



# Tail dimension reduction for extreme quantile estimation

Laurent Gardes

## ► To cite this version:

| Laurent Gardes. Tail dimension reduction for extreme quantile estimation. 2016. hal-01322374v1

**HAL Id: hal-01322374**

**<https://hal.science/hal-01322374v1>**

Preprint submitted on 27 May 2016 (v1), last revised 21 Jun 2017 (v2)

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

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

# Tail dimension reduction for extreme quantile estimation

Laurent Gardes

Université de Strasbourg & CNRS, IRMA, UMR 7501, 7 rue René Descartes,  
67084 Strasbourg cedex, France.  
`gardes@unistra.fr`

## Abstract

In a regression context where a response variable  $Y \in \mathbb{R}$  is recorded with a covariate  $X \in \mathbb{R}^p$ , two situations can occur simultaneously in some applications: (a) we are interested in the tail of the conditional distribution and not on the central part of the distribution and (b) the number  $p$  of regressors is large. Up to our knowledge, these two situations have only been considered separately in the literature. The aim of this paper is to propose a new dimension reduction approach adapted to the tail of the distribution and to introduce an extreme conditional quantile estimator. A simulation experiment and an illustration on a real data set were presented.

**Keywords** – Regression, extreme quantile, dimension reduction, kernel smoothing.

**AMS Subject classifications** – 62G32, 62G08, 62G05, 62G20.

## 1 Introduction

This work takes place in a regression context where a real response variable  $Y$  is recorded with a random vector  $X \in E \subset \mathbb{R}^p$  of explanatory variables. In the literature, several ways for examining how the regressor  $X$  influence the distribution of  $Y$  have been considered. The most common approach consists to summarize the relationship between  $Y$  and  $X$  by the regression function  $\mathbb{E}(Y|X)$  which is the conditional expectation of  $Y$  given  $X$ . Several estimators of the regression function are available, the probably most known being the kernel estimator introduced independently by Nadaraya [26] and Watson [30]. In the same spirit, one can also mention the estimator introduced by Gasser and Müller [18]. Another way to understand the link between  $Y$  and  $X$  is to use a conditional quantile of fixed order  $1 - \alpha \in (0, 1)$ . For instance, Koenker and Basset [23] introduced the notion of quantile regression assuming that the conditional quantile of  $Y$  given  $X$  is a linear combination of the explanatory variables. This approach presents the advantage to be more robust against outliers than the regression function. Concerning the estimation of conditional quantile, local linear approaches were considered by Yu and Jones [34] while a fully nonparametric estimator can be found in the paper of Chaudhuri [3].

In many applications such as climatology, finance, insurance to name a few, two situations can occur simultaneously in a regression context.

(a) We are interested in the tail distribution of  $Y$  given  $X$  instead of the central part of the conditional distribution. In this case, regression function and conditional quantile of fixed order  $1 - \alpha$  are clearly irrelevant tools.

(b) The dimension  $p$  of the regressor is large. In this situation, inference on the conditional distribution of  $Y$  given  $X$  becomes difficult since the space is sparsely populated by data points. This is the well known curse of dimensionality problem.

A motivating example is the study of the influence of various pollutants (sulphur dioxide, nitrogen dioxide, carbon monoxide, ...) and weather conditions (temperature, humidity, ...) on extreme values of ozone concentration (see Section 6). Another example can be found in hydrology where the understanding of the influence of the geographical position and the altitude on return periods of large amount of rain is a problem of primary interest (see Gardes and Girard [15]). Despite this large range of applications, the situations (a) and (b) mentioned before have been considered separately in the literature.

To make inference on the tail distribution of  $Y$  given  $X$ , one solution is to use a conditional quantile of order  $1 - \alpha_n$  where  $\alpha_n \rightarrow 0$  as the sample size goes to infinity. Such a quantile is said to be extreme. Estimation of extreme conditional quantiles has been considered for instance by Daouia *et al.* ([7] and [8]), Gardes and Girard [16], Davison and Ramesh [9] among many others.

To deal with high dimensional covariates, one classical way consists in assuming the existence of a  $p \times q$  full rank matrix  $B$  (with  $q < p$ ) such that in the study of the relationship between  $Y$  and  $X$ , the random vector  $X \in \mathbb{R}^p$  can be replaced by  $B^\top X \in \mathbb{R}^q$  without loss of information. More precisely, it is assumed that  $X$  and  $Y$  are independent conditionally to  $B^\top X$  (in symbols  $X \perp\!\!\!\perp Y | B^\top X$ ). Recall that, for a given random vector  $Z$ ,  $X \perp\!\!\!\perp Y | Z$  if the distribution of  $Y$  given  $(X, Z)$  is equal to the one of  $Y$  given  $Z$  (or equivalently if for all real-valued measurable functions  $h_1(\cdot)$  and  $h_2(\cdot)$  such that  $h_1(X)$  and  $h_2(Y)$  are integrable,  $\mathbb{E}(h_1(X)h_2(Y)|Z) = \mathbb{E}(h_1(X)|Z)\mathbb{E}(h_2(Y)|Z)$ ). For a comprehensive discussion on conditional independence see Basu and Pereira [2]. Hence, if  $X \perp\!\!\!\perp Y | B^\top X$ , the distribution of  $Y$  given  $X$  is equal to the one of  $Y$  given  $B^\top X$ . In the literature, this model is referred to the multiple-index model (single-index model if  $q = 1$ ) and the subspace spanned by the columns of  $B$  is called the Dimension Reduction (DR) subspace. Among the contributions on the estimation of the DR subspace, one can cite the Sliced Inverse Regression (SIR) method introduced by Li [24], the Slice Average Variance Estimation (SAVE) method proposed by Cook and Weisberg [6] and the Principal Hessian Directions (PHD) method (see Li [25]). The existence of a DR subspace is assumed by many authors in order to study the link between  $Y$  and the explanatory variables  $X$ . For instance, to estimate conditional quantiles of fixed order, Wu *et al.* [32] use a single-index model while a combination of SIR and kernel estimation is considered by Gannoun *et al.* [13].

Up to our knowledge, the use of adapted dimension reduction methods for estimating extreme conditional quantiles has not been considered yet in the literature. This is the purpose of the present paper which can be divided into three main parts. 1) We adapt the classical definition of conditional independence to an extreme value context. More specifically, we introduce the notion of Tail Conditional Independence of  $Y$  and  $X$  given  $Z$  which, roughly speaking, consists in assuming that the tail distribution of  $Y$  given  $(X, Z)$  is equivalent to the one of  $Y$  given  $Z$ . This new definition permits us to deal with situations where the covariate dimension can be reduced only in the tail of the distribution. 2) We assume the existence of a  $p \times q$  full rank matrix  $B$  such that  $Y$  and  $X$  are tail conditionally independent given  $B^\top X$ . The subspace spanned by the columns of  $B$  is called the Tail Dimension Reduction (TDR) subspace. We then proposed a kernel based estimator of extreme conditional quantiles by taking into account the existence of the TDR subspace and assuming that the matrix  $B$  is known. 3) We propose a procedure for the estimation of the matrix  $B$ .

The paper is organized as follows. The notion of Tail Conditional Independence is defined rigorously in Section 2. In Section 3, assuming the existence of a TDR subspace, the estimation of extreme conditional quantiles is addressed and the theoretical asymptotic behavior of the proposed estimator is established. The estimation of the TDR subspace is considered in Section 4. The finite sample properties of the proposed estimators is investigated through a simulation study in Section 5 and on a real data set in Section 6. Section 7 collects the proofs.

## 2 Tail conditional independence

Let  $(X, Y, Z) \in \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^q$  be a random vector defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . The goal of this section is to introduce the notion of *tail conditional independence (TCI)* of  $Y$  and  $X$  given  $Z$ . First, let us give some notations used in the rest of the paper. For any random variable  $W : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$  where  $m \in \mathbb{N}^*$ , let  $\text{supp}(W)$  be the support of its distribution. We denote by  $\mathbb{P}(\cdot|W = \cdot) : \mathcal{F} \times \text{supp}(W) \mapsto [0, 1]$  a regular version of the conditional expectation  $\mathbb{E}(\mathbb{I}_{\{\cdot\}}|W)$  where the dot denotes any element of  $\mathcal{F}$  and  $\mathbb{I}_{\{\cdot\}}$  is the indicator function. Finally, the conditional quantile of  $Y$  given  $W$  of order  $\alpha \in [0, 1]$  is the measurable function  $Q(\alpha|W = \cdot) := \inf\{y \in \mathbb{R}; \mathbb{P}(Y > y|W = \cdot) \leq \alpha\}$  where  $\mathbb{R}$  denotes the extended real number line. The TCI property is defined below.

**Definition 1.** *The random variable  $Y$  is tail conditionally independent of  $X$  given  $Z$  (in symbols  $Y \perp\!\!\!\perp X|Z$ ) if for all  $\varepsilon > 0$  there exists  $\kappa > 0$  such that for all  $\delta \in (0, \kappa]$ ,*

$$\mathbb{P} \left[ \left| \frac{\mathbb{P}(Y > \mathcal{Y}_\delta(Z)|X, Z)}{\mathbb{P}(Y > \mathcal{Y}_\delta(Z)|Z)} - 1 \right| \leq \varepsilon \right] = 1, \quad (1)$$

where for  $\delta > 0$ ,  $\mathcal{Y}_\delta(Z = \cdot)$  is a measurable function defined on  $\text{supp}(Z)$  and given by  $\mathcal{Y}_\delta(Z = \cdot) := Q(0|Z = \cdot) - \delta$  if  $Q(0|Z = \cdot) < +\infty$  and  $\mathcal{Y}_\delta(Z = \cdot) := \delta^{-1}$  if  $Q(0|Z = \cdot) = +\infty$ .

Note that, as a direct consequence of (1), if  $Y \perp\!\!\!\perp X|Z$  then  $Q(0|X, Z) = Q(0|Z)$  a.s. that is to say that the distributions of  $Y$  given  $(X, Z)$  and  $Y$  given  $Z$  share the same right endpoint.

For a better understanding of Definition 1, one can remark that if  $Y \perp\!\!\!\perp X|Z$ , there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^p) \otimes \mathcal{B}(\mathbb{R}^q)$  with  $\mathbb{P}[(X, Z) \in \mathcal{A}] = 1$  such that for all  $(x, z) \in \mathcal{A}$ ,

$$\lim_{y \uparrow Q(0, Z=z)} \frac{\mathbb{P}(Y > y|X = x, Z = z)}{\mathbb{P}(Y > y|Z = z)} = 1. \quad (2)$$

Hence, if  $Y \perp\!\!\!\perp X|Z$ , the tail of the conditional distribution of  $Y$  given  $(X, Z)$  is equivalent to the one of  $Y$  given  $Z$  and thus, to make inference on the tail of  $Y$  given  $(X, Z)$ , only the knowledge of  $Z$  is required.

