 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 $\varphi_y : \mathbb{R} \to \mathbb{R}$ such that, as $y \to \infty$,

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

The notation $\xrightarrow{a.s.u.}$ stands for the almost surely uniform convergence [1], see for instance Lukács [22] or Rambaud [24, Proposition 1]. Roughly speaking under (TCI), inference on extreme conditional quantiles 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 the distance d_0.

Note that if there exist $\phi : \mathbb{R} \to \mathbb{R}$, $\phi \neq Id$ and $\tilde{g}_0 : \mathbb{R}^p \to \mathbb{R}$ such that $g_0 = \phi \circ \tilde{g}_0$ then if g_0 satisfies (TCI) same holds for the function $\tilde{g}_0$. To ensure that g_0 is the only function satisfying (TCI), 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 $g : \mathbb{R}^p \to \mathbb{R} \in \mathcal{G}$, there are no functions $\phi : \mathbb{R} \to \mathbb{R}$ (with $\phi \neq Id$) and $\tilde{g} : \mathbb{R}^p \to \mathbb{R} \in \mathcal{G}$ such that $g = \phi \circ \tilde{g}$.

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

$$\mathcal{G}_L := \left\{ g : \mathbb{R}^p \to \mathbb{R};\ g(x) = b^\top x;\ b \in \Theta_p \right\}, \tag{7}$$

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

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 [10]. One can also consider sets of non-linear functions (see Section 3.3 for an example). The function g_0 satisfying (TCI) is unknown and has to be estimated. This is done in the next section. Note that in Gardes [10], an estimator has been proposed but the procedure is computationally expensive and can be used only for a linear function $g_0 \in \mathcal{G}_L$.

3.2 Estimation of g_0

To explain our estimation procedure, let us first assume that the function φ_y involved in (TCI) is such that for all $y \in \mathbb{R}$

$$\arg\max_{z\in\mathbb{R}} \varphi_y(z) = z^*, \tag{8}$$

where z^* does not depend on y. Since under (TCI), $\mathbb{P}[Y > y \mid X] \approx \varphi_y(g_0(X))$ for y large enough, condition (8) entails that the largest observations of Y are more likely to be observed when $g_0(X)$ is close to z^*. In other word, given that Y is large, the dispersion of $g_0(X)$ around z^* must be small. One way to quantify such a dispersion is to consider a Gini-type dispersion measure given for a large threshold $y \in \mathbb{R}$ by $\mathbb{E}[|g_0(X) - g_0(X^*)| \mid \min(Y, Y^*) > y]$ where (X^*, Y^*) is an independent copy of (X, Y). This measure is estimated by replacing the expectation by its empirical counterpart and by taking for the threshold y the order statistic $Y_{n-\lfloor n\beta_n\rfloor,n}$ where (β_n) is a sequence tending to 0 as the sample size increases. An estimator of g_0 is then obtained by solving

$$\arg\min_{g\in\mathcal{G}} \frac{1}{(\lfloor n\beta_n\rfloor)^2} \sum_{i,j=0}^{\lfloor n\beta_n\rfloor-1} \left|g(X_{(i)}) - g(X_{(j)})\right|, \tag{9}$$

where $X_{(i)}$ is the concomitant of the order statistic $Y_{n-i,n}$. This estimation procedure is only reliable if (8) holds. This quite restrictive condition can be weakened if for each $g \in \mathcal{G}$ we assume the existence of $H \in \mathbb{N} \setminus \{0\}$ non-overlapping intervals $\mathcal{S}_{1,g}, \ldots, \mathcal{S}_{H,g}$ covering the support of $g(X)$ and such that for all $h \in \{1, \ldots, H\}$

$$\arg\max_{z\in\mathcal{S}_{h,g}} \varphi_y(z) = z^*_{h,g}. \tag{10}$$

Hence, as explained above, for each $h \in \{1, \ldots, H\}$, the Gini-type dispersion measure

$$E_{h,g_0}(y) := \mathbb{E}\left[|g_0(X) - g_0(X^*)| \mid \min(Y, Y^*) > y; (g_0(X), g_0(X^*)) \in \mathcal{S}^2_{h,g_0}\right],$$

is expected to be small for a large threshold y. Let

$$n_{h,g} := \sum_{i=0}^{\lfloor n\beta_n\rfloor-1} \mathbb{I}_{\mathcal{S}_{h,g}}(g(X_{(i)})),$$

and, if $n_{h,g} \neq 0$, let

$$\widehat{E}_{h,g}(\beta_n) := \frac{1}{n_{h,g}^2} \sum_{i,j=0}^{\lfloor n\beta_n \rfloor - 1} \left| g(X_{(i)}) - g(X_{(j)}) \right| \mathbb{I}_{\mathcal{S}_{h,g}}(g(X_{(i)})) \mathbb{I}_{\mathcal{S}_{h,g}}(g(X_{(j)})),$$

be the estimator of the Gini-type measure $E_{h,g}(u)$ obtained by replacing the expectation by its empirical counterpart and y by the order statistic $Y_{n-\lfloor n\beta_n \rfloor, n}$. An estimator of g_0 can be defined as the solution of

$$\arg\min_{g \in \mathcal{G}} \sum_{h \in J_{H,g}} \widehat{E}_{h,g}(\beta_n), \tag{11}$$

where $J_{H,g} := \{h \in \{1, \ldots, H\};\ n_{h,g} > 0\}$ is the set of indices $h \in \{1, \ldots, H\}$ for which the $\mathbb{R}^2$ subset $\mathcal{S}_{h,g} \times [Y_{n-\lfloor n\beta_n \rfloor, n}, \infty)$ is non-empty. The obtained estimator of g_0 can be significantly improved by adding to (11) a penalty term in order to minimize the number $\mathrm{card}(J_{H,g})$ of non-empty sets $\mathcal{S}_{h,g} \times [Y_{n-\lfloor n\beta_n \rfloor, n}, \infty)$. Our final estimator of g_0 is thus defined by

$$\widehat{g}_{n,0} := \arg\min_{g \in \mathcal{G}} \left\{ \sum_{h \in J_{H,g}} \widehat{E}_{h,g}(\beta_n) + \lambda \mathrm{card}(J_{H,g}) \right\}, \tag{12}$$

for some penalty coefficient $\lambda > 0$. In practice, $\widehat{g}_{n,0}$ is computed by taking the non-overlapping intervals $\mathcal{S}_{h,g} := [\widehat{\xi}_{h-1,g}, \widehat{\xi}_{h,g}]$ where $\widehat{\xi}_{h,g}$ is the $(h/H)-$sample quantile of $\{g(X_i),\ i = 1, \ldots, n\}$. When H is large enough, one can reasonably assume that (10) is satisfied. The setting parameters of our procedure of estimation are the sequence β_n, the number H of intervals and finally the penalty coefficient λ. In the simulation study below, these parameters are set to $\beta_n = 5/(3\sqrt{n})$, $\lambda = 1$ and $H = 20$. A more theoretical justification of (12) is provided in Appendix B.

3.3 Illustration on simulated data

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ 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 goal of this section is to assess the finite sample performance of the confidence interval for the conditional quantiles of Y given $X = x_0$ when the dimension p is large. Assuming that condition (TCI) holds for some function g_0, we propose the following two step procedure for the construction of the confidence interval.

i) Estimate the function g_0 by $\widehat{g}_{n,0} = \widehat{g}_0$ given in (12);

ii) Select the nearest neighbors with the estimated distance defined for $(x, y) \in \mathbb{R}^{2p}$ by $\widehat{d}_0(x, y) = |\widehat{g}_0(x) - \widehat{g}_0(y)|$ and construct the confidence interval (3).

We recall that the estimator of g_0 is only adapted to the right tail of the conditional distribution. We thus focus on extreme conditional quantiles with α close to 0. Throughout this simulation study, the point of interest x_0 is the vector of ones and the 3 following models are considered for the distribution of the random vector (X, Y):

– **Model 1**: as defined in Section 2.2.

For the value of the dimension p and for the function g_0, 4 different settings are investigated: (S_2), (S_3), (S_4) and

(S_5) $p = 4$ with for $x \in \mathbb{R}^4$, $g_0(x) = |(1, 0, 0, 1)^\top x^2|^{1/2}/16$, where the components of x^2 are the square of the components of x.

It can be shown that **Model 1** satisfies condition (TCI) with $\varphi_y(z) = y^{-1/\xi(z)}$.

– **Model 2**: as defined in Section 2.2,

with the same choices for p and for the function g_0 as before. Of course, condition (TCI) also holds for **Model 2** with $\varphi_y(z) = \exp(-y^{-1/\xi(z)})$.

– **Model 3**: The p components of the random vector X are independent and distributed as a normal random variable with mean $1/2$ and standard deviation $1/3$. The conditional quantile of Y given X is, for $\alpha \in (0, 1)$

$$Q(\alpha \mid X) = \left[\ln\left(\frac{1}{1-\alpha}\right)\right]^{-\xi_1(g_0(x))} \left[1 + \xi_2(g_1(x)) \exp\left(-\frac{1}{\alpha}\right)\right]^{-1},$$

with for $z \in \mathbb{R}$,

$$\xi_1(z) := \frac{1}{2} + \frac{3}{2} \min\left\{1; \left(\frac{\exp(2z) - 1}{\exp(8/3) - 1}\right)_+\right\}$$

and $\xi_2(z) := \exp\left(5 \min\{2; (z)_+\}\right)$, $(\cdot)_+$ being the notation for the positive part function.