The TCI property is obviously less restrictive than the classical conditional independence property since the last one assumes that  $\mathbb{P}(Y > y|X, Z) = \mathbb{P}(Y > y|Z)$  a.s. for all  $y \in \mathbb{R}$ . Note also that the conditional independence property is symmetric (*i.e.*  $X \perp\!\!\!\perp Y|Z \Leftrightarrow Y \perp\!\!\!\perp X|Z$ ) but not the tail conditional independence property.

As it is the case for the conditional independence property, the TCI property can be characterized in different equivalent ways. This is the purpose of the next result.

**Theorem 1.** *The following statements are equivalent:*

(i)  $Y \underline{\mathbb{L}} X|Z$ .

(ii) For all  $\varepsilon > 0$  there exists  $\kappa > 0$  such that for all  $\delta \in (0, \kappa]$  and for all non-zero bounded and positive measurable function  $h(\cdot)$ ,

$$\mathbb{P} \left[ \left| \frac{\mathbb{E}(\mathbb{I}_{\{Y > \mathcal{Y}_\delta(Z)\}} h(X)|Z)}{\mathbb{P}(Y > \mathcal{Y}_\delta(Z)|Z) \mathbb{E}(h(X)|Z)} - 1 \right| \leq \varepsilon \right] = 1.$$

(iii) For all  $\delta > 0$ ,  $\mathbb{P}(Y > \mathcal{Y}_\delta(Z)|X, Z) = s_\delta(Z)(1 + \eta_\delta(X, Z))$  a.s., where  $s_\delta : \text{supp}(Z) \mapsto \mathbb{R}$  and  $\eta_\delta : \text{supp}(X, Z) \mapsto \mathbb{R}$  are two measurable functions and, for all  $\varepsilon \in (0, 1)$ , there exists  $\kappa > 0$  such that for all  $\delta \in (0, \kappa]$ ,  $\mathbb{P}[|\eta_\delta(X, Z)| \leq \varepsilon] = 1$ .

Note that uniform convergence in (1) (i.e. the fact that  $\kappa$  does not depend on the values of the random vector  $(X, Z)$ ) is essential to prove the second and third statements in Theorem 1. Furthermore, since the conditional expectation is almost surely unique, Theorem 1 can not be established if we ask that (2) holds everywhere.

The second statement gives us another way to understand the TCI property. It entails that for all bounded and positive measurable function  $h(\cdot)$ , there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^q)$  with  $\mathbb{P}(Z \in \mathcal{A}) = 1$  such that for all  $z \in \mathcal{A}$ ,  $\mathbb{E}(\mathbb{I}_{\{Y > y\}} h(X)|Z = z) \sim \mathbb{P}(Y > y|Z = z) \mathbb{E}(h(X)|Z = z)$  as  $y \uparrow Q(0|Z = z)$ . In the particular situation considered later where  $Z = B^\top X$  with  $B$  a full rank  $p \times q$  matrix ( $q < p$ ), this property will be the starting point of the estimation of  $B$ .

The third statement is a very useful tool for anyone who wants to propose conditional distributions of  $Y$  given  $(X, Z)$  satisfying Definition 1. Some examples are given hereafter.

**Example 1.** Let  $\mu_1(\cdot|(X, Z) = \cdot) : \mathcal{B}(\mathbb{R}) \times \text{supp}(X, Z) \mapsto ([0, 1], \mathcal{B}([0, 1]))$  and  $\mu_2(\cdot|Z = \cdot) : \mathcal{B}(\mathbb{R}) \times \text{supp}(Z) \mapsto ([0, 1], \mathcal{B}([0, 1]))$  be two functions such that for all  $(x, z) \in \text{supp}(X, Z)$ ,  $\mu_1(\cdot|(X, Z) = (x, z))$  and  $\mu_2(\cdot|Z = z)$  are two probability measures on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and, for all  $A \in \mathcal{B}(\mathbb{R})$ , the functions  $\mu_1(A|(X, Z) = \cdot)$  and  $\mu_2(A|Z = \cdot)$  are measurable. We assume in addition that uniformly on  $(x, z) \in \text{supp}(X, Z)$ ,

$$\lim_{\delta \rightarrow 0} \frac{\mu_1(I_\delta(z)|(X, Z) = (x, z))}{\mu_2(I_\delta(z)|Z = z)} = 0.$$

where  $I_\delta(z) = [\mathcal{Y}_\delta(Z = z), \infty)$ . Regardless the distribution of  $(X, Z)$ , if the conditional distribution of  $Y$  given  $(X, Z)$  is the mixture distribution defined by

$$\mathbb{P}(Y \in \cdot|X, Z) = \theta(Z)\mu_1(\cdot|X, Z) + (1 - \theta(Z))\mu_2(\cdot|Z) \text{ a.s.},$$

where  $\theta(\cdot)$  is a  $[0, 1]$ -valued measurable function, then, from Theorem 1 point (iii), it is easy to see that  $Y \underline{\mathbb{L}} X|Z$ .

**Example 2.** Consider the semi-parametric model

$$\mathbb{P}(Y > y|X) = y^{-\exp(b_0^\top X)} \mathcal{L}(y|X) \text{ a.s.},$$

where  $b_0 \in \mathbb{R}^p$  and  $\mathcal{L}(\cdot|X = \cdot) : \mathbb{R} \times \text{supp}(X) \mapsto (0, \infty)$  is a function such that for all  $t > 0$  and  $x \in \text{supp}(X)$ ,  $\mathcal{L}(ty|X = x)/\mathcal{L}(y|X = x) \rightarrow 1$  as  $y \rightarrow \infty$ . Note that this model was introduced by Wang and Tsai [29] where a maximum likelihood method to estimate  $b_0$  is considered. If we assume that  $\mathcal{L}(y|X = x)$  converges to  $c(b_0^\top x)$  uniformly on  $x \in \text{supp}(X)$  as  $y \rightarrow \infty$  where  $c(\cdot)$  is a positive and measurable function then, using the third statement of Theorem 1, it is easy to check that  $Y \underline{\mathbb{L}} X|b_0^\top X$ .

### 3 Extreme quantile estimation under a TDR model

Let  $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$  be a random vector defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . In what follows, the distribution of  $(X, Y)$  is assumed to be absolutely continuous with respect to the Lebesgue measure. A probability distribution function of  $X$  is denoted by  $f_X(\cdot)$  and its support is given by  $\text{supp}(X) := \{x \in \mathbb{R}^p; f_X(x) > 0\}$  which is assumed to be an open set.

The aim of this section is to propose an estimator of the conditional quantile  $Q(\alpha|X)$  when  $\alpha$  is closed to 0 (extreme conditional quantile) and when the dimension  $p$  of the covariate  $X$  is large ( $p \geq 3$ ). To reach this goal, the existence of a *tail dimension reduction (TDR)* subspace is assumed. The definition of such a subspace is given in the next paragraph. The estimator and its asymptotic properties are given in the second paragraph.

#### 3.1 Tail dimension reduction (TDR) subspace

Using the notion of tail conditional independence presented in the previous section, we give the definition of a TDR subspace. In what follows, for a full rank  $p \times q$  matrix  $B$  with  $q < p$ , the space spanned by the columns of  $B$  is denoted  $\mathcal{S}(B)$ .

**Definition 2.** *If there exists a full rank  $p \times q$  matrix  $B_0$  with  $q < p$  such that  $Y \perp\!\!\!\perp X|B_0^\top X$  then  $\mathcal{S}(B_0)$  is a Tail Dimension Reduction (TDR) subspace for  $Y$  given  $X$ .*

The TDR subspace is an adaptation of the DR subspace introduced by Li [24]. Roughly speaking, if there exists a TDR subspace  $\mathcal{S}(B_0)$ , the tail of the conditional distribution of  $Y$  given  $X$  can be reasonably approximated by the tail of the conditional distribution of  $Y$  given  $B_0^\top X$ . Obviously, for any random vector  $(X, Y)$ ,  $\mathcal{S}(I_p) = \mathbb{R}^p$  is a TDR subspace and thus a TDR subspace is not unique. Since our goal is to reduce the dimension, the notion of *minimum TDR subspace* is defined below by analogy with the definition of the minimum DR subspace (see for instance [4]).

**Definition 3.** *For a full rank matrix  $B_0$ , the subspace  $\mathcal{S}(B_0)$  is a minimum TDR subspace if its dimension is lower or equal than the dimension of any other TDR subspace.*

Note that for any regular  $q \times q$  matrix  $D$ ,  $Y \perp\!\!\!\perp X|B_0^\top X$  is equivalent to  $Y \perp\!\!\!\perp X|(B_0 D)^\top X$ . Hence, the matrix  $B_0$  is not uniquely defined. From now on, we assume that  $B_0$  belongs to the set  $\mathcal{B}$  defined in the following way:  $B \in \mathcal{B}$  if  $B$  is given by the first normalized  $q$  linearly independent columns of the projection matrix  $P_B := B(B^\top B)^{-1}B^\top$  on  $\mathcal{S}(B)$ . In this section, the matrix  $B_0$  is supposed to be known. Its estimation will be considered in Section 4. In the next paragraph, an estimator of the conditional quantile  $Q(\cdot|X)$  is proposed using the information that  $\mathcal{S}(B_0)$  is a minimum TDR subspace.

#### 3.2 Extreme quantile estimation

Given  $n$  independent copies  $(X_1, Y_1), \dots, (X_n, Y_n)$  of the random vector  $(X, Y)$ , the aim of this section is the estimation of the extreme conditional quantile  $Q(\alpha_n|X = x)$  for  $x \in \text{supp}(X)$  where  $(\alpha_n)$  is a sequence in  $(0, 1)$  converging to 0 as  $n$  goes to infinity. Without assuming the existence of a minimum TDR subspace, this question has already been investigated by several authors (see for instance Araújo Santos *et al.* [1], Daouia *et al.* [7], Gardes and Girard [16] among many others). Unfortunately, these estimators often fail to approximate correctly the extreme conditional quantile in a large dimension setting (usually  $p \geq 3$ ) since in this situation, the space  $\text{supp}(X)$

is only sparsely populated by data points. As a consequence, only few points can be reasonably considered to estimate  $Q(\alpha_n|X = x)$  and, unless the sample size is very large, classical estimators become inefficient for large values of  $p$ .

The first step in the estimation of  $Q(\alpha_n|X = x)$  is often the estimation of the conditional survival function  $\mathbb{P}(Y > y|X = x)$  for large value of  $y$ . A classical way to reach this goal is to use a kernel based estimator (see for instance Daouia *et al.* [7]). In this paper, taking into account the existence of a minimum TDR subspace  $\mathcal{S}(B_0)$ , a new kernel estimator of  $Q(\alpha_n|X = x)$  is proposed. The basic idea is that the conditional survival function  $\mathbb{P}(Y > y|X = x)$  can be approximated for  $y$  large enough by  $\mathbb{P}(Y > y|B_0^\top X = B_0^\top x)$ . If  $B_0$  is known, we propose the following statistics as an estimator of  $\mathbb{P}(Y > y|X = x)$ :

$$\hat{S}_n(y|B_0, x) := \sum_{i=1}^n \mathbb{I}_{\{Y_i > y\}} K(H_n^{-1} B_0^\top (x - X_i)) \bigg/ \sum_{i=1}^n K(H_n^{-1} B_0^\top (x - X_i)). \quad (3)$$

Here,  $H_n$  is a  $q \times q$  positive definite matrix and  $K(\cdot)$  is a probability distribution function on  $\mathbb{R}^q$ . From now on, we assume that  $K(\cdot)$  is bounded with support the unit ball of  $\mathbb{R}^q$ . Note that  $\hat{S}_n(y|B_0, x)$  is the classical kernel estimator of  $\mathbb{P}(Y > y|B_0^\top X = B_0^\top x)$  that is considered here as an estimator of  $\mathbb{P}(Y > y|X = x)$ .

A first attempt to estimate  $Q(\alpha_n|X = x)$  is to use the generalized inverse of  $\hat{S}_n(\cdot|B_0, x)$  leading to the statistics

$$\hat{Q}_n(\alpha_n|B_0, x) := \inf\{y; \hat{S}_n(y|B_0, x) \leq \alpha_n\}. \quad (4)$$

Unfortunately, such an estimator fails to estimate an extreme quantile with an order  $\alpha_n$  as small as we like. Indeed, it is shown in Proposition 4 that the condition  $n|H_n|\alpha_n \rightarrow \infty$ , where  $|H_n|$  stands for the determinant of  $H_n$ , is required to establish the consistency of (4). As a consequence,  $Q(\alpha_n|X = x)$  cannot be consistently estimated by (4) when  $\alpha_n$  is too small. To overcome this drawback, an additional information on the tail distribution of  $Y$  given  $X$  is necessary.

**Regular varying condition.** In the unconditional case, when dealing with the right-tail of a real random variable  $Y$ , it is commonly assumed that  $Y$  belongs to the maximum domain of attraction of an extreme value distribution (see Fisher and Tippett [11] and Gnedenko [19] for a definition). According to de Haan and Ferreira [20, Theorem 1.1.6], this is equivalent to assume the existence of a positive auxiliary function  $a_Y(\cdot)$  and a parameter  $\gamma_Y \in \mathbb{R}$  such that for all  $u > 0$ ,  $[Q_Y(u\alpha) - Q_Y(\alpha)]/a_Y(\alpha^{-1}) \rightarrow L_{\gamma_Y}(1/u)$  as  $\alpha \rightarrow 0$  where  $Q_Y(\alpha) = \inf\{y; \mathbb{P}(Y > y) \leq \alpha\}$  and with for all  $v \geq 1$  and  $s \in \mathbb{R}$ ,  $L_s(v) := \int_1^v u^{s-1} du$ . The function  $Q_Y(\cdot)$  is then said to be of extended regular variation.

The same kind of assumption is made in our conditional setting. We assume that for all  $x \in \text{supp}(X)$ , the function  $Q(\cdot|X = x)$  is of extended regular variation *i.e.* that there exist a positive function  $a(\cdot|x) : (0, \infty) \mapsto (0, \infty)$  and a real-valued function  $\gamma(\cdot)$  such that for all  $x \in \text{supp}(X)$  and  $u > 0$ ,

$$\text{ERV}(\alpha, u|x) := \left| \frac{Q(u\alpha|X = x) - Q(\alpha|X = x)}{a(\alpha^{-1}|x)} - L_{\gamma(x)}(1/u) \right| \rightarrow 0, \quad (5)$$

as  $\alpha \rightarrow 0$ . Note that this convergence holds locally uniformly on  $u \in (0, \infty)$ . The function  $\gamma(\cdot)$  is referred to as the conditional extreme value index function. The function  $a(\cdot|x)$  is called the auxiliary function and satisfied  $a(\alpha^{-1}|x)/Q(\alpha|X = x) - \gamma_+(x) \rightarrow 0$  as  $\alpha$  goes to 0 where  $(\cdot)_+$  and  $(\cdot)_-$  are the positive and negative part functions (see Fraga Alves *et al.* [12, Lemma 3.1]).

Condition (5) is equivalent to assume that the distribution function  $\mathbb{P}(Y \leq \cdot | X = x)$  belongs to the maximum domain of attraction of an extreme value distribution with extreme value index  $\gamma(x)$ . Condition (5) is also satisfied for the conditional quantile of  $Y$  given  $B_0^\top X$  as it is shown in the following result.

**Proposition 1.** *Let  $B_0$  be a full rank matrix such that  $\mathcal{S}(B_0)$  is a TDR subspace. If for all  $x \in \text{supp}(X)$ , the conditional quantile  $Q(\cdot | X = x)$  satisfies (5), there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^p)$  with  $\mathbb{P}(X \in \mathcal{A}) = 1$ , a positive function  $\tilde{a}(\cdot | B_0^\top x)$  and a real-valued function  $\tilde{\gamma}(\cdot)$  such that for all  $u > 0$  and  $x \in \mathcal{A}$ ,*

$$\lim_{\alpha \rightarrow 0} \frac{Q(u\alpha | B_0^\top X = B_0^\top x) - Q(\alpha | B_0^\top X = B_0^\top x)}{\tilde{a}(\alpha^{-1} | B_0^\top x)} = L_{\tilde{\gamma}(B_0^\top x)}(1/u).$$

In addition, for all  $x \in \mathcal{A}$ ,  $\gamma(x) = \tilde{\gamma}(B_0^\top x)$  and  $\tilde{a}(\alpha^{-1} | B_0^\top x)/a(\alpha^{-1} | x) \rightarrow 1$  as  $\alpha \rightarrow 0$ .

To estimate  $Q(\beta_n | X = x)$  for an arbitrary sequence  $(\beta_n)$  converging to 0, we start with (5) which suggests the approximation  $Q(\beta_n | X = x) \approx Q(\alpha_n | X = x) + a(\alpha_n^{-1} | x)L_{\gamma(x)}(\alpha_n/\beta_n)$ . The sequence  $(\alpha_n)$  is chosen not too small so that  $Q(\alpha_n | X = x)$  can be consistently estimated by the kernel estimator  $\hat{Q}_n(\alpha_n | B_0, x)$  defined in (4). Assuming as before that  $\mathcal{S}(B_0)$  is a TDR subspace for a given full rank matrix  $B_0$ , an estimator of  $Q(\beta_n | X = x)$  is thus given by

$$\check{Q}_n(\beta_n | B_0, x) := \hat{Q}_n(\alpha_n | B_0, x) + \hat{a}_n(B_0, x)L_{\hat{\gamma}_n(B_0, x)}(\alpha_n/\beta_n), \quad (6)$$

where  $\hat{\gamma}_n(B_0, x)$  and  $\hat{a}_n(B_0, x)$  are consistent estimators of  $\gamma(x)$  and  $a(\alpha_n^{-1} | x)$ . In this paper, the estimator  $\hat{\gamma}_n(B_0, x)$  is given by

$$\begin{aligned} \hat{\gamma}_n(B_0, x) &:= \hat{\gamma}_{n,+}(B_0, x) + \hat{\gamma}_{n,-}(B_0, x) \\ &= \mathcal{T}_{\alpha_n}^{(1)}(\hat{Q}_n(\cdot | B_0, x)) + \Psi^\leftarrow \left( \max \left\{ \frac{[\mathcal{T}_{\alpha_n}^{(1)}(\hat{Q}_n(\cdot | B_0, x))]^2}{\mathcal{T}_{\alpha_n}^{(2)}(\hat{Q}_n(\cdot | B_0, x))}; \Psi(0) \right\} \right). \end{aligned} \quad (7)$$

For  $\nu \in (0, 1)$  and  $\varphi(\cdot)$  a positive and bounded function on  $[\nu, 1]$ ,  $\Psi(\cdot)$  is the decreasing function defined for  $s \leq 0$  by

$$\Psi(s) := \left( \int_\nu^1 \varphi(u)L_s(1/u)du \right)^2 \Big/ \int_\nu^1 \varphi(u)L_s^2(1/u)du,$$

and  $\Psi^\leftarrow(t) = \inf\{s; \Psi(s) \leq t\}$  is the generalized inverse of  $\Psi(\cdot)$ . In addition, for all  $\delta \in \mathbb{N}$ , for all non-increasing and right-continuous function  $U(\cdot)$  and all  $\alpha \in (0, 1)$ , the quantity  $\mathcal{T}_\alpha^{(\delta)}(U)$  is defined by

$$\mathcal{T}_\alpha^{(\delta)}(U) := \int_\nu^1 \varphi(u) \left( \ln \frac{U(u\alpha)}{U(\alpha)} \right)^\delta du \Big/ \left( \int_\nu^1 \varphi(u)L_0(1/u)du \right)^\delta.$$

Note that this estimator of  $\gamma(x)$  belongs to the class of estimators introduced in Gardes [14]. Concerning the estimation of  $a(\alpha_n^{-1} | x)$ , we consider the statistics

$$\hat{a}_n(B_0, x) = \tilde{\mathcal{T}}_{\alpha_n} \left( \hat{Q}_n(\cdot | B_0, x); \hat{\gamma}_{n,-}(B_0, x) \right), \quad (8)$$

where  $\tilde{\mathcal{T}}_\alpha(U, \gamma_-)$  is given for all non-increasing and right-continuous function  $U(\cdot)$ , for all  $\gamma_- \leq 0$  and for all  $\alpha \in (0, 1)$  by

$$\tilde{\mathcal{T}}_\alpha(U, \gamma_-) := U(\alpha) \int_\nu^1 \varphi(u) \ln \frac{U(u\alpha)}{U(\alpha)} du \Big/ \int_\nu^1 \varphi(u)L_{\gamma_-}(1/u)du.$$

Note that the previous defined estimators depend on the choice of a parameter  $\nu \in (0, 1)$  and a positive and bounded function  $\varphi(\cdot)$  on  $[\nu, 1]$ . In order to not overload the notations, this dependence has been omitted. The expressions of  $\hat{\gamma}_n(B_0, x)$  and  $\hat{a}_n(B_0, x)$  are motivated by the following result.

**Proposition 2.** *Let  $x \in \text{supp}(X)$ . If condition (5) holds then*

$$\lim_{\alpha \rightarrow 0} \mathcal{T}_\alpha^{(1)}(Q(\cdot|X=x)) = \gamma_+(x), \quad \lim_{\alpha \rightarrow 0} \frac{[\mathcal{T}_\alpha^{(1)}(Q(\cdot|X=x))]^2}{\mathcal{T}_\alpha^{(2)}(Q(\cdot|X=x))} = \Psi(\gamma_-(x)),$$

and  $\lim_{\alpha \rightarrow 0} \frac{\tilde{\mathcal{T}}_\alpha(Q(\cdot|X=x), \gamma_-(x))}{a(\alpha^{-1}|x)} = 1.$

The proof of this Proposition is a direct consequence of [14, Lemma 3]. Its proof is thus omitted. The study of the asymptotic behavior of  $\check{Q}_n(\beta_n|B_0, x)$  is done under the assumptions given below.