For the dimension p and the functions g_0 and g_1 involved in this model, the following choices are considered:

(S_2^*) setting (S_2) for p and g_0 and $g_1(x) = (1, 0)^\top x/2$,

(S_3^*) setting (S_3) for p and g_0 and $g_1(x) = (1, 0, 0, 1)^\top x/2$,

(S_4^*) setting (S_4) for p and g_0 and $g_1(x) = (1, 0, 0, 1, 1, 1, 0, 0)^\top x/2$,

(S_5^*) setting (S_5) for p and g_0 and $g_1(x) = (0, 1, 0, 1)^\top x/2$.

For all $x \in \mathcal{X}$ and $t > 0$,

$$\lim_{\alpha \to 0} \frac{Q(t\alpha \mid x)}{Q(\alpha \mid x)} = t^{-\xi_1(g_0(x))},$$

and thus the conditional distribution in **Model 3** is heavy tailed with conditional extreme value index $\xi_1(g_0(x))$. Again, one can show that for this model, condition (TCI) is satisfied with $\varphi_y(z) = y^{-1/\xi_1(z)}$.

i) Estimation of g_0 − Let us first take a look at the finite sample performance of the estimator of the function g_0 defined in (12). The optimization problem is solved by using a coordinate search method, see Hooke and Jeeves [18] and Appendix C for more details. For the settings (S_1) to (S_4), the function g_0 is linear i.e., $g_0(x) = b_0^\top x$ for $x \in \mathbb{R}^p$. In this case, the minimization (12) is achieved over the set $\mathcal{G} = \mathcal{G}_L$ given in (7). For the setting (S_5), $g_0(x) = |b_0^\top x^2|^{1/2}$ and the minimization is achieved over the set

$$\mathcal{G}_{NL} := \left\{ g : \mathbb{R}^p \to \mathbb{R}^p; g(x) = |b^\top x^2|^{1/2};\ b \in \Theta_p \right\},$$

with $\Theta_p := \{b \in \mathbb{R}^p \text{ with } b^\top b = 1 \text{ and } b^\top u_p > 0\}$. In all cases, the function g_0 only depends on a vector $b_0 \in \mathbb{R}^p$ and the minimization (12) is *in fine* achieved over the set Θ_p. We denote by $\widehat{b}_0$ the obtained estimator of b_0. The distance d_0 is then estimated by

$$\widehat{d}_0(x, y) = |\widehat{g}_0(x) - \widehat{g}_0(y)|,$$

where, for the settings (S_1) to (S_4), $\widehat{g}_0(x) = \widehat{b}_0^\top x$ and, for the setting (S_5), $\widehat{g}_0(x) = |\widehat{b}_0^\top x^2|^{1/2}$.
Our estimator of b_0 is compared to the one obtained with another dimension reduction approach: the Slice Inverse Regression (SIR) method, see Li [21]. The assumption behind SIR is the existence of a direction $b_{SIR} \in \mathbb{R}^p$ such that the projection $b_{SIR}^\top X$ captures all the information on Y. In other word, the conditional distribution of Y given X is supposed to be the same as the conditional distribution of Y given $b_{SIR}^\top X$. The estimator $\widehat{b}_{SIR}$ of b_{SIR} is the eigen vector associated to the largest eigen value of the matrix $\widehat{\Sigma}^{-1}\widehat{\Gamma}$ where $\widehat{\Sigma}$ is the sample covariance matrix of X and $\widehat{\Gamma}$ is the sample version of $\mathrm{Cov}(\mathbb{E}(Y \mid X))$. The SIR method is implemented in R, see `https://cran.rproject.org/package=dr`. Roughly speaking, $\widehat{b}_{SIR}^\top X$ is the linear combination providing the best available information on Y. A natural idea is then to select the nearest neighbors with the data driven pseudo-distance

$$\widehat{d}_{SIR}(x, y) = \left| \widehat{b}_{SIR}^\top (x - y) \right|.$$

To measure the performance of $\widehat{b}_0$ and $\widehat{b}_{SIR}$ as estimators of b_0, we use the criterion

$$\delta(\widehat{b}, b_0) := (\widehat{b} - b_0)^\top (\widehat{b} - b_0),$$

where $\widehat{b}$ is either $\widehat{b}_0$ or $\widehat{b}_{SIR}$. We replicate $N = 2\ 000$ times the original sample of size $n = 1000$ in order to compute the empirical mean and standard deviation of this criterion. The results are gathered in Tables 3 to 5. For the 3 models, the estimator $\widehat{b}_0$ obtained by our approach is more accurate than the SIR estimator $\widehat{b}_{SIR}$. For **Model 1** and **Model 3**, the linear combination $b_0^\top X$ captures the information on the tail distribution of Y but not on the whole distribution. This can explains the difficulty for SIR to estimate the vector b_0. For **Model 2**, the conditional distribution of Y given X is the same as the

one of Y given $b_0^\top X$. Despite of this, the performance of SIR is very poor. This is due to the fact that in this case, $\mathbb{E}(Y \mid X)$ is a symmetric function of $b_0^\top X$ and it is well known, see e.g., Cook and Weisberg [3], that SIR fails to recover the true direction b_0 in this case. Finally, SIR is clearly not adapted to setting (S_5) since g_0 is a non linear function in this case.

ii) Behavior of the confidence interval – Our goal is to assess the finite sample performance of the confidence interval for $Q(\alpha_n \mid x_0)$ defined in (3) when the estimated distance $\widehat{d}_0$ is used to select the nearest neighbors. First, we are interested in the influence of the number k_n of nearest neighbors on the coverage probability. For a sample of size $n = 1000$ generated from **Model 1** with settings (S_2) to (S_5), the coverage probabilities are represented as a function of $k_n \in \{10, \ldots, 200\}$ in Figure 4. The k_n nearest neighbors are selected with 3 different distances: our data driven distance $\widehat{d}_0$, the ideal but unknown distance d_0 and the Euclidean distance. The conditional quantile level is fixed to $\alpha = 8\ln(n)/n \approx 0.055$. It appears that the choice of k_n is really less crucial when one use the distances $\widehat{d}_0$ or d_0. We can also check again that the selection of k_n by the procedure described in Section 2.2 provides confidence intervals with a coverage probability close to γ.
Let us now look at the influence of the sample size n. We generate samples from **Model 1** with $n \in \{100, \ldots, 2000\}$ and under setting (S_3). The coverage probabilities for the conditional quantile $Q(\alpha_n \mid x_0)$ with $\alpha_n = [n^{-3/10}\ln(n)]^3/14$ are represented on Figure 5 as a function of n. The estimated distance $\widehat{d}_0$ and the Euclidean distance are considered for the selection of the nearest neighbors. As expected, when the distance $\widehat{d}_0$ is used, the coverage probability converges to γ as the sample size increases. This is not the case for the Euclidean distance.
Finally, in Tables 6 to 8, we compare the coverage probabilities obtained under the 3 models and the 4 different settings for p and g_0. The value of the sample size is fixed to $n = 1000$ and the conditional quantile level to $\alpha = 8\ln(n)/n \approx 0.055$. The nearest neighbors are selected with 4 distances: $\widehat{d}_0$, $\widehat{d}_{SIR}$, d_0 and the Euclidean distance d_e. For settings (S_2) to (S_4), replacing the Euclidean distance by the estimated distance $\widehat{d}_0$ leads to a significant improvement in the coverage probability. Note that for setting (S_5), the estimation of the non linear function g_0 is more challenging, especially in **Model 1**, but the obtained coverage probability remains better than the one obtained with the Euclidean distance. Of course the best results are obtained for the unknown distance d_0 but they are generally close to the ones obtained with the estimated distance $\widehat{d}_0$. Finally, the coverage probabilities obtained by using the distance $\widehat{d}_{SIR}$ are far from the preselected probability γ except for **Model 3**, setting (S_2). This was expected in view of the results presented in paragraph i).

4 Chicago air pollution data set

The Chicago air pollution data set, 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 data set has been studied by several authors in a dimension reduction context (e.g., Scrucca [26] and Xia [29]) and, in an extreme value context, by Gardes [10]. 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 centered and normalized daily maximum concentrations of PM_{10}, SO_2, NO_2 and CO. As in Gardes [10], we assume that condition (TCI) holds with $g_0(x) = b_0^\top x$ for $x \in \mathbb{R}^4$.

The first step is the estimation of the vector $b_0 \in \mathbb{R}^4$. Two estimators are considered: $\widehat{b}_0$ as defined in (12) and the SIR estimator $\widehat{b}_{SIR}$. The obtained estimated vector are:

$$\begin{aligned} \widehat{b}_0 &= (0.198, -0.155, 0.963, 0.093)^\top \\ \text{and } \widehat{b}_{SIR} &= (0.327, -0.085, 0.910, -0.238)^\top . \end{aligned}$$

These two vectors are quite different but both of them show that the covariate NO_2 bring the most important information on large values of ozone concentration. This point has also been noted by Scrucca [26] or Gardes [10].