**Condition on the TDR subspace.** Recall that, from Theorem 1,  $\mathcal{S}(B_0)$  is a TDR subspace if and only if for all  $\delta > 0$  there exist two measurable functions  $s_\delta(\cdot)$  and  $\eta_\delta(\cdot)$  such that  $\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X)|X) = s_\delta(B_0^\top X)(1 + \eta_\delta(X))$  almost surely. As shown in the proof of Theorem 1, (see equation (21)), the function  $\eta_\delta(\cdot)$  controls the rate of convergence of the ratio  $\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X)|X)/\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X)|B_0^\top X)$  to 1. In the sequel, we assume that there exist  $\kappa > 0$  and a decreasing function  $\bar{\eta}(\cdot)$  converging to 0 at infinity such that for all  $\delta \in (0, \kappa]$

$$\mathbb{P}[|\eta_\delta(X)| \leq \bar{\eta}(\delta^{-1})] = 1. \quad (9)$$

**Regularity condition.** Since the distribution of  $X$  is absolutely continuous with respect to the Lebesgue measure and that  $B_0$  is a full rank matrix, the random variable  $B_0^\top X$  is also absolutely continuous with a probability distribution function  $f_{B_0^\top X}(\cdot)$  such that  $f_{B_0^\top X}(B_0^\top x) > 0$  for all  $x \in \text{supp}(X)$ . The following regularity condition on  $f_{B_0^\top X}(\cdot)$  is needed. For all  $(s, t) \in (\text{supp}(X))^2$ , there exists a constant  $c_0 > 0$  such that

$$|f_{B_0^\top X}(B_0^\top s) - f_{B_0^\top X}(B_0^\top t)| \leq c_0 \|B_0^\top(s - t)\|_\infty, \quad (10)$$

where  $\|\cdot\|$  denotes any norm on  $\mathbb{R}^q$ .

We can now state the main result of this paragraph. The following notations are used. For all positive definite  $q \times q$  matrix  $M$  and  $s \in \mathbb{R}^q$ , let  $D(s, M) := \{t \in \mathbb{R}^q; \|M^{-1}(s - t)\|_\infty \leq 1\}$ . For  $v \geq 1$  and  $s \in \mathbb{R}$ , let  $\tilde{L}_s(v) := \int_1^v u^{s-1} \ln u \, du$ . Finally, for all random variable  $W$ , let  $\delta_\alpha(W = \cdot)$  be the measurable function defined for  $t \in \text{supp}(W)$  and  $\alpha \in (0, 1)$  by

$$\delta_\alpha(W = t) := \begin{cases} Q(0|W=t) - Q(\alpha|W=t) & \text{if } Q(0|W=t) < \infty, \\ 1/Q(\alpha|W=t) & \text{if } Q(0|W=t) = \infty. \end{cases} \quad (11)$$

**Theorem 2.** *Assume that there exists a full rank matrix  $B_0$  such that  $\mathcal{S}(B_0)$  is a TDR subspace and suppose conditions (9) and (10) hold. Let  $(\alpha_n)$ ,  $(\beta_n)$  and  $(H_n)$  be sequences such that  $\alpha_n \rightarrow 0$ ,  $\beta_n/\alpha_n \rightarrow 0$ ,  $n|H_n|\alpha_n \rightarrow \infty$ ,  $\tau_n \ln^2(\alpha_n/\beta_n) \rightarrow 0$  and  $\tau_n^{-1}\|H_n\|_\infty \rightarrow 0$ , where  $\tau_n := (n|H_n|\alpha_n)^{-1/2}[\ln(n|H_n|\alpha_n)]^{1/2}$ . If there exists  $\xi \in (0, 1)$  such that for all  $x \in \text{supp}(X)$ ,*

$$\max \left\{ \sup_{(t, \zeta) \in A_n} \left| \frac{\mathbb{P}(Y > Q(\zeta|B_0^\top X = B_0^\top x)|B_0^\top X = t)}{\zeta} - 1 \right| ; \bar{\eta}(\xi \delta_{\alpha_n}^{-1}(X = x)) \right\} = \mathcal{O}(\tau_n), \quad (12)$$

where  $A_n := D(B_0^\top x, H_n) \times [\xi \nu \alpha_n, \xi^{-1} \alpha_n]$ , if  $\tau_n^{-1} \text{ERV}(\alpha_n, u | X = x) \rightarrow 0$  locally uniformly on  $u \in (0, \infty)$  and

$$\lim_{n \rightarrow \infty} \tau_n^{-1} \max \left\{ \left| \frac{a(\alpha_n^{-1} | x)}{Q(\alpha_n | X = x)} - \gamma_+(x) \right| ; \frac{\text{ERV}(\alpha_n, \beta_n / \alpha_n | X = x)}{\tilde{L}_{\gamma(x)}(\alpha_n / \beta_n)} \right\} = 0, \quad (13)$$

then, there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^p)$  with  $\mathbb{P}(X \in \mathcal{A}) = 1$  and such that for all  $x \in \mathcal{A}$ ,

$$\frac{\tilde{Q}(\beta_n | B_0, x) - Q(\beta_n | X = x)}{a(\alpha_n^{-1} | X = x) \tilde{L}_{\gamma(x)}(\alpha_n / \beta_n)} = \mathcal{O}_{\mathbb{P}}(\tau_n).$$

### Remarks.

- Because of the definition (1) of the conditional tail independence, the convergence of the estimator of the conditional quantile cannot be obtained for every  $x \in \text{supp}(X)$ .
- In (12), two conditions are gathered:

The first one is a classical regularity condition of the function  $\mathbb{P}(Y > y | B_0^\top X = \cdot)$  for large values of  $y$ . This condition is essential in a conditional framework (see for instance Daouia *et al.* [7] and Gardes [14]). A careful reading of the proof shows that this condition only need to be satisfied for  $x \in \mathcal{A}$ . This condition involves the conditional distribution of  $Y$  given  $B_0^\top X$ . In Lemma 3 (see appendix), it is shown that it can be replaced by a regularity condition on the conditional distribution of  $Y$  given  $X$  which is often more convenient to check. Note also that the parameter  $\nu$  in the set  $A_n$  is the one used in the estimators (7) and (8) of  $\gamma(x)$  and  $a(\alpha_n^{-1} | x)$ .

The second condition is required to deal with the TDR subspace. Roughly speaking, this condition ensures that the ratio  $\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X) | X) / \mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X) | B_0^\top X)$  converges to 1 sufficiently fast in order to obtain a consistent estimator of  $Q(\alpha_n | X = x)$  with rate of convergence  $\tau_n$ .

- We would like now to show that if  $q < p$  (*i.e* if the dimension of the covariate can be reduced), the new estimator  $\tilde{Q}_n(\beta_n | B_0, x)$  is more efficient than the estimator  $\tilde{Q}_n(\beta_n | x) := \tilde{Q}_n(\beta_n | I_p, x)$  which does not take into account the existence of the TDR subspace. Assume that conditions of Theorem 2 are satisfied for sequences  $(\alpha_n)$ ,  $(\beta_n)$  and a bandwidth matrix  $H_n = h_n I_q$  where  $(h_n)$  is a positive sequence tending to 0. Let  $M_n := h_n^{q/p} I_p$ . If there exists  $\xi \in (0, 1)$  such that

$$\sup_{(t, \zeta) \in \tilde{A}_n} \left| \frac{\mathbb{P}(Y > Q(\zeta | X = x) | X = t)}{\zeta} - 1 \right| = \mathcal{O}(\tau_n),$$

where  $\tilde{A}_n := D(x, M_n) \times [\xi \alpha_n, \xi^{-1} \alpha_n]$  and if  $\tau_n h_n^{q/p} \rightarrow 0$  with  $\tau_n := [n h_n^q \alpha_n / \ln(n h_n^q \alpha_n)]^{-1/2}$ , then,  $\tilde{Q}_n(\beta_n | x)$  computed with the bandwidth matrix  $M_n$  and  $\tilde{Q}_n(\beta_n | B_0, x)$  computed with  $H_n$  are both consistent estimators of  $Q(\alpha_n | X = x)$  with the same rate of convergence  $\tau_n$ . Since  $h_n / h_n^{q/p} \rightarrow 0$  (when  $q < p$ ), estimator  $\tilde{Q}_n(\beta_n | B_0, x)$  shall performs better in practice because, to keep the same rate of convergence, estimator  $\tilde{Q}_n(\beta_n | x)$  have to use observations located far away from the target leading to an important bias. To illustrate this fact, let  $B_0 = (1, 2)^\top / \sqrt{5}$  and  $x_0 = (1/2, 1/2)^\top$ . Suppose that  $\mathcal{S}(B_0)$  is a TDR subspace and that one is interested in the estimation of a conditional extreme quantile of  $Y$  given  $X = x_0$  for a random vector  $X = (X^{(1)}, X^{(2)})^\top$  with  $X^{(1)}$  and  $X^{(2)}$  independent and uniformly distributed on  $(0, 1)$ . In this situation, the points of interest are those located on the set  $\{x = (x^{(1)}, x^{(2)}); B_0^\top x = B_0^\top x_0\}$ . On Figure 1, observations among  $n = 500$  independent copies  $X_1, \dots, X_n$  of  $X$  for which  $K(h_n^{-1} B_0^\top (x_0 - X_i)) > 0$  (left panel) and  $K(M_n^{-1} (x_0 - X_i)) > 0$  (right panel) are represented for  $h_n = 1/10$ . It clearly appears that the observations used to compute  $\tilde{Q}_n(\cdot | B_0, x)$  (right panel) are more relevant than the ones used in  $\tilde{Q}_n(\cdot | x)$  (left panel).

## 4 Estimation of the TDR subspace

In this section, a procedure to estimate the subspace  $\mathcal{S}(B_0)$  (*i.e.* the matrix  $B_0$ ) is proposed. In what follows, it is assumed that  $\text{supp}(X) = \mathbb{R}^p$ . In this case, we have the following result.

**Proposition 3.** *If  $\text{supp}(X) = \mathbb{R}^p$  then a minimum TDR subspace is unique.*

The starting point for the estimation of  $B_0$  is a result showing that  $B_0$  can be seen as the solution of a minimization problem. Let us first introduce some notations. For  $J \in \mathbb{N}^*$  and for any matrix  $B \in \mathcal{B}$ , let  $\{\Pi_j(B^\top X), j = 1, \dots, J\}$  be a random partition of  $\text{supp}(X)$ . In the sequel, it is assumed that for all  $B \in \mathcal{B}$  and all  $j = 1, \dots, J$ ,  $\mathbb{P}(X \in \Pi_j(B^\top X) | B^\top X) > 0$  a.s. (an example of such a partition is given in section 5). Let us also introduce the function  $T : (0, 1) \times \mathcal{B} \mapsto [0, \infty]$  defined by

$$T(\alpha, B) := \sum_{j=1}^J \left\{ \mathbb{E} \left[ \frac{\mathbb{P}(\{Y > Q(\alpha | B^\top X)\} \cap \{X \in \Pi_j(B^\top X)\} | B^\top X)}{\alpha \mathbb{P}(X \in \Pi_j(B^\top X) | B^\top X)} \right] - 1 \right\}^2.$$

According to the second statement of Theorem 1, the quantity  $T(\alpha, B_0)$  is close to 0 for small values of  $\alpha$ . This argument suggests that an approximation of  $B_0$  can be obtained by minimizing the function  $T(\alpha, B)$  with  $\alpha$  small. More precisely, introducing the notation

$$\tilde{B}_0(\alpha) := \arg \min_{B \in \mathcal{B}} T(\alpha, B), \quad (14)$$

we have the following result.

**Theorem 3.** *Assume that there exists a full rank matrix  $B_0 \in \mathcal{B}$  such that  $\mathcal{S}(B_0)$  is a minimum TDR subspace and that  $\text{supp}(X) = \mathbb{R}^p$ . If for all  $\kappa > 0$ , there exists  $\alpha_0 \in (0, 1)$  such that for all  $\alpha \in (0, \alpha_0)$ ,  $\mathbb{P}[\delta_\alpha(B_0^\top X) < \kappa] = 1$  and if there exists a function  $\tilde{L}(\cdot) : \mathcal{B} \mapsto [0, \infty)$  such that for all  $\varepsilon > 0$ , there exists  $\alpha_0 \in (0, 1)$  such that for all  $\alpha \in (0, \alpha_0)$ ,*

$$\sup_{B \in \mathcal{B}} \left| \min \left\{ \left| \frac{\mathbb{P}(Y > Q(\alpha | B^\top X) | X)}{\alpha} - 1 \right| ; \left| \frac{\mathbb{P}(Y > Q(\alpha | B^\top X) | X)}{\alpha} - 1 \right|^{-1} \right\} - \tilde{L}(B) \right| \leq \varepsilon, \quad (15)$$

*then  $\|\tilde{B}_0(\alpha) - B_0\| \rightarrow 0$  as  $\alpha \rightarrow 0$  where  $\|\cdot\|$  is any matrix norm.*

The condition on the random variable  $\delta_\alpha(B_0^\top X)$  (which was defined in (11)) entails that  $Q(\alpha | B_0^\top X = B_0^\top x)$  converges to the endpoint  $Q(0 | B_0^\top X = B_0^\top x)$  uniformly on  $x \in \text{supp}(X)$  as  $\alpha$  goes to 0. Condition (15) ensures that, uniformly on  $B \in \mathcal{B}$ ,  $|\alpha^{-1} \mathbb{P}(Y > Q(\alpha | B^\top X) | X) - 1|$  admits a positive (possibly infinite) limit as  $\alpha$  goes to 0.

Theorem 3 suggests to estimate  $B_0$  by replacing  $T(\alpha, B)$  in (14) by a reasonable estimator with  $\alpha$  sufficiently small. To construct this estimator, let us introduce a sequence  $(\alpha_n)$  converging to 0 with the sample size. The sample analog of  $T(\alpha_n, B)$  is given by:

$$\frac{1}{n^2} \sum_{j=1}^J \left\{ \sum_{i=1}^n \left( \frac{\Phi_{n,j}(B^\top X_i)}{\alpha_n p_j(B^\top X_i)} - 1 \right) \right\}^2, \quad (16)$$

with for  $B \in \mathcal{B}$ ,  $z \in \text{supp}(B^\top X)$  and  $j \in \{1, \dots, J\}$ ,  $p_j(z) := \mathbb{P}(X \in \Pi_j(B^\top X) | B^\top X = z) f_{B^\top X}(z)$  (where  $f_{B^\top X}(\cdot)$  is the probability density function of  $B^\top X$ ) and

$$\frac{\Phi_{n,j}(z)}{f_{B^\top X}(z)} := \mathbb{P}(\{Y > Q(\alpha | B^\top X = z)\} \cap \{X \in \Pi_j(B^\top X)\} | B^\top X = z).$$

Obviously, in practice, the random variables  $\Phi_{n,j}(B^\top X_i)$  and  $p_j(B^\top X_i)$  are not observed and need to be replaced by their respective kernel estimators:

$$\hat{\Phi}_{n,j}(B^\top X_i) := \sum_{\ell \neq i} \mathbb{I}_{\{Y_\ell > \hat{Q}_{n,-i}(\alpha_n | B^\top X_i)\}} \mathbb{I}_{\{X_\ell \in \Pi_B^{(j)}(B^\top X_i)\}} K(H_n^{-1} B^\top (X_i - X_\ell)),$$

where  $\hat{Q}_{n,-i}(\alpha_n | B^\top X_i)$  is the conditional quantile estimator defined in (4) computed without the couple  $(X_i, Y_i)$  and

$$\hat{p}_j(B^\top X_i) := \sum_{\ell \neq i} \mathbb{I}_{\{X_\ell \in \Pi_B^{(j)}(B^\top X_i)\}} K(H_n^{-1} B^\top (X_i - X_\ell)).$$

We can now introduce our estimator of  $B_0$ :

$$\hat{B}_{0,n}(\alpha_n) := \arg \min_{B \in \mathcal{B}} \hat{T}_n(B), \quad (17)$$

where  $\hat{T}_n(B)$  is obtained by replacing in (16) the unobserved random variables  $\Phi_{n,j}(B^\top X_i)$  and  $p_j(B^\top X_i)$  by  $\hat{\Phi}_{n,j}(B^\top X_i)$  and  $\hat{p}_j(B^\top X_i)$ .

An important task consists in showing the consistency of  $\hat{B}_{0,n}(\alpha_n)$  as an estimator of  $B_0$ . A possible way is to follow the lines of the proof of [22, Theorem 5.1] where an  $M$ -estimator of the direction in a single-index model is proposed. This proof requires in particular to show the uniform consistency on  $B \in \mathcal{B}$  of the estimator  $\hat{T}_n(B)$  of  $T(B)$ . This is of course an interesting but non-trivial result that is beyond the scope of the present paper. In what follows, we only focus on the finite-sample behavior of the TDR estimator  $\hat{B}_{0,n}(\alpha_n)$  and also of the plug-in conditional quantile estimator  $\hat{Q}_n(\beta_n | \hat{B}_{0,n}(\alpha_n), x)$ .

## 5 Simulation study

### 5.1 A procedure for estimating the TDR subspace

Let  $(X, Y)$  be a random vector and assume that there exists a full rank matrix  $B_0$  of rank  $q < p$  such that  $\mathcal{S}(B_0)$  is a TDR subspace. The aim of this paragraph is to present a procedure that can be used to compute the estimator (17) of the direction  $B_0$ .

First, for  $B \in \mathcal{B}$  and  $x \in \text{supp}(X)$ , the partition  $\{\Pi_j(B^\top X = B^\top x), j = 1, \dots, J\}$  of  $\text{supp}(X)$  is constructed in the following way: let  $D = [d_1, \dots, d_p]$  be a  $p \times p$  orthogonal matrix such the first  $q$  columns spanned  $\mathcal{S}(B)$  (such a matrix is obtained by using the Gram-Schmidt process). Let  $m(B^\top x)$  be the conditional marginal median of  $X$  given  $B^\top X = B^\top x$ . For  $\ell \in \{1, \dots, p - q\}$ , introducing the half spaces

$$E_\ell(B^\top x) := \{s \in \mathbb{R}^p; d_{\ell+q}^\top s > d_{\ell+q}^\top m(B^\top x)\}$$

and  $\bar{E}_\ell(B^\top x) := \{s \in \mathbb{R}^p; d_{\ell+q}^\top s \leq d_{\ell+q}^\top m(B^\top x)\},$

an element of the partition  $\{\Pi_j(B^\top X = B^\top x), j = 1, \dots, J\}$  is the intersection of  $p - q$  half spaces. More specifically, an element of the partition is a set  $E_1^* \cap \dots \cap E_{p-q}^*$  where for  $\ell \in \{1, \dots, p - q\}$ ,  $E_\ell^* \in \{E_\ell(B^\top x), \bar{E}_\ell(B^\top x)\}$ . There is thus  $J = 2^{p-q}$  elements in the partition. Obviously, if  $\text{supp}(X) = \mathbb{R}^p$ , then, for all  $x \in \mathbb{R}^p$  and for all  $j \in \{1, \dots, J\}$ ,  $\mathbb{P}(X \in \Pi_j(B^\top X = B^\top x) | B^\top X = B^\top x) > 0$ . In practice, since the conditional marginal median  $m(B^\top x)$  is unknown, it is replaced by its empirical estimator  $\hat{m}_n(B^\top x) := (\hat{m}_{n,j}(B^\top x), j = 1, \dots, J)^\top$  where for  $j = 1, \dots, J$ ,  $\hat{m}_{n,j}(B^\top x)$  is the empirical median of the  $j$ -th component of the observations falling in  $D(B^\top x, H_n)$ . This choice of

random partition ensures that, for all  $x \in \text{supp}(X)$ , the number of available observations in each element of  $D(B^\top x, H_n) \cap \Pi_j(B^\top X = B^\top x)$  is approximatively the same.

For a better view of the previously constructed partition, let us give an example. Let  $X = (X^{(1)}, X^{(2)})^\top$  be a random vector where  $X^{(1)}$  and  $X^{(2)}$  are independent and uniformly distributed on  $[0, 1]$ . In Figure 2, the partition  $\{\Pi_1(B^\top X = B^\top x), \Pi_2(B^\top X = B^\top x)\}$  of  $\text{supp}(X) = [0, 1]^2$  is represented for  $B = (1, 2)^\top / \sqrt{5}$  and  $x = (3/5, 1/5)^\top$ . Using  $n = 500$  independent copies of  $X$  and taking  $H_n = 50/n = 0.1$ , the conditional marginal median  $m(B^\top x)$  is estimated by  $\hat{m}_n(B^\top x) = (0.416, 0.280)^\top$ .

Using the whole sample  $\{X_1, \dots, X_n\}$  to compute the sample analog (16) of  $T(\alpha_n, B)$  is time-consuming since the partition must be computed for each observations  $X_i$ ,  $i = 1, \dots, n$ . We thus choose to use only a subsample of size  $n_0 := \lfloor an \rfloor$  for a given  $a \in (0, 1)$  to compute (16). Extensive simulation studies show that this procedure does not affect too much the quality of the obtained estimator of  $B_0$  (see paragraph 5.3).