We construct the confidence interval for $Q(\alpha \mid x_0)$ given in (3). For the selection of the nearest neighbors, two distances are investigated: $\widehat{d}_0(x, y) = |\widehat{b}_0^\top(x - y)|$ and $\widehat{d}_{SIR}(x, y) = |\widehat{b}_{SIR}^\top(x - y)|$. About the point of interest x_0, two situations are investigated. For $\tau \in (0, 1)$, let $x_0^{\mathrm{PM}_{10}}(\tau)$, $x_0^{\mathrm{SO}_2}(\tau)$, $x_0^{\mathrm{NO}_2}(\tau)$ and $x_0^{\mathrm{CO}}(\tau)$ be the sample quantiles of order $1 - \tau$ of the values of PM_{10}, SO_2, NO_2 and CO.
Situation 1 – $x_0 = (x_0^{\mathrm{PM}_{10}}(0.5), x_0^{\mathrm{SO}_2}(0.5), x_0^{\mathrm{NO}_2}(0.5), x_0^{\mathrm{CO}}(0.5))^\top$. 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 – $x_0 = (x_0^{\mathrm{PM}_{10}}(0.5), x_0^{\mathrm{SO}_2}(0.25), x_0^{\mathrm{NO}_2}(0.05), x_0^{\mathrm{CO}}(0.05))^\top$ corresponding to large values for NO_2 and CO.
The quantile level α is taken between $8\ln(n)/n \approx 0.014$ and $64\ln(n)/n \approx 0.112$. The number of nearest neighbors is chosen by the data driven procedure (6). For instance for $\alpha = 8\ln(n)/n$ and under situation 2, the number of nearest neighbors is 242. The value 0.014 for the quantile level is thus close to 0 in that sense that $242 \times 0.014 \approx 3.39$. Keep in mind that condition (4) in Theorem 1 entails that $k_n\alpha_n \to \infty$. The confidence

intervals with preselected probability $\gamma = 0.9$ are represented on Figure 6 as a function of α for the 2 values of x_0 and by using the two distances $\widehat{d}_0$ and $\widehat{d}_{SIR}$. It appears that for α close to 0, the confidence intervals obtained with $\widehat{d}_0$ and $\widehat{d}_{SIR}$ are different. This difference is more important for situation 2 corresponding to large values for NO_2 and CO. When the distance $\widehat{d}_0$ is used, the length of the confidence interval increases when the quantile level α decreases. This is an expected behavior of confidence interval for extreme conditional quantiles. This is no longer the case when the distance $\widehat{d}_{SIR}$ is considered. Since our method is dedicated to right tail of the distribution and in view of the results presented in the simulation study, the use of $\widehat{d}_0$ is preferable. Note also that, as pointed out in Gardes [10] or Han et al. [17], very important ozone concentrations are more likely to be observed for large concentrations of NO_2 and CO.

Finally, the confidence intervals obtained with the distance $\widehat{d}_0$ are represented on Figure 7 as a function of $\widehat{b}_0^\top X$ along with the concentrations Y of ozone. Two quantile levels are consider: $\alpha_n = 0.02$ and $\alpha_n = 0.05$. Be aware that what is represented in Figure 7 are the point wise confidence intervals and not the confidence bands for the function $x \mapsto Q(\alpha_n \mid x)$, see the discussion in the next section.

5 Concluding remarks

As illustrated in the simulation study presented in Section 2.2, the construction of confidence intervals for extreme conditional quantiles with large-dimensional covariates is a difficult task. The main contribution of this paper is to propose a method to construct confidence intervals in such situations. First, based on the condition (TCI) introduced in Gardes [10], we reduce the dimension of the covariate. This dimension reduction method is dedicated to the right tail of the conditional distribution. Second, a nearest neighbors version of the order statistics approach is used to obtain the confidence intervals. The nearest neighbors are selected with a distance based on the reduced covariate rather than the classical Euclidean distance. The results obtained on simulated data show that the dimension reduction step improve substantially the performance of the confidence intervals when the quantile level is close to 0. This work can be continued in at least two directions.

1) Condition (4) in Theorem 1 entails that $k_n\alpha_n(1-\alpha_n) \to \infty$. As a consequence, α_n cannot tend to 0 or 1 too fast and in particular, the conditional quantile must be located inside the range of the k_n nearest neighbors. In this situation, the endpoints of the confidence interval are order statistics that can be seen as nonparametric estimators of the conditional quantiles $Q(\alpha_{n,L} \mid x_0)$ and $Q(\alpha_{n,R} \mid x_0)$ with $\alpha_{n,L} = 1-\mathcal{L}_\gamma(k_n, \alpha_n)/k_n$ and $\alpha_{n,R} = 1-\mathcal{R}_\gamma(k_n, \alpha_n)/k_n$. For an extreme conditional quantile $Q(\alpha_n \mid x_0)$ located outside the data range,

i.e., when $k_n\alpha_n(1-\alpha_n) \to c \in [0,\infty)$, the endpoints of the confidence interval can no longer be order statistics. In such a case, a possible solution to construct confidence intervals is to assume that the conditional distribution of Y given $X = x_0$ belongs to a given maximum domain of attraction. The endpoints of the confidence intervals can then be obtained by extrapolating the conditional quantiles $Q(\alpha_{n,L} \mid x_0)$ and $Q(\alpha_{n,R} \mid x_0)$ outside the data range. Extrapolated estimators can be found for instance in Daouia et al. [4]. The main difficulty is to establish the convergence of the coverage probability to γ; this is a work in progress.
2) In this paper, we focus on point wise confidence intervals since x_0 is fixed. It would also be interesting to obtain confidence bands for extreme conditional quantiles. Here the problem is to find a collection $\{(A_{n,\gamma}(x), B_{n,\gamma}(x)),\ x \in \mathcal{X}\}$ of random variables such that

$$\lim_{n\to\infty} \mathbb{P}\left\{[A_{n,\gamma}(x), B_{n,\gamma}(x)] \ni Q(\alpha_n \mid x),\ \text{for all } x \in \mathcal{X}\right\} = \gamma,$$

or equivalently

$$\lim_{n\to\infty} \mathbb{P}\left\{\max_{x\in\mathcal{X}} (Q(\alpha_n \mid x) - B_{n,\gamma}(x)) < 0; \min_{x\in\mathcal{X}} (A_{n,\gamma}(x) - Q(\alpha_n \mid x)) > 0\right\} = \gamma.$$

Proving such a convergence result is a difficult mathematical problem. As a departure points, one can try to adapt some elements of the proof of Theorem 1 in Gardes and Stupfler [15] where a uniform consistency result is proven in an extreme value framework.

6 Proofs

6.1 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 Γ is the gamma function.

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

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

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

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

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 $f_{\text{beta}}\left(\alpha_{n,\tau}; \lfloor m_n\alpha_{n,\tau}\rfloor + 1, m_n - \lfloor m_n\alpha_{n,\tau}\rfloor\right)$, we use the Stirling's bounds given for all $r \in \mathbb{N}\setminus\{0\}$ by $\sqrt{2\pi}r^{r+1/2}\mathrm{e}^{-r} \leq r! \leq r^{r+1/2}\mathrm{e}^{1-r}$ (see for instance Feller [8, Paragraph 2.9]).

First, taking $r = m_n$ leads to

$$\sqrt{2\pi}m_n^{1/2+m_n}\mathrm{e}^{-m_n} \leq m_n! \leq m_n^{1/2+m_n}\mathrm{e}^{1-m_n}. \tag{13}$$

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

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

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,

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

Note that the first inequality is due to the fact that

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

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

$$\sqrt{\frac{\pi}{2}}\frac{(m_n\alpha_{n,\tau})^{1/2+\lfloor m_n\alpha_{n,\tau}\rfloor}}{\mathrm{e}^{\lfloor m_n\alpha_{n,\tau}\rfloor+1}} \leq (\lfloor m_n\alpha_{n,\tau}\rfloor)! \leq \frac{(m_n\alpha_{n,\tau})^{1/2+\lfloor m_n\alpha_{n,\tau}\rfloor}}{\mathrm{e}^{\lfloor m_n\alpha_{n,\tau}\rfloor-1}}. \quad (14)$$

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

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

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

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

we obtain the inequality

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

leading to

$$\begin{aligned}&\sqrt{\frac{\pi}{2}}\frac{(m_n(1-\alpha_{n,\tau}))^{m_n-\lfloor m_n\alpha_{n,\tau}\rfloor-1/2}}{\mathrm{e}^{m_n-\lfloor m_n\alpha_{n,\tau}\rfloor-1/2}} \leq (m_n-\lfloor m_n\alpha_{n,\tau}\rfloor-1)!\\ \leq\ &\frac{(m_n(1-\alpha_{n,\tau}))^{m_n-\lfloor m_n\alpha_{n,\tau}\rfloor-1/2}}{\mathrm{e}^{m_n-\lfloor m_n\alpha_{n,\tau}\rfloor-2}}. \quad (15)\end{aligned}$$

Gathering (13), (14) and (15) yields to

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

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

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

$$\delta_n^2:=\frac{\ln^2(m_n)}{m_n\alpha_n(1-\alpha_n)}\to 0, \tag{16}$$

as $n\to\infty$. 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,

$$F_{\mathrm{beta}}(\alpha_n;m_n-\mathcal{L}_\gamma(m_n,\alpha_n)+1,\mathcal{L}_\gamma(m_n,\alpha_n))=\frac{1-\gamma}{2}+\mathcal{O}\left(\delta_n\right) \tag{17}$$

and $$F_{\mathrm{beta}}(\alpha_n;m_n-\mathcal{R}_\gamma(m_n,\alpha_n)+1,\mathcal{R}_\gamma(m_n,\alpha_n))=1-\frac{1-\gamma}{2}+\mathcal{O}\left(\delta_n\right). \tag{18}$$

Proof – Remark that since $m_n\to\infty$, condition (16) entails that $m_n(\alpha_n\wedge(1-\alpha_n))\to\infty$. Hence, since the function $j\to F_{\mathrm{beta}}(\alpha;m_n-j+1,j)$ is increasing for all $\alpha\in(0,1)$

$$\max_{j=1,\ldots,m_n}F_{\mathrm{beta}}(\alpha_n;n-j+1,j)=F_{\mathrm{beta}}(\alpha_n;1,m_n)=1-(1-\alpha_n)^{m_n}\to 1,$$