We are now in position to compute the estimator  $\hat{B}_{0,n}(\alpha_n)$  *i.e.* to find the matrix  $B \in \mathcal{B}$  minimizing the quantity  $\hat{T}_n(B)$ . To solve this optimization problem, we choose to use the coordinate search method (see for instance [21]). Starting with a matrix  $B_0$ , this method tries to find a new matrix  $B_1$  such that the value of the objective function at  $B_1$  is smaller than the one at  $B_0$ . The new matrix  $B_1$  is obtained by adding or subtracting the search distance  $\delta$  to a single coordinate in each row of  $B_0$ . If such a matrix can be found then  $B_0$  is replaced by  $B_1$ , the search distance  $\delta$  is increased and the previous procedure is repeated. If not, the previous procedure is repeated with  $B_0$  and a smaller value of  $\delta$ . More specifically, the coordinate search method used in this paper is described below:

- Initialization: Take  $B^{(0)} \in \mathcal{B}$ ,  $\delta^{(0)} > 0$  and  $\delta^{\text{tol}} > 0$ .
- Step  $k$ : let  $(e_1, \dots, e_p)$  be the canonical basis of  $\mathbb{R}^p$  and, for  $i = 1, \dots, 2p$ , let  $\tilde{e}_i := (-1)^i e_{\lceil i/2 \rceil}$ . For  $(i_1, \dots, i_q) \in \{1, \dots, 2p\}^q$ , compute the matrix  $\tilde{B}_{(i_1, \dots, i_q)} \in \mathcal{B}$  whose columns are the first  $q$  linearly independent columns of the projection matrix on the set spanned by the columns of the matrix  $B^{(k-1)} + \delta^{(k-1)} \tilde{E}$  with  $\tilde{E} = [\tilde{e}_{i_1}, \dots, \tilde{e}_{i_q}]$ . Denoting by  $\mathcal{B}^{(k)}$  the set of the  $(2p)^q$  matrices  $\tilde{B}_{(i_1, \dots, i_q)}$ ,  $(i_1, \dots, i_q) \in \{1, \dots, 2p\}^q$ , there are two possibilities

- if  $\hat{T}_n(B^{(k-1)}) \leq \min\{\hat{T}_n(B), B \in \mathcal{B}^{(k)}\}$ , then

$$B^{(k)} = B^{(k-1)} \text{ and } \delta^{(k)} = \delta^{(k-1)} / 2,$$

- if  $\hat{T}_n(B^{(k-1)}) > \min\{\hat{T}_n(B), B \in \mathcal{B}^{(k)}\}$ , then

$$B^{(k)} = \arg \min_{B \in \mathcal{B}^{(k)}} \hat{T}_n(B) \text{ and } \delta^{(k)} = 2\delta^{(k-1)}.$$

- If  $\delta^{(k)} > \delta^{\text{tol}}$ , go to step  $k + 1$ , else, the algorithm is stopped.

The procedure of estimation presented here is used in paragraph 5.3 to estimate the TDR subspace. The next paragraph is dedicated to the presentation of the models considered in the simulation study.

## 5.2 Model setting

In what follows, the  $p$  components of the random vector  $X$  are independent and distributed as a Gaussian random variable with mean  $1/2$  and variance  $\sigma^2$ . In this simulation study, the dimension of the explanatory variable  $X$  is fixed to  $p = 4$ . The response  $Y$  is generated according to the following models: for positive functions  $g_0(\cdot)$ ,  $g_1(\cdot)$ ,  $g_2(\cdot)$ , for  $B_0 \in \mathbb{R}^4$  and  $B_1 \in \mathbb{R}^4$ , the conditional quantile of  $Y$  given  $X = x$  is given for  $\alpha \in [0, 1]$  and  $x \in \mathbb{R}^p$  by:

$$\textbf{Model 1} - Q_1(\alpha|X = x) := [\ln(1/(1 - \alpha))]^{-g_0(B_0^\top x)} [1 + g_1(B_1^\top x) \exp(-\alpha^{-1})]^{-1}.$$

$$\textbf{Model 2} - Q_2(\alpha|X = x) := g_2(B_0^\top x) - [\ln(1/(1 - \alpha))]^{g_0(B_0^\top x)} [1 + g_1(B_1^\top x) \exp(-\alpha^{-1})].$$

$$\textbf{Model 3} - Q_3(\alpha|X = x) := [\ln(1/\alpha)]^{g_0(B_0^\top x)} [1 + g_1(B_1^\top x) \exp(-\alpha^{-1})]^{-1}.$$

In the simulation process considered here, we take  $g_0(\cdot) := \tilde{g}(z; 1/3, 8/3)$  and  $g_2(\cdot) := \tilde{g}(z; 1, 10)$  where for  $a < b$  and  $z \in \mathbb{R}$ ,

$$\tilde{g}(z; a, b) = a\mathbb{I}_{(-\infty, 0)}(z) + \left( a + b \frac{\exp(2z) - 1}{\exp(6/\sqrt{5}) - 1} \right) \mathbb{I}_{[0, 3/\sqrt{5})}(z) + (a + b)\mathbb{I}_{[3/\sqrt{5}, \infty)}(z).$$

Note that for all  $z \in \mathbb{R}$ ,  $\tilde{g}(z; a, b) \in [a, a + b]$ . The function  $g_1(\cdot)$  is defined for  $z \in \mathbb{R}$  by  $g_1(z) = \mathbb{I}_{(-\infty, 0)}(z) + \exp(5z)\mathbb{I}_{[0, 2)}(z) + \exp(10)\mathbb{I}_{[2, \infty)}(z)$ . Finally, we take  $B_0^\top := (2, 1, 0, 0)/\sqrt{5}$  and  $B_1^\top := (0, 0, 1, 1)$ . It can be shown (see appendix, Lemma 4) that for all these models,  $\mathcal{S}(B_0)$  is a TDR subspace (and thus  $q = 1$ ) and that condition (9) is satisfied. Obviously, in practice, the dimension  $q$  of the minimum TDR subspace is unknown and need to be estimated. In a non extreme-value context, some approaches for the estimation of  $q$  can be found in the literature (see for instance [6] where a general permutation test is proposed and [5] for the use of information criterion). The adaptation of these methods to an extreme-value context is beyond the scope of the present paper. In what follows, we thus apply our methodology for the estimation of a TDR subspace of dimension 1.

As shown in the Appendix, denoting by  $S_1(y|x)$  (resp.  $S_2(y|x)$ ,  $S_3(y|x)$ ) the conditional survival function  $\mathbb{P}(Y > y|X = x)$  when  $(X, Y)$  is distributed from Model 1 (resp. Model 2, Model 3), one has for all  $x \in \mathbb{R}^p$  and  $\delta > 0$  that

$$\begin{aligned} S_1(\delta^{-1}|x) &= \delta^{1/g_0(B_0^\top x)} L_1(\delta^{-1}|x), \quad S_2(g_2(B_0^\top x) - \delta|x) = S_1(\delta^{-1}|x) \\ \text{and} \quad S_3(\delta^{-1}|x) &= \exp \left[ -\delta^{-1/g_0(B_0^\top x)} L_3(\delta^{-1}|x) \right], \end{aligned}$$

where  $L_1(y|x)$  and  $L_3(y|x)$  converge to 1 as  $y \rightarrow \infty$ . Hence, from [20, Theorem 1.2.1],  $S_1(\cdot|x)$  is in the maximum domain of attraction of Fréchet with positive extreme value index  $\gamma(x) = g_0(B_0^\top x)$  and, as a consequence, condition (5) is satisfied by  $Q_1(\cdot|X = x)$ . Using again [20, Theorem 1.2.1],  $S_2(\cdot|x)$  is in the maximum domain of attraction of Weibull with a negative extreme value index  $\gamma(x) = -g_0(B_0^\top x)$  and a right endpoint  $Q_2(0|X = x) = g_2(B_0^\top x)$ . Finally, for all  $x \in \mathbb{R}^p$ ,  $S_3(\cdot|x)$  is a Weibull-tail distribution (see for instance [17] for a definition). As a consequence,  $S_3(\cdot|x)$  belongs to the Gumbel maximum domain of attraction and thus  $Q_3(\cdot|X = x)$  satisfies (5) with an extreme value index equal to 0.

## 5.3 TDR subspace and conditional quantile estimation

In this paragraph, we are both interested in the performance of the TDR subspace estimator defined in (17) and of the following large quantile estimators of  $Q(\beta_n|X = x)$ : (a) estimator  $\check{Q}_n(\beta_n|B_0, x)$

in the (unrealistic) situation where the TDR direction  $B_0$  is known, (b) the more useful plug-in estimator  $\check{Q}_n(\beta_n|\hat{B}_{0,n}(\alpha_n), x)$  and (c) the classical estimator  $\check{Q}_n(\beta_n|x) := \check{Q}_n(\beta_n|I_p, x)$  when the existence of a TDR subspace is not taken into account.

To this aim, for each model introduced in paragraph 5.2,  $N = 100$  samples of size  $n = 4000$  are generated and the conditional quantile order is taken as  $\beta_n = 2/n$ .

To compute the TDR subspace estimator, we take  $\alpha_n = n^{-1/3}$  and a subsample of size  $n_0 = 100$ . To appreciate the influence of the size  $n_0$ , the estimation of  $B_0$  has been considered under model 1 with  $\sigma = 1/5$  for  $n_0 \in \{2, 4, \dots, 20, 25, 50, 75, 100, 150, 200, 250, 300\}$ . In figure 3, the empirical quantiles of order 0.05, 0.5 and 0.95 of  $\|\hat{B}_{0,n}(\alpha_n) - B_0\|_F$  (where for all matrix  $B$ ,  $\|B\|_F := [\text{tr}(B^\top B)]^{1/2}$  is the Frobenius norm) are represented as a function of  $n_0$ . It appears that for  $n_0 > 75$ , the quality of the estimation does not clearly depends on  $n_0$ . Recall also that an initialization is required in the coordinate search method used to compute the TDR subspace estimator. In this simulation study, the initial vector  $B^{(0)}$  is randomly chosen. More specifically,  $B^{(0)} = (u_1, \dots, u_p)/\sqrt{u_1^2 + \dots + u_p^2}$  where  $u_1, \dots, u_p$  are realizations of  $p$  independent random variables uniformly distributed on  $[-1, 1]$ . The estimator  $\hat{B}_{0,n}(\alpha_n)$  of  $B_0$  is compared with the one obtained by the SIR method. Recall that the SIR estimator, denoted  $\hat{B}_{0,n}^{\text{SIR}}$ , corresponds to the  $q$  eigen vectors associated to the  $q$  largest eigen values of the matrix  $\hat{\Sigma}_n^{-1}\hat{\Gamma}_n$  with

$$\begin{aligned}\hat{\Sigma}_n &:= \frac{1}{n} \sum_{i=1}^n \left( X_i - \frac{1}{n} \sum_{i=1}^n X_i \right) \left( X_i - \frac{1}{n} \sum_{i=1}^n X_i \right)^\top \\ \text{and } \hat{\Gamma}_n &:= \sum_{j=1}^J \frac{n_j}{n} \left( \frac{1}{n_j} \sum_{i: Y_i \in R_j} X_i - \frac{1}{n} \sum_{i=1}^n X_i \right) \left( \frac{1}{n_j} \sum_{i: Y_i \in R_j} X_i - \frac{1}{n} \sum_{i=1}^n X_i \right)^\top,\end{aligned}$$

where  $\{R_1, \dots, R_J\}$  are non-overlapping slices that cover the range of  $Y$  and  $n_j$  is the number of  $Y_i$ 's lying in  $R_j$ . To compute the SIR estimator, the R package `dr` (see [31]) was used with the default method to construct the slices. The SIR method focus only on the central part of the conditional distribution of  $Y$  given  $X$  and is thus not adapted for the estimation of the TDR subspace. The accuracy of the TDR direction estimators is measured by the Frobenius norm of their difference with the true direction  $B_0$ .