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

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

Hence, for n large enough,

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

and

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

This conclude the first part of the proof.
We now prove (17). The proof of (18) is similar and is thus omitted. The definition of $\mathcal{L}_\gamma(m_n,\alpha_n)$ ensures that

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

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

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

Hence to prove (17) it suffices to show that

$$\max_{j=1,\ldots,m_n} D_n(j) = \mathcal{O}\left(\frac{\ln(m_n)}{[m_n\alpha_n(1-\alpha_n)]^{1/2}}\right). \tag{19}$$

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

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

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

$$D_n(m_n) = o\left(\frac{\ln(m_n)}{[m_n\alpha_n(1-\alpha_n)]^{1/2}}\right). \tag{20}$$

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

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

Let

$$\begin{aligned} A_n &:= \left\{\max_{j=2,\ldots,m_n}(U_{j,m_n} - U_{j-1,m_n}) \leq 2\frac{\ln(m_n)}{m_n}\right\} \\ \text{and} \quad \overline{A}_n &:= \left\{\max_{j=2,\ldots,m_n}(U_{j,m_n} - U_{j-1,m_n}) > 2\frac{\ln(m_n)}{m_n}\right\}. \end{aligned}$$

It is easy to check that

$$\begin{aligned} D_n(m_n - j + 1) &\leq \mathbb{P}\left[\left\{U_{j,m_n} \leq \alpha_n + \max_{j=2,\ldots,m_n}(U_{j,m_n} - U_{j-1,m_n})\right\} \cap A_n\right] \\ &- \mathbb{P}[U_{j,m_n} \leq \alpha_n] + \mathbb{P}(\overline{A}_n) \\ &\leq D_n^{(1)}(m_n - j + 1) + \mathbb{P}(\overline{A}_n), \end{aligned} \tag{21}$$

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 $j = 2, \ldots, m_n$, 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 (16), the second part of Lemma 1 entails that

$$\max_{j=2,\ldots,m_n} D_n^{(1)}(m_n - j + 1) = \mathcal{O}\left(\frac{\ln(m_n)}{[m_n\alpha_n(1-\alpha_n)]^{1/2}}\right). \tag{22}$$

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

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

Let $T_{m_n+1} := (E_1 + \ldots + E_{m_n+1})/(m_n+1)$. 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 $n \geq N_\eta$

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

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

$$\mathbb{P}(\overline{A}_n) = \mathcal{O}\left(\frac{\ln(m_n)}{[m_n\alpha_n(1-\alpha_n)]^{1/2}}\right). \tag{23}$$

By gathering (20), (21), (22) and (23) we get (19) and the proof is complete. ■

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

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

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

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

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

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

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

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

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

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

and the proof is complete. ■

6.2 Proofs of main results

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

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

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

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

one has that

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

Remarking that from Lemma 3, ii)

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

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

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

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

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

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

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

uniformly on $j=1,\ldots,k_n$. 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 Dony and Einmahl [6, Corollary 2.1] (see also Gardes et al. [14, Lemma 2]), since $nh_n^p/[\ln\ln n]\to\infty$, one as $N_n/(nh_n^p)\xrightarrow{a.s.}f_X(x_0)$. 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. ■

References

[1] Basu, D. and Pereira, C.A.B. (1983). Conditional independence in statistics, *Sankhyā*, **45(3)**, Series A, 324–337.

[2] Cook, R.D. and Li, B. (2002). Dimension reduction for conditional mean in regression, *The Annals of Statistics*, **30**, 455–474.

[3] Cook, R.D. and Weisberg, S. (1991). Discussion of "Sliced Inverse Regression for dimension reduction", *Journal of the American Statistical Association*, **86**, 328–332.

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

[5] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, Third edition, Wiley Series in Probability and Statistics.

[6] 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.

[7] Fan, Y. and Liu, R. (2016). A direct approach to inference in nonparametric and semiparametric quantile models, *Journal of Econometrics*, **191(1)**, 196–216.

[8] Feller, W. (1968). *An introduction to probability theory and its applications*, Volume 1, third edition, New York, Wiley.

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

[10] Gardes, L. (2018). Tail Dimension Reduction for extreme quantile estimation, *Extremes*, **21(1)**, 57–95.

[11] 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.

[12] Gardes, L. and Girard, S. (2016). On the estimation of the functional Weibull tail-coefficient, *Journal of Multivariate Analysis*, **146**, 29–45.

[13] 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.

[14] 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.

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

[16] 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.

[17] Han, S., Bian, H., Feng, Y., Liu, A., Li, ., Zeng, F. and Zhang, X. (2011). Analysis of the relationship between O_3, NO and NO_2 in Tianjin, China, *Aerosol and Air Quality Research*, **11**, 128–139.

[18] Hooke, R. and Jeeves, T.A. (1961). Direct search solution of numerical and statistical problems, *Journal of the ACM*, **8(2)**, 212–229.

[19] Hutson, A.D. (1999). Calculating nonparametric confidence intervals for quantiles using fractional order statistics, *Journal of Applied Statistics*, **26(3)**, 343–353.

[20] Kocherginsky, M., He, X. and Mu, Y. (2005). Practical confidence intervals for regression quantiles, *Journal of Computational and Graphical Statistics*, **14(1)**, 41–55.

[21] Li, K. C. (1991). Sliced Inverse Regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316–342.

[22] Lukács, E. (1975). *Stochastic convergence*, 2nd edition, Academic Press, New York.

[23] Parzen, M.I., Wei, L.J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions, *Biometrika*, **81(2)**, 341–350.

[24] 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(3)**, 215–221.

[25] Samarov, A.M. (1993). Exploring regression structure using nonparametric functional estimation, *Journal of the American Statistical Association*, **88**, 836–847.

[26] Scrucca, L. (2011). Model-based SIR for dimension reduction, *Computational Statistics and Data Analysis*, **55**, 3010–3026.

[27] Stute, W. (1984). Asymptotic normality of nearest neighbor regression function estimates, *The Annals of Statistics*, **12**, 917–926.

[28] Thompson, W.R. (1936). On confidence ranges for the median and other expectation distributions for populations of unknown distribution form, *Annals of Mathematical Statistics*, **7**, 122–128.

[29] Xia, Y. (2009). Model checking in regression via dimension reduction, *Biometrika*, **96(1)**, 133–148.

[30] 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.

[31] 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: Values of the coverage probabilities for **Model 1** with different settings.

p	γ	Settings	$\alpha_1 \approx 0.945$	$\alpha_2 = 1/2$	$\alpha_3 \approx 0.055$
1	0.9	(S_1)	0.3130	0.8770	0.9125
	0.95		0.3300	0.9320	0.9410
2	0.9	(S_2)	0.0900	0.8700	0.8040
	0.95		0.1265	0.9385	0.8530
4	0.9	(S_3)	0.8675	0.7865	0.6265
	0.95		0.9110	0.8670	0.6975
8	0.9	(S_4)	0.7045	0.9115	0.5215
	0.95		0.7330	0.9550	0.6095

Table 2: Values of the coverage probabilities for **Model 2** with different settings.

p	γ	Settings	$\alpha_1 \approx 0.945$	$\alpha_2 = 1/2$	$\alpha_3 \approx 0.055$
1	0.9	(S_1)	0.9300	0.9270	0.9350
	0.95		0.9525	0.9640	0.9630
2	0.9	(S_2)	0.7140	0.9280	0.8310
	0.95		0.7985	0.9630	0.8930
4	0.9	(S_3)	0.2885	0.9245	0.6430
	0.95		0.3565	0.9640	0.7200
8	0.9	(S_4)	0.3070	0.8805	0.6230
	0.95		0.3785	0.9400	0.7190

Table 3: For **Model 1**, values of the empirical mean and empirical standard deviation (into brackets) of the criterion $\delta(\widehat{b}, b_0)$ for $\widehat{b} = \widehat{b}_0$ and $\widehat{b} = \widehat{b}_{SIR}$.

p	Settings	$\delta(\widehat{b}_0, b_0)$	$\delta(\widehat{b}_{SIR}, b_0)$
2	(S_2)	0.0163 (0.0233)	0.8239 (0.7690)
4	(S_3)	0.0265 (0.0235)	1.4173 (0.8272)
4	(S_5)	0.1047 (0.0905)	1.3996 (0.7996)
8	(S_4)	0.0595 (0.0412)	1.9960 (0.7113)

Table 4: For **Model 2**, values of the empirical mean and empirical standard deviation (into brackets) of the criterion $\delta(\widehat{b}, b_0)$ for $\widehat{b} = \widehat{b}_0$ and $\widehat{b} = \widehat{b}_{SIR}$.

p	Settings	$\delta(\widehat{b}_0, b_0)$	$\delta(\widehat{b}_{SIR}, b_0)$
2	(S_2)	0.0113 (0.0161)	0.8491 (0.7659)
4	(S_3)	0.0213 (0.0189)	1.4474 (0.8163)
4	(S_5)	0.0869 (0.0793)	1.3904 (0.8045)
8	(S_4)	0.0763 (0.1918)	1.7531 (0.0959)

Table 5: For **Model 3**, values of the empirical mean and empirical standard deviation (into brackets) of the criterion $\delta(\widehat{b}, b_0)$ for $\widehat{b} = \widehat{b}_0$ and $\widehat{b} = \widehat{b}_{SIR}$.

p	Settings	$\delta(\widehat{b}_0, b_0)$	$\delta(\widehat{b}_{SIR}, b_0)$
2	(S_2)	0.1293 (0.2071)	0.3404 (0.1012)
4	(S_3)	0.269 (0.3925)	1.4690 (0.1597)
4	(S_5)	0.4637 (0.5193)	1.6310 (0.1455)
8	(S_4)	0.4501 (0.3996)	1.7002 (0.1203)

Table 6: For **Model 1**, values of the coverage probabilities for $n = 1000$ and $\alpha = 8\ln(n)/n \approx 0.055$.

p	Settings	γ	$\widehat{d}_0$	$\widehat{d}_{SIR}$	d_0	d_e
2	(S_2)	0.9	0.8805	0.2675	0.9255	0.8040
		0.95	0.9210	0.2980	0.9550	0.8530
4	(S_3)	0.9	0.8665	0.0545	0.9235	0.6225
		0.95	0.9060	0.0680	0.9510	0.6975
4	(S_5)	0.9	0.7965	0.0000	0.8910	0.2470
		0.95	0.8370	0.0000	0.9065	0.2990
8	(S_4)	0.9	0.8895	0.1385	0.9345	0.5215
		0.95	0.9260	0.1755	0.9615	0.6095

Table 7: For **Model 2**, values of the coverage probabilities for $n = 1000$ and $\alpha = 8\ln(n)/n \approx 0.055$.

p	Settings	γ	$\widehat{d}_0$	$\widehat{d}_{SIR}$	d_0	d_e
2	(S_2)	0.9	0.9205	0.4160	0.9275	0.8310
		0.95	0.9445	0.4660	0.9530	0.8930
4	(S_3)	0.9	0.8910	0.2190	0.9280	0.6430
		0.95	0.9310	0.2810	0.9515	0.7200
4	(S_5)	0.9	0.8550	0.0330	0.8820	0.4940
		0.95	0.8840	0.0515	0.9030	0.5775
8	(S_4)	0.9	0.9025	0.2655	0.9385	0.6230
		0.95	0.9270	0.2950	0.9615	0.7190

Table 8: For **Model 3**, values of the coverage probabilities for $n = 1000$ and $\alpha = 8\ln(n)/n \approx 0.055$.

p	Settings	γ	$\widehat{d}_0$	$\widehat{d}_{SIR}$	d_0	d_e
2	(S_2^*)	0.9	0.8400	0.8135	0.9129	0.7960
		0.95	0.8735	0.8520	0.9395	0.8095
4	(S_3^*)	0.9	0.7510	0.4480	0.9030	0.5265
		0.95	0.7890	0.4880	0.9340	0.5430
4	(S_5^*)	0.9	0.8110	0.5000	0.9305	0.6235
		0.95	0.8460	0.5645	0.9545	0.6375
8	(S_4^*)	0.9	0.8840	0.7050	0.9330	0.6550
		0.95	0.9055	0.7485	0.9505	0.6555

Figure 1: Simulated data. For **Model 1**, coverage probabilities as a function of k_n with $\alpha \approx 0.9447$ (left panels), $\alpha = 1/2$ (middle panels) and $\alpha \approx 0.0553$ (right panels). Choice of p and g_0: top panels: setting (S_1), center panels: setting (S_2), bottom panels: setting (S_3). The horizontal full line is the nominal probability $\gamma = 0.9$ and the dashed horizontal line represents the coverage probability obtained with the selected value of k_n.

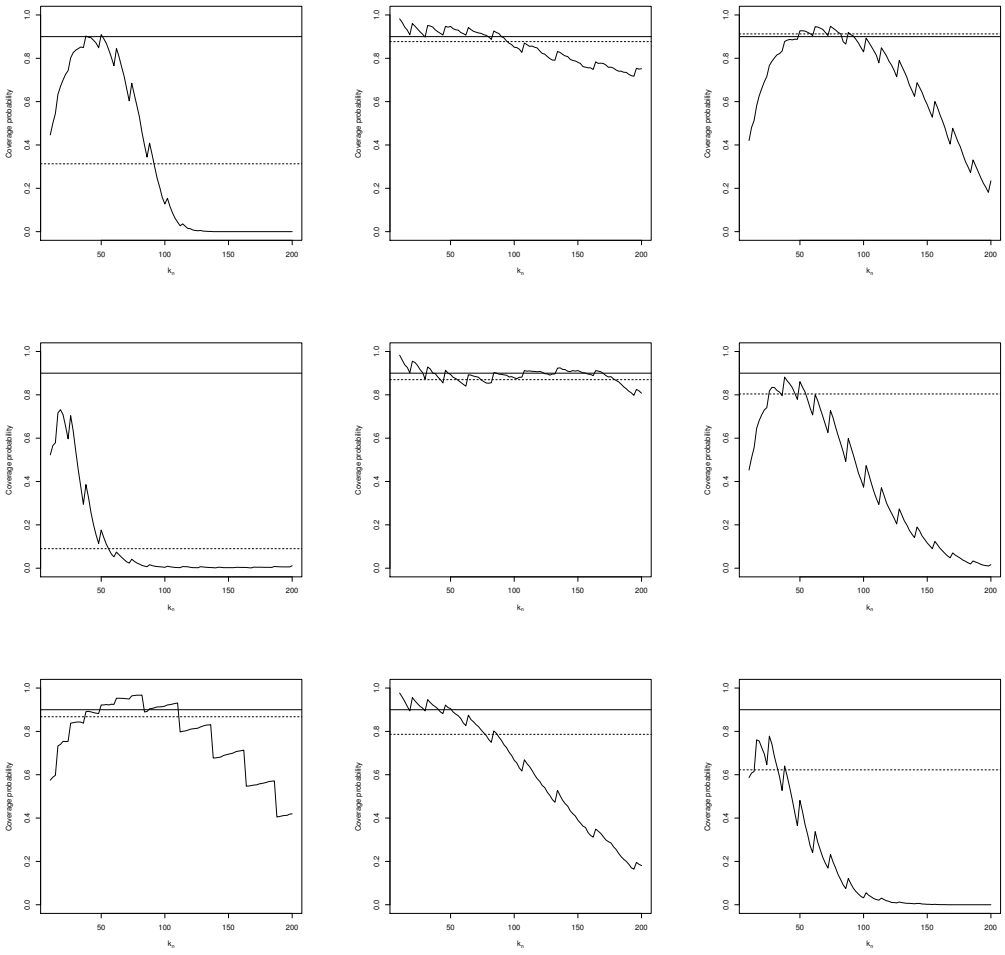


Figure 2: Simulated data. For **Model 1**, coverage probabilities as a function of the sample size n with $\alpha_n = 1 - [n^{-3/10}\ln(n)]^3/14$ (dotted line), $\alpha_n = 1/2$ (dashed line) and $\alpha_n = [n^{-3/10}\ln(n)]^3/14$ (full line). Left panel: setting (S_1), right panel: setting (S_3). The horizontal full line is the nominal probability $\gamma = 0.9$.

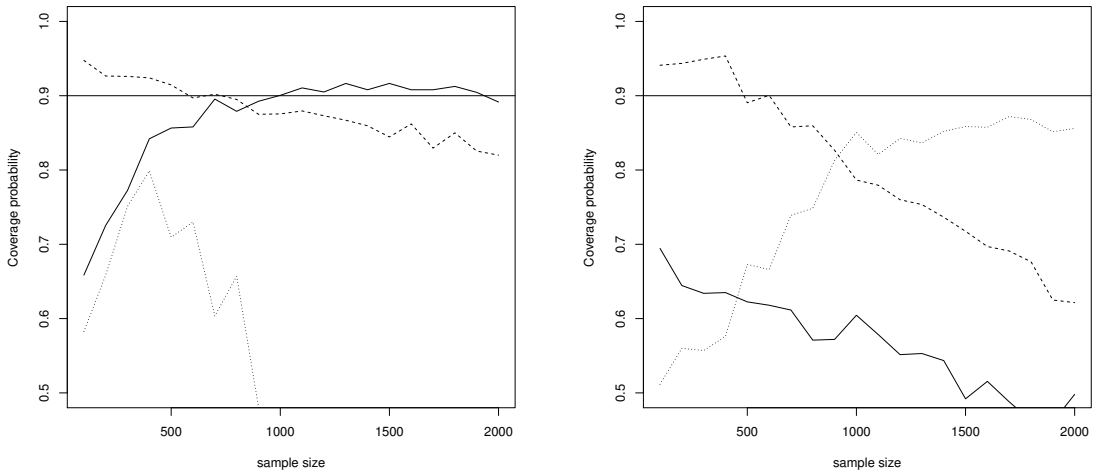


Figure 3: Simulated data. For **Model 1**, coverage probabilities as a function of x_0 with $\alpha_n = 1 - [n^{-3/10}\ln(n)]^3/14$ (dotted line), $\alpha_n = 1/2$ (dashed line) and $\alpha_n = [n^{-3/10}\ln(n)]^3/14$ (full line). The sample size is $n = 1000$. Left panel: setting (S_1), right panel: setting (S_3). The horizontal full line is the nominal probability $\gamma = 0.9$.