Concerning the conditional quantile estimators, we use the following hyperparameters:  $\alpha_n = n^{-1/3}$  and  $H_n = n^{-2/9}$ . To compute the conditional tail index estimator (7) and the auxiliary function estimator (8) we take  $\nu = 0.02$  and  $\varphi(\cdot) = \ln(1/\cdot)$  (this choice was suggested in [14]). In order to measure the quality of the estimations, we compute for each replication the error given by

$$E_Q(\check{B}) := \frac{1}{\text{card}(\mathcal{L})} \sum_{x \in \mathcal{L}} \left( \ln \frac{\check{Q}(\beta_n|\check{B}, x)}{\check{Q}(\beta_n|X = x)} \right)^2,$$

where  $\mathcal{L}$  is the set  $\{(x_1, \dots, x_4); x_i \in \{1/4, 1/2, 3/4\} \text{ for } i = 1, \dots, 4\}$  with  $\text{card}(\mathcal{L}) = 3^4 = 81$  and  $\check{B} \in \{\hat{B}_{0,n}, \hat{B}_{0,n}^{\text{SIR}}, I_p\}$ . The results are gathered in Table 1. As expected, the extreme conditional quantile estimator  $\check{Q}_n(\beta_n|B_0, \cdot)$  is the best among the three statistics considered. Unfortunately, this estimator can only be used in the ideal situation where  $B_0$  is known. Nevertheless, the plug-in estimator  $\check{Q}_n(\beta_n|\hat{B}_{0,n}, \cdot)$  provides similar results and outperforms in each case the classical estimator  $\check{Q}_n(\beta_n|I_p, \cdot)$  (for which the existence of a TDR subspace is not assumed). Note also that the TDR estimator  $\hat{B}_{0,n}$  provides good estimation of the true direction  $B_0$  while, as expected, the

SIR estimator  $\hat{B}_{0,n}^{\text{SIR}}$  is not able to find the true direction. It appears in fact that  $\hat{B}_{0,n}^{\text{SIR}}$  is often close to the “central” direction  $B_1 = (0, 0, 1, 1)^\top$ . For instance, when  $\sigma = 1/3$ ,  $\|\hat{B}_{0,n}^{\text{SIR}} - B_1\|_F = 0.016$  for Model 1,  $\|\hat{B}_{0,n}^{\text{SIR}} - B_1\|_F = 0.066$  for Model 2 and  $\|\hat{B}_{0,n}^{\text{SIR}} - B_1\|_F = 0.110$  for Model 3. This confirms the fact that classical dimension reduction methods are not adapted to the study of the tail of the conditional distribution. Finally, note that the value of the standard deviation  $\sigma$  has an influence on the estimation of  $B_0$ . Except for Model 2, the performance of  $\hat{B}_{0,n}$  deteriorates as  $\sigma$  decreases. One possible explanation is that when  $\sigma$  is small, the variance of  $\tilde{\gamma}(B_0^\top X)$  is also small is thus the true direction is difficult to capture.

## 6 Data analysis

According to the world health organization, atmospheric pollutants may cause serious effects on public health and on environment. One of the most dangerous pollutant is ozone. It is generated in the air when others pollutants (called primary pollutants) react with atmospheric oxygen and weather conditions (temperature, humidity, etc ...).

Our goal is to study the effect of some primary pollutants (namely particular matter with diameter smaller than 10 microns (PM<sub>10</sub>), sulphur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>) and carbon monoxide (CO)) on the daily maximum surface concentration of ozone in parts per billion ( $Y$ ). We considered daily data collected in Chicago from 1987 to 2000. These data are available at <http://www.ihapss.jhsph.edu/data/data.htm> and were considered by many authors for illustrate dimension reduction methods (see for instance [33] and [28]). The number of observations is  $n = 4841$ . To compute the TDR estimator proposed in this paper we take  $h_n = 0.15$ ,  $\alpha_n = 0.05$  and  $n_0 = 100$ . Note that the coordinate search method was initialized with different vectors  $B^{(0)}$ , all these initialization leading to very similar values of  $\hat{B}_{0,n}(\alpha_n)$ .

First, the components of the covariate vector  $X$  are the daily **maximum** values of the primary pollutants PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub> and CO. For numerical convenience, the predictors values are centered and normalized. We thus have  $p = 4$  and, as in [33], we assume the existence of a TDR subspace of dimension  $q = 1$ . The obtained estimator of the TDR direction is given by:  $\hat{B}_{0,n}(\alpha_n) = (0.187, -0.045, 0.965, -0.179)^\top$ . The direction obtained with the SIR method is  $\hat{B}_{0,n}^{\text{SIR}}(\alpha_n) = (0.327, -0.085, 0.910, -0.238)^\top$ . It appears that the SIR estimator is closed from our TDR estimator ( $\|\hat{B}_{0,n}(\alpha_n) - \hat{B}_{0,n}^{\text{SIR}}(\alpha_n)\|_F = 0.027$ ). Hence, the influence of the daily maximum values of the 4 primary pollutants on the ozone concentration seems to be the same in the central part of the distribution and in the tail distribution of  $Y$ . As mentioned in [28], the level of ozone concentration is mainly determined by NO<sub>2</sub>.

Next, we look at the influence of the daily **median** values of the primary pollutants on ozone concentration. In this case, our TDR estimator is  $\hat{B}_{0,n}(\alpha_n) = (0.957, -0.051, 0.130, -0.254)^\top$ , while the SIR estimator is given by  $\hat{B}_{0,n}^{\text{SIR}}(\alpha_n) = (0.747, -0.254, 0.441, -0.429)^\top$ . In this situation, the directions estimated by SIR and by our proposed method are quite different ( $\|\hat{B}_{0,n}(\alpha_n) - \hat{B}_{0,n}^{\text{SIR}}(\alpha_n)\|_F = 0.212$ ). A possible interpretation is that large values of ozone concentration are essentially associated with large values of daily median values of PM<sub>10</sub> while an increase in mean values of  $Y$  is associated with increasing values of PM<sub>10</sub> and NO<sub>2</sub>. In Figure 4, the conditional quantile estimators  $\check{Q}(1/n|\hat{B}_{0,n}(\alpha_n), x_\tau)$  and  $\check{Q}(1/n|\hat{B}_{0,n}^{\text{SIR}}(\alpha_n), x_\tau)$  are represented as a function of  $\tau \in [0.4, 0.65]$ . Here  $x_\tau := (x_\tau^{\text{PM}_{10}}, x_\tau^{\text{SO}_2}, x_\tau^{\text{NO}_2}, x_\tau^{\text{CO}})^\top$  where  $x_\tau^{\text{PM}_{10}}$  is the empirical quantile of

order  $1 - \tau$  of the daily median values of  $\text{PM}_{10}$  and so on ... It appears that for  $\tau$  around 0.5, the estimators  $\check{Q}(1/n|\hat{B}_{0,n}(\alpha_n), x_\tau)$  and  $\check{Q}(1/n|\hat{B}_{0,n}^{\text{SIR}}(\alpha_n), x_\tau)$  are quite close but differ strongly for small values of  $\tau$  (corresponding to large values of the covariate  $\text{PM}_{10}$ ).

## 7 Proofs

### 7.1 Preliminary results

The first Lemma is a probability result that will be helpful in the proof Theorem 1.

**Lemma 1.** *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. Let  $U$  be a positive random variable and  $V$  be a  $\mathbb{R}^q$ -valued random vector ( $U$  and  $V$  are both defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ ). Denoting by  $\sigma(U)$  (resp.  $\sigma(V)$ ) the  $\sigma$ -algebra generated by  $U$  (resp.  $V$ ), it is assumed that  $\sigma(V) \subset \sigma(U)$ . If there exists  $\theta \geq 1$  such that for all  $F \in \sigma(U)$ ,*

$$\int_F \mathbb{E}(U|V)d\mathbb{P} \leq \theta \int_F U d\mathbb{P} \quad \left( \text{or } \int_F \mathbb{E}(U|V)d\mathbb{P} \geq \theta^{-1} \int_F U d\mathbb{P} \right), \quad (18)$$

*then  $\mathbb{E}(U|V) \leq \theta U$  a.s. (or  $\mathbb{E}(U|V) \geq \theta^{-1}U$  a.s.).*

**Proof** – Assume that for all  $F \in \sigma(U)$

$$\int_F \mathbb{E}(U|V)d\mathbb{P} \leq \theta \int_F U d\mathbb{P},$$

(the proof for the other case is similar). First, we suppose that  $U$  is a positive simple function. More specifically, let  $\{A_1, \dots, A_k\}$  be  $k$  disjoint elements of  $\mathcal{F}$  with  $0 < \mathbb{P}(A_i) < 1$  for all  $i \in \{1, \dots, k\}$ . We assume that

$$U = \sum_{i=1}^k c_i \mathbb{I}_{A_i} + c_{k+1} \mathbb{I}_{(A_1 \cup \dots \cup A_k)^c},$$

where for all  $A \subset \Omega$ ,  $A^c$  is the complement of  $A$  in  $\Omega$  and  $c_i > 0$  for all  $i \in \{1, \dots, k\}$ . Since  $\sigma(V) \subset \sigma(U)$ , one can assume without loss of generality that  $\sigma(V) = \sigma(\{A_1, \dots, A_\ell\})$  with  $1 \leq \ell < k$  (the situation where  $\ell = k$ , i.e.  $\sigma(U) = \sigma(V)$  is trivial). It is then easy to check that

$$\mathbb{E}(U|V) = \sum_{i=1}^{\ell} c_i \mathbb{I}_{A_i} + \xi \mathbb{I}_{(A_1 \cup \dots \cup A_\ell)^c},$$

where

$$\xi := \sum_{i=\ell+1}^k c_i \frac{\mathbb{P}(A_i)}{\mathbb{P}[(A_1 \cup \dots \cup A_\ell)^c]} + c_{k+1} \frac{\mathbb{P}[(A_1 \cup \dots \cup A_k)^c]}{\mathbb{P}[(A_1 \cup \dots \cup A_\ell)^c]}.$$

We thus have to show that for every  $i \in \{\ell+1, \dots, k\}$ ,  $\xi \leq \theta c_i$  and that, if  $\mathbb{P}[(A_1 \cup \dots \cup A_k)^c] \neq 0$ ,  $\xi \leq \theta c_{k+1}$ . From (18), for all  $F \in \sigma(U) = \sigma(\{A_1, \dots, A_k\})$ ,

$$\sum_{i=1}^{\ell} c_i \mathbb{P}(A_i \cap F) + \xi \mathbb{P}[(A_1 \cup \dots \cup A_\ell)^c \cap F] \leq \theta \sum_{i=1}^k c_i \mathbb{P}(A_i \cap F) + c_{k+1} \mathbb{P}[(A_1 \cup \dots \cup A_k)^c \cap F].$$

For  $j \in \{\ell+1, \dots, k\}$ , taking  $F = A_j$  in the previous inequality leads to  $\xi \leq \theta c_j$  since  $\mathbb{P}(A_j) > 0$ . Furthermore, taking  $F = (A_1 \cup \dots \cup A_k)^c$  entails that  $\xi \mathbb{P}[(A_1 \cup \dots \cup A_k)^c] \leq \theta \mathbb{P}[(A_1 \cup \dots \cup A_k)^c]$ . The result is thus proved for all positive simple functions. Since any positive measurable function is the pointwise limit of an increasing sequence of positive simple function, we conclude the proof by using the Lebesgue's monotone convergence theorem.  $\blacksquare$

The next lemma is a technical result used in the proof of Theorem 4.