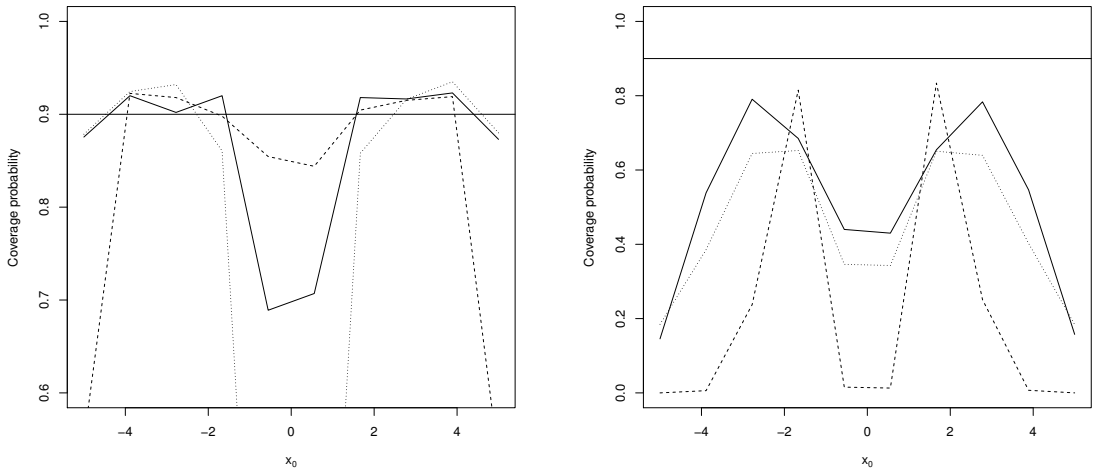


Figure 4: Simulated data. For **Model 1**, values of the coverage probabilities as a function of k_n with $n = 1000$ and $\alpha_n = 8\ln(n)/n \approx 0.055$. The distances used for the nearest neighbors selection are the Euclidean distance (Full line), the estimated distance $\widehat{d_0}$ (dashed line) and d_0 (dotted line). Top panels: settings (S_2) and (S_3), bottom panels: settings (S_4) and (S_5). The horizontal full line is the nominal probability $\gamma = 0.9$ and the dashed horizontal line represents the coverage probability obtained with the selected value of k_n.

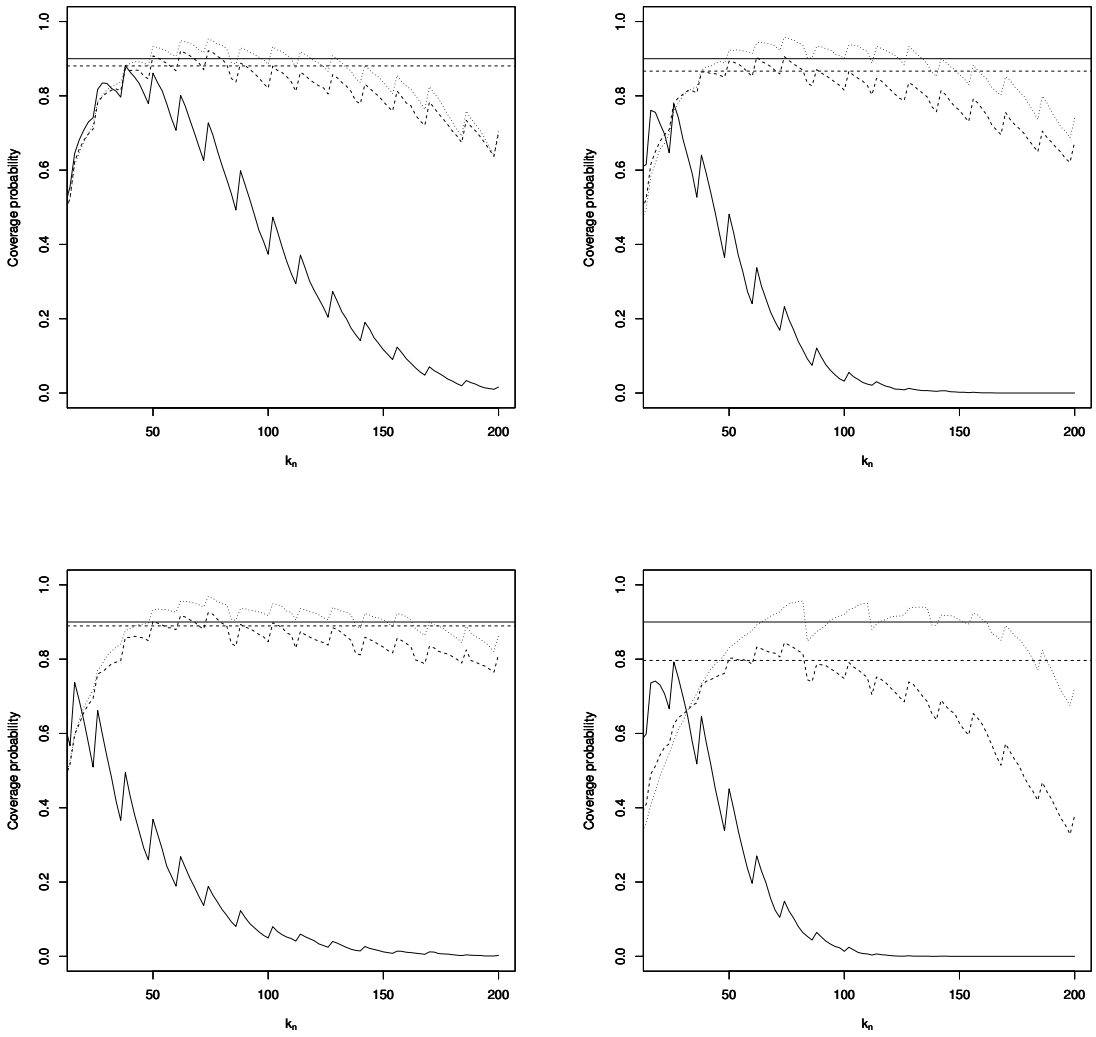


Figure 5: Simulated data. For **Model 1** with setting (S_3), values of the coverage probabilities as a function of the sample size n with $\alpha_n = [n^{-3/10}\ln(n)]^3/14$. The nearest neighbors are selected with the estimated distance $\widehat{d}_0$ (full line) and the Euclidean distance (dashed line). The horizontal full line is the nominal probability $\gamma = 0.9$.

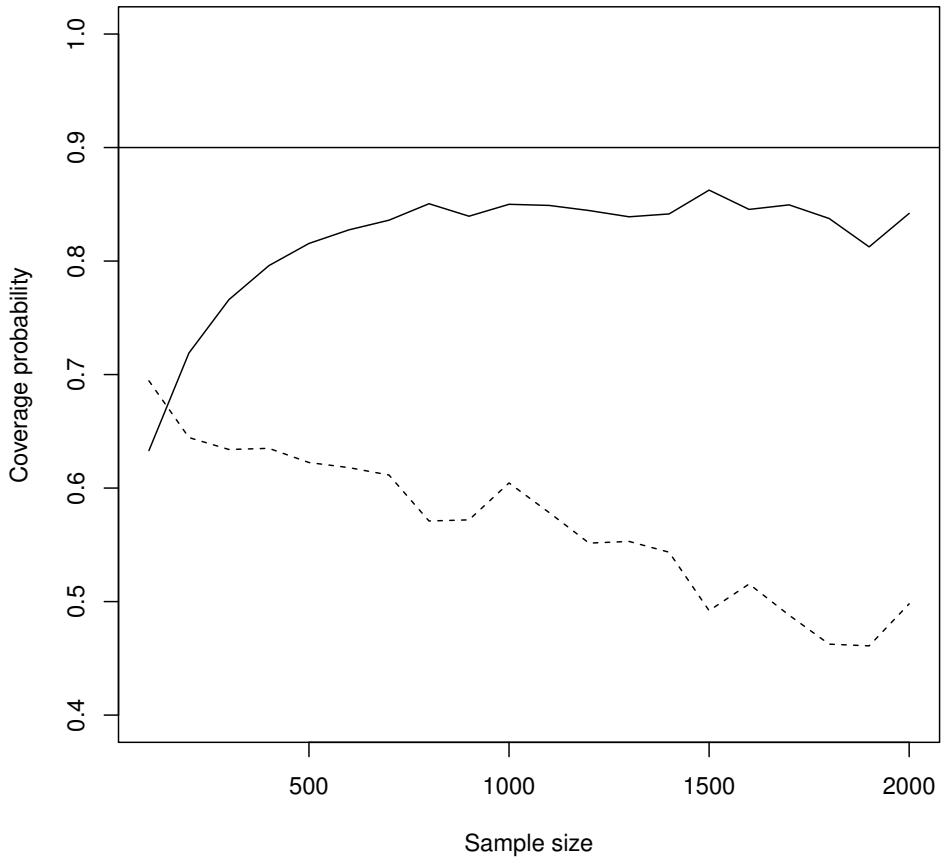


Figure 6: Real data. Confidence intervals for $Q(\alpha \mid x_0)$ with nominal probability $\gamma = 0.9$ as a function of $\alpha \in [0.014, 0.112]$. The nearest neighbors are selected with $\widehat{d}_0$ (full line) and $\widehat{d}_{SIR}$ (dashed line). The left panel corresponds to situation 1 and the right panel to situation 2.

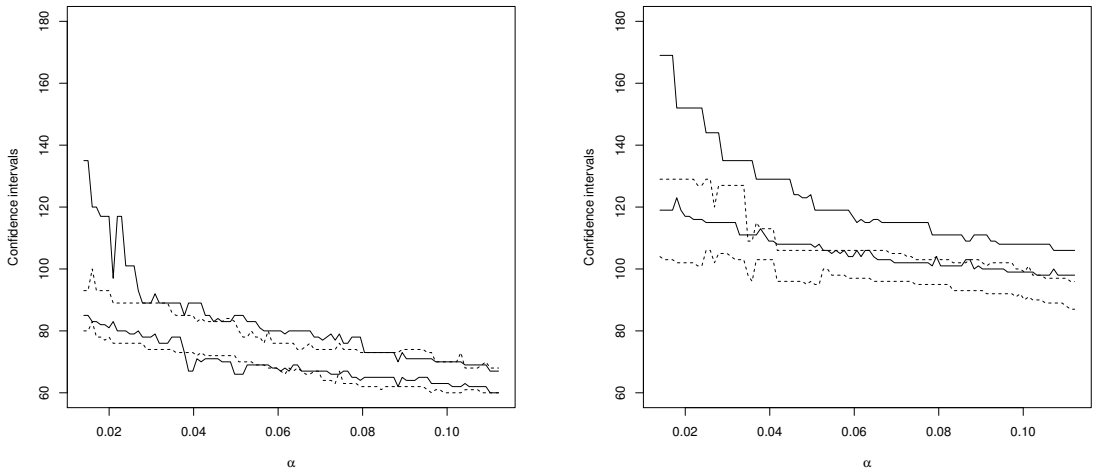
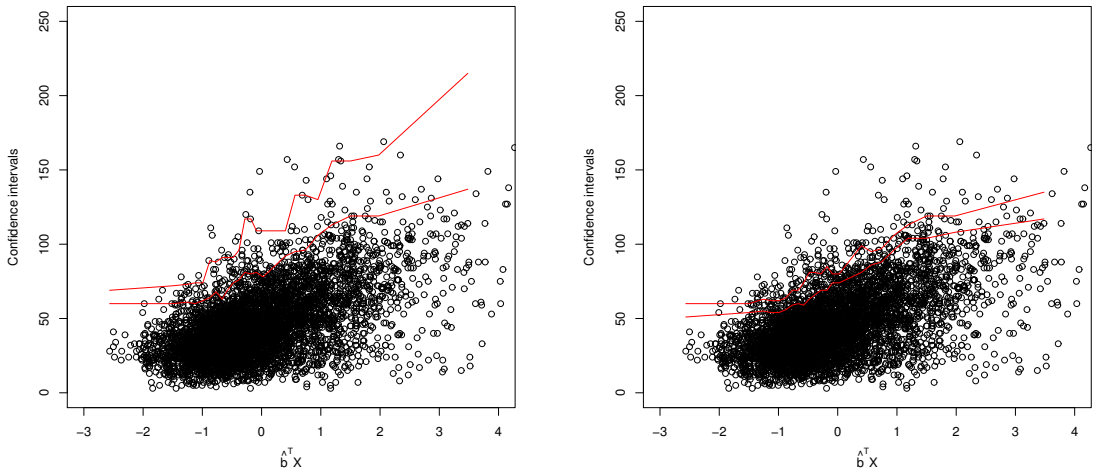


Figure 7: Real data. Confidence intervals for $Q(\alpha \mid X)$ (in red) with nominal probability $\gamma = 0.9$ as a function of $\widehat{b}_0^\top X$. The nearest neighbors are selected with $\widehat{d}_0$. The points are the values of Y versus $\widehat{b}_0^\top X$. The left panel corresponds to $\alpha = 0.02$ and the right panel to $\alpha = 0.05$.



Appendix B: Theoretical justification of $\widehat{g}_{n,0}$

B1 – Main results

In this section, we give a theoretical justification for the expression of the estimator of g_0 defined in (12). The proofs of the results given in Section B1 are postponed to Section B2. Let us first introduce the following additional condition on the function φ_y involved in (TCI). In what follows, we denote by $\mathcal{X}_0$ the support of $g_0(X)$ where g_0 satisfies condition (TCI).

(H1) The function φ_y is continuous. Furthermore, there exist $J \in \mathbb{N} \setminus \{0\}$, a collection $I_1, \ldots, I_J$ of non-overlapping intervals covering $\mathcal{X}_0$ and $y_0 \in \mathbb{R}$, such that for all $y \geq y_0$ and $j \in \{1, \ldots, J\}$, the function φ_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 φ_y admits a finite number of local maximum points. Assuming that (H1) holds, the following condition on the non-overlapping intervals $\mathcal{S}_{1,g}, \ldots, \mathcal{S}_{H,g}$ is required.

(H2) There exists $H_0 \in \mathbb{N} \setminus \{0\}$ such that for all $H \geq H_0$ and $j \in \{1, \ldots, J\}$,

$$z_j^* \in \bigcup_{h=1}^{H} \mathring{\mathcal{S}}_{h,g_0}$$

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

Hence, under (H1) and (H2) and for a sufficiently large number of intervals $\mathcal{S}_{h,g_0}$, the J local maximum points of φ_y belong to the interior of an interval $\mathcal{S}_{h,g_0}$. 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 $p_y(\cdot \mid \mathcal{C})$ be the survival function of $S(y \mid X)$ given $X \in \mathcal{C}$ ($p_y(t \mid \mathcal{C}) := \mathbb{P}[S(y \mid X) > t \mid X \in \mathcal{C}]$). The associated quantile function is denoted by $q_y(\cdot \mid \mathcal{C}) := \inf\{t; p_y(t \mid \mathcal{C}) \leq \cdot\}$.

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

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

Condition (H3) entails that the observations of $g_0(X)$ given that $Y > y$ and $g_0(X) \in \mathcal{C}$ are located, for y large enough, on a small probability. More specifically, we have the following result.

Lemma 4 *Let g_0 be a function satisfying condition (TCI). For a given interval $I_0 \subset \mathbb{R}$ such that $\mathbb{P}(X \in g_0^{-1}(I_0)) > 0$, assume that condition (H3) holds for $\mathcal{C}_0 := g_0^{-1}(I_0)$ then, for all $\varepsilon \in (0,1)$,*

$$\begin{aligned} &\mathbb{P}[g_0(X) \in B_{y,\varepsilon} \mid X \in \mathcal{C}_0] \leq \varepsilon \\ \textit{and} \quad &\lim_{y\to\infty} \mathbb{P}\left[g_0(X) \in B_{y,\varepsilon} \mid \{X \in \mathcal{C}_0\} \cap \{Y > y\}\right] = 1, \end{aligned}$$

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

Condition (H3) is satisfied for instance by conditional heavy-tailed distributions defined for all $x \in \mathcal{X}$ by $S(y \mid x) := y^{-1/\gamma(x)}\mathcal{L}(y \mid x)$, where γ is a positive function and for all $x \in \mathcal{X}$, $\mathcal{L}(\cdot \mid x)$ is a slowly varying function. This is the object of the following result.

Lemma 5 *Let us consider the random vector (X, Y) such that for $y > 0$ and $x \in \mathcal{X} \subset \mathbb{R}^p$, $S(y \mid x) = y^{-1/\gamma(x)}\mathcal{L}(y, x)$, where γ is a positive function defined on $\mathcal{X}$ and for all $x \in \mathcal{X}$, $\mathcal{L}(\cdot \mid x)$ 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*

$$\lim_{y\to\infty} \sup_{x\in\mathcal{C}} \frac{\ln \mathcal{L}(y, x)}{\ln y} = 0 \tag{24}$$

then condition (H3) holds.

We are now in position to provide a theoretical justification of (12). As in Section 3.2, we denote by $E_{h,g_0}(y)$ the Gini-type measure

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

where (X^*, Y^*) is an independent copy of (X, Y).

Proposition 2 *Assume that there exists a function g_0 satisfying condition (TCI) 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 all $\upsilon > 0$, one has under (H1) and (H2) that*

$$\lim_{y\to\infty} \sum_{h\in\mathcal{J}_{H,y}(\upsilon)} E_{h,g_0}(y) = 0 \text{ and } \lim_{y\to\infty} \text{card}(\mathcal{J}_{H,y}(\upsilon)) \leq J, \tag{25}$$

where $\mathcal{J}_{H,y}(\upsilon) := \{h \in \{1,\ldots,H\};\ \mathbb{P}(g_0(X) \in \mathcal{S}_{h,g_0} \mid Y > y) > \upsilon\}$.

The result given by (25) appears clearly as a theoretical justification of our estimation procedure.

B2 − Proofs of the results

Proof of Lemma 4 − To not overload the equations, we write in the rest of the proof $p_y(\cdot \mid \mathcal{C}_0) =: p_{y,0}(\cdot)$. 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 $\mathbb{P}(g_0(X) \in B_{y,\varepsilon} \mid g_0(X) \in I_0) \leq \varepsilon$. We start with

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

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

$$1 - \delta \leq \frac{S(y \mid X)}{\varphi_y(g_0(X))} \leq 1 + \delta, \tag{26}$$

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

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

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

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

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 $\varepsilon \geq p_{y,0}\left((1+\delta)q_{y,0}(\varepsilon/2)\right)$. In the same way, $\varepsilon/4 \leq p_{y,0}\left((1+\delta)q_{y,0}(\varepsilon/2)\right)$. As a consequence, for $y \geq \max(y_0, y_1)$,

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

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

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

Using (26) and the fact that $\varphi_y(g_0(x)) \leq q_{y,0}(\varepsilon/2)$ for $g(x) \in B^{\mathrm{C}}_{y,\varepsilon}$, one has for $y \geq \max(y_0, y_1)$ that

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

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

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

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

Proof of Lemma 5 – For all y and $t \in (0,1)$, let us introduce the set $A_y(t) := \{x \in \mathcal{X}; S(y \mid x) > t\}$. 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 (24) entails that for all $\delta > 0$, there exists y_δ such that for all $y > y_\delta$, $A^-_y(t) \subset A_y(t) \subset A^+_y(t)$ with