**Lemma 2.** *Assume that there exists a full rank  $p \times q$  matrix  $B_0$  such that  $\mathcal{S}(B_0)$  is a TDR subspace for  $Y$  given  $X$  and that condition (5) holds. For  $x \in \text{supp}(X)$ , let  $(y_n(u|x))$  be a sequence such that  $[y_n(u|x) - Q(u\alpha_n|X = x)]/a(\alpha_n^{-1}|X = x) \rightarrow 0$  locally uniformly on  $u \in (0, \infty)$ . The following statements hold almost everywhere for  $x \in \text{supp}(X)$ :*

- (i) *Let  $\nu \in (0, 1)$ . For all  $u \in [\nu, 1]$ ,  $y_n(u|x) = Q(\beta|B_0^\top X = B_0^\top x)$  with  $\beta \in [\xi\nu\alpha_n, \xi^{-1}\alpha_n]$  for  $\xi < 1$  as near as you like to 1.*
- (ii) *Let  $u \in (0, 1)$  and  $\xi < 1$  as near as we like to 1. If  $Q(0|X = x) < \infty$  then  $Q(0|X = x) - y_n(u|x) \leq \xi^{-1}\delta_{\alpha_n}(X = x)$  and if  $Q(0|X = x) = +\infty$  then  $[y_n(u|x)]^{-1} \leq \xi^{-1}\delta_{\alpha_n}(X = x)$ .*

**Proof** – We start by proving the first statement. From (5), one has for  $x \in \text{supp}(X)$  that  $[y_n(u|x) - Q(\alpha_n|X = x)]/a(\alpha_n^{-1}|X = x) - L_{\gamma(x)}(1/u) \rightarrow 0$  locally uniformly on  $u \in (0, \infty)$ . Now, since  $\mathcal{S}(B_0)$  is a TDR subspace, [20, Lemma 1.2.12] entails that  $[Q(\alpha_n|X = x) - Q(\alpha_n|B_0^\top X = B_0^\top x)]/a(\alpha_n^{-1}|X = x)$  converges to 0 almost everywhere for  $x \in \text{supp}(X)$ . As a first conclusion,

$$y_n(u|x) = Q(\alpha_n|B_0^\top X = B_0^\top x) + a(\alpha_n^{-1}|x) [L_{\gamma(x)}(1/u) + o(1)],$$

locally uniformly on  $u \in (0, \infty)$  and almost everywhere for  $x \in \text{supp}(X)$ . From Lemma 1, the distribution function of  $Y$  given  $B_0^\top X$  belongs to a maximum domain of attraction with an auxiliary function equivalent to  $a(\cdot|x)$ . Thus, according to [20, Theorem 1.1.6], one can find  $\xi < 1$  as close as we like to 1 such that for all  $u \in [\nu, 1]$ ,  $\mathbb{P}(Y > y_n(u|x)|B_0^\top X = B_0^\top x) \geq \nu\xi\alpha_n$ . Hence,

$$y_n(u|x) = Q(\mathbb{P}(Y > y_n(u|x)|B_0^\top X = B_0^\top x)|B_0^\top X = B_0^\top x) \leq Q(\nu\xi\alpha_n|B_0^\top X = B_0^\top x). \quad (19)$$

Mimicking the proof of (19), we show that  $y_n(u|x) \geq Q(\xi^{-1}\alpha_n|B_0^\top X = B_0^\top x)$  and thus conclude the proof of the first statement.

Let us now focus on the second statement. Assume first that  $Q(0|X = x) < \infty$  (this implies that  $\gamma(x) \leq 0$ ). Using the first statement with  $B_0 = I_p$ , one has for all  $u \leq 1$  and for  $\tilde{\xi}$  as near as we like to 1 that  $Q(0|X = x) - y_n(u|x) \leq Q(0|X = x) - Q(\tilde{\xi}^{-1}\alpha_n|X = x)$ . Now from [20, Lemma 1.2.9], one has that  $\delta_{\alpha_n}^{-1}(X = x)a(\alpha_n^{-1}|X = x) \rightarrow -\gamma(x)$  as  $n$  goes to infinity. Hence, from (5), one can find  $\xi < 1$  as near as we want to 1 such that

$$\frac{Q(0|X = x) - y_n(u|x)}{\delta_{\alpha_n}(X = x)} \leq 1 + \frac{Q(\alpha_n|X = x) - Q(\tilde{\xi}^{-1}\alpha_n|X = x)}{\delta_{\alpha_n}(X = x)} \leq \xi^{-1}.$$

Now, suppose that  $Q(0|X = x) = +\infty$  (and thus that  $\gamma(x) \geq 0$ ). Applying again the first statement with  $B_0 = I_p$ , one has for all  $u \leq 1$  that  $y_n(u|x) \geq Q(\tilde{\xi}^{-1}\alpha_n|X = x)$  where  $\tilde{\xi} < 1$  is as close as we want to 1. Since  $\delta_{\alpha_n}(X = x)a(\alpha_n^{-1}|X = x) = a(\alpha_n^{-1}|X = x)/Q(\alpha_n|X = x) \rightarrow \gamma(x)$  as  $n$  goes to infinity, condition (5) entails that  $\delta_{\alpha_n}(X = x)Q(\tilde{\xi}^{-1}\alpha_n|X = x) = 1 + \delta_{\alpha_n}(X = x)[Q(\tilde{\xi}^{-1}\alpha_n|X = x) - Q(\alpha_n|X = x)] \rightarrow \tilde{\xi}^{\gamma(x)}$ . Hence, one can find  $\xi < 1$  as close as we want to 1 and such that for  $n$  large enough,  $y_n(u|x) \geq \xi\delta_{\alpha_n}^{-1}(X = x)$  and the proof is complete. ■

## 7.2 Proofs of main results

**Proof of Theorem 1** – First, we prove that (i) implies (ii). Let  $\varepsilon > 0$ . There exists  $\kappa > 0$  such that for all  $\delta \in (0, \kappa]$ ,

$$\Delta_\delta(X, Z) := \left| \frac{\mathbb{P}(Y > \mathcal{Y}_\delta(Z)|X, Z)}{\mathbb{P}(Y > \mathcal{Y}_\delta(Z)|Z)} - 1 \right| \leq \varepsilon.$$

Remarking that for all non-zero bounded and positive function  $h(\cdot)$ ,

$$\mathbb{E}(\mathbb{I}_{\{Y > \mathcal{Y}_\delta(Z)\}} h(X) | Z) = \mathbb{E}[h(X) \mathbb{E}(\mathbb{I}_{\{Y > \mathcal{Y}_\delta(Z)\}} | X, Z) | Z].$$

it is easy to check that, almost surely,

$$\left| \frac{\mathbb{E}(\mathbb{I}_{\{Y > \mathcal{Y}_\delta(Z)\}} h(X) | Z)}{\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z) \mathbb{E}(h(X) | Z)} - 1 \right| \leq \mathbb{E} \left[ \left| \frac{\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z)}{\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z)} - 1 \right| \frac{h(X)}{\mathbb{E}(h(X) | Z)} | Z \right] \leq \varepsilon. \quad (20)$$

Let us now show that (ii) implies (i). We thus assume that for all  $\varepsilon > 0$ , there exist  $\kappa > 0$  such that for all  $\delta \in (0, \kappa]$ , inequality (20) holds. Let  $A \in \mathcal{B}(\mathbb{R}^p)$  and  $B \in \mathcal{B}(\mathbb{R}^q)$ , we have

$$\begin{aligned} & \int_{\{X \in A\} \cap \{Z \in B\}} \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z) d\mathbb{P} = \mathbb{E} \{ \mathbb{E}[\mathbb{I}_{\{X \in A\}} \mathbb{E}(\mathbb{I}_{\{Y > \mathcal{Y}_\delta(Z)\}} | Z) | Z] \} \\ &= \int_{\{Z \in B\}} \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z) \mathbb{P}(X \in A | Z) d\mathbb{P} \\ &= \int_{\{Z \in B\}} \mathbb{P}[\{Y > \mathcal{Y}_\delta(Z)\} \cap \{X \in A\} | Z] \frac{\mathbb{P}(X \in A | Z) \mathbb{P}(X \in A | Z)}{\mathbb{P}[\{Y > \mathcal{Y}_\delta(Z)\} \cap \{X \in A\} | Z]} d\mathbb{P}. \end{aligned}$$

Let  $\varepsilon \in (0, 1)$ . Applying inequality (20) with  $h(X) = \mathbb{I}_A(X)$  leads to

$$\int_{\{X \in A\} \cap \{Z \in B\}} \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z) d\mathbb{P} \leq \frac{1}{1 - \varepsilon} \int_{\{Z \in B\}} \mathbb{P}[\{Y > \mathcal{Y}_\delta(Z)\} \cap \{X \in A\} | Z] d\mathbb{P}$$

and

$$\int_{\{X \in A\} \cap \{Z \in B\}} \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z) d\mathbb{P} \geq \frac{1}{1 + \varepsilon} \int_{\{Z \in B\}} \mathbb{P}[\{Y > \mathcal{Y}_\delta(Z)\} \cap \{X \in A\} | Z] d\mathbb{P}.$$

Since

$$\int_{\{Z \in B\}} \mathbb{P}[\{Y > \mathcal{Y}_\delta(Z)\} \cap \{X \in A\} | Z] d\mathbb{P} = \int_{\{X \in A\} \cap \{Z \in B\}} \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z) d\mathbb{P},$$

and by the monotone class theorem we thus have for all  $F \in \sigma(X, Z)$  that

$$\frac{1}{1 + \varepsilon} \int_F \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z) d\mathbb{P} \leq \int_F \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z) d\mathbb{P} \leq \frac{1}{1 - \varepsilon} \int_F \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z) d\mathbb{P}.$$

We conclude the proof by applying Lemma 1 with  $U := \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z)$  and  $V = Z$ .

Let us now prove that (i) is equivalent to (iii). Obviously, if (i) holds then (iii) also holds with  $s_\delta(Z) = \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z)$  and  $\eta_\delta(X, Z) = \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z) / \mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z) - 1$ .

To show that (iii) implies (i), first remark that the “tower property” of conditional expectations entails that  $\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z) = \mathbb{E}[\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z) | Z] = s_\delta(Z)[1 + \mathbb{E}(\eta_\delta(X, Z) | Z)]$  a.s. Hence

$$\frac{\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z)}{\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z)} = \frac{1 + \eta_\delta(X, Z)}{1 + \mathbb{E}(\eta_\delta(X, Z) | Z)} \text{ a.s.} \quad (21)$$

By assumption