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

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

$$G\left[\left(-\frac{\ln t}{\ln y} - \delta \right)^{-1} \right] \leq p_y(t \mid \mathcal{C}) \leq G\left[\left(-\frac{\ln t}{\ln y} + \delta \right)^{-1} \right]. \tag{27}$$

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 $p_y(y^{-1/G^{\leftarrow}(d\eta)-\delta} \mid \mathcal{C}) \geq d\eta$. Applying the function $q_y(\cdot \mid \mathcal{C})$ (the inverse of $p_y(\cdot \mid \mathcal{C})$) conducts us to the inequality $y^{-1/\overline{G}^{\leftarrow}(d\eta)-\delta} \leq q_y(d\eta \mid \mathcal{C})$. Similarly, using the second inequality in (27), one has for $\eta \in (0,1)$ that $y^{-1/\overline{G}^{\leftarrow}(\eta)+\delta} \geq q_y(\eta \mid \mathcal{C})$. Gathering these inequalities yields

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

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. ■

Proof of Proposition 2 – Let $j \in \{1, \ldots, J\}$ where J is defined in condition (H1) and for $y \in \mathbb{R}$ let

$$E_j(y) := \mathbb{E}\left[|g_0(X) - g_0(X^*)| \mid \min(Y, Y^*) > y; (g_0(X), g_0(X^*)) \in I_j^2\right],$$

The first step of the proof consists in showing that

$$\lim_{y\to\infty} E_j(y) = 0. \tag{28}$$

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 copy of (X, Y). For all $\varepsilon > 0$, let $B_{y,\varepsilon} = \{z \in I_j;\ \varphi_y(z) \geq q_y(\varepsilon/2 \mid \mathcal{C}_j)\}$ where $\mathcal{C}_j := g_0^{-1}(I_j)$. Finally, let $\mathcal{B}_{j,\in} := \{g_0(X) \in B_{y,\varepsilon}\}$. Before proving (28), let us give some results on the previous defined sets. From Lemma 4,

$$\mathbb{P}[\mathcal{B}_{j,\in} \mid \mathcal{B}_j] \leq \varepsilon, \tag{29}$$

and

$$\lim_{y\to\infty} \mathbb{P}[\mathcal{B}_{j,\in} \mid \mathcal{B}_j \cap \mathcal{A}_y] = 1. \tag{30}$$

Since on I_j, φ_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 (29) 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,

$$\mathbb{P}\left[|g_0(X) - z_j^*| \leq \frac{\varepsilon\mathbb{P}(\mathcal{B}_j)}{m}\middle| \mathcal{B}_{j,\in}\right] = 1. \tag{31}$$

We are now in position to prove (28). For $y \in \mathbb{R}$,

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

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

$$E_j(y) = \frac{2}{[\mathbb{P}(\mathcal{A}_y \cap \mathcal{B}_j)]^2}\left[T_{1,y} - T_{2,y}\right], \tag{32}$$

where

$$\begin{aligned} T_{1,y} &:= \mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)\mathbb{E}[g_0(X)\mathbb{I}_{\mathcal{B}_j\cap\mathcal{A}_y} \\ \text{and } T_{2,y} &:= \mathbb{E}[\min(g_0(X), g_0(X^*))\mathbb{I}_{\mathcal{B}_j\cap\mathcal{B}_j^*\cap\mathcal{A}_y\cap\mathcal{A}_y^*}]. \end{aligned}$$

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

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

From (30), for all $\varepsilon > 0$, there exists $y_{1,\varepsilon} \in \mathbb{R}$ such that for all $y > y_{1,\varepsilon}$, $1 - \varepsilon \leq \mathbb{P}[\mathcal{B}_{j,\in} \mid \mathcal{B}_j \cap \mathcal{A}_y] \leq 1 + \varepsilon$. Furthermore, since $B_{y,\varepsilon}$ is a closed interval, there exists $c_j > 0$ such that $|z_j^*| \leq c_j$ and thus, $z_j^* \mathbb{P}(\mathcal{B}_{j,\in} \mid \mathcal{B}_j \cap \mathcal{A}_y) \leq z_j^* + \varepsilon c_j$. Hence,

$$\mathbb{E}(g_0(X)\mathbb{I}_{\mathcal{B}_{j,\in} \cap \mathcal{A}_y}) \leq \mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y) \left(z_j^* + \varepsilon c_j + \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \right). \tag{33}$$

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

$$\begin{aligned}
\mathbb{E}(g_0(X)\mathbb{I}_{\mathcal{B}_{j,\notin} \cap \mathcal{A}_y}) &\leq c_j \mathbb{P}[\mathcal{B}_{j,\notin} \cap \mathcal{A}_y] = c_j \mathbb{P}[\mathcal{B}_j \cap \mathcal{A}_y] \mathbb{P}[\mathcal{B}_{j,\notin} \mid \mathcal{B}_j \cap \mathcal{A}_y] \\
&\leq c_j \varepsilon \mathbb{P}[\mathcal{B}_j \cap \mathcal{A}_y] \qquad (34)
\end{aligned}$$

Gathering (33) and (34) yield to

$$T_{1,y} \leq [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2 \left[z_j^* + \varepsilon \left(2c_j + \frac{\mathbb{P}(\mathcal{B}_j)}{m} \right) \right] \tag{35}$$

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

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

Using the same arguments than those leading to (33), we obtain

$$T_{2,y}^{(1)} \geq [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2 \left(z_j^* - \varepsilon c_j - \frac{\varepsilon \mathbb{P}(\mathcal{B}_j)}{m} \right). \tag{36}$$

Now, using (29), one has for $y > y_{1,\varepsilon}$,

$$\begin{aligned} T_{2,y}^{(2)} &\geq -c_j \mathbb{P}(\mathcal{B}_{j,\in} \cap \mathcal{A}_y)\mathbb{P}(\mathcal{B}_{j,\notin}^* \cap \mathcal{A}_y^*) \\ &= -c_j [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2 \mathbb{P}(\mathcal{B}_{j,\in} \mid \mathcal{B}_j \cap \mathcal{A}_y)\mathbb{P}(\mathcal{B}_{j,\notin} \mid \mathcal{B}_j \cap \mathcal{A}_y) \\ &\geq c_j \varepsilon [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2. \end{aligned} \tag{37}$$

Finally, from (31), one has for $y > y_{1,\varepsilon}$

$$T_{2,y}^{(3)} \geq -c_j [\mathbb{P}(\mathcal{B}_{j,\in} \cap \mathcal{A}_y)]^2 \geq -c_j \varepsilon^2 [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2. \tag{38}$$

Collecting (36), (37) and (38) yield to

$$T_{2,y} \geq [\mathbb{P}(\mathcal{B}_j \cap \mathcal{A}_y)]^2 \left(z_j^* - \varepsilon \frac{\mathbb{P}(\mathcal{B}_j)}{m} + \varepsilon c_j + c_j \varepsilon^2 \right). \tag{39}$$

for all $\varepsilon > 0$ and $y > y_{1,\varepsilon}$. Gathering (32), (35) and (39) conduct to

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

proving (28) since ε can be chosen arbitrarily small.
Let us show now that $\text{card}(\mathcal{J}_{H,y}(\upsilon)) \leq J$ as $y \to \infty$. Conditions (H1) and (H2) entail that there exist $h_1^*, \ldots, h_J^*$ such that for all $j \in \{1, \ldots, J\}$, $z_j^* \in \mathcal{S}_{h_j^*, g_0} \subset I_j$. Furthermore, taking

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

conditions (29) and (30) entail that for all $\upsilon > 0$, there exists $y_2 \in \mathbb{R}$ such that for all $y \geq y_2$ and $h \notin \{h_1^*, \ldots, h_J^*\}$,

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

showing that $\text{card}(\mathcal{J}_{H,y}(\upsilon)) \leq J$.
Finally, mimicking the proof of (28), it is easy to check that $E_{h,g_0}(y) \to 0$ as $y \to \infty$ for all $h \in \{h_1^*, \ldots, h_J^*\} \supset \mathcal{J}_{H,y}(\upsilon)$. ■

Appendix C: Coordinate search method

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

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

where $\Phi : \mathbb{R}^p \to \mathbb{R}$ 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 $i \in \{1, \ldots, 2p\}$, we denote by D_i the ith 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
$$\Phi(x_k^*) \leq \min_{i=1,\ldots,2p} \Phi\left(x_k^* + \alpha_k D_i\right)$$
then $x_{k+1}^* = x_k^*$ and $\alpha_{k+1} = \zeta\alpha_k$. Go to Step $k+1$.
- if
$$\Phi(x_k^*) > \min_{i=1,\ldots,2p} \Phi\left(x_k^* + \alpha_k D_i\right)$$
then $\alpha_{k+1} = \zeta^{-1}\alpha_k$ and
$$x_{k+1}^* = \arg\min\left\{\Phi(x);\ x \in \{x_k^* + \alpha_k D_1, \ldots, x_k^* + \alpha_k D_{2p}\}\right\}.$$
Go to Step $k+1$.

This algorithm is used to solve (12). 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 Θ_p. Denoting by g_b the function belonging to $\mathcal{G}$ with parameter $b \in \Theta_p$, the solution of (12) is obtained by applying the coordinate search method to the function Φ defined for all $\tilde{b} \in \mathbb{R}^p$ by
$$\Phi(\tilde{b}) := \sum_{J_{H,g_b}} \widehat{E}_{h,g_b}(\beta_n) + \lambda \text{card}(J_{H,g_b}).$$

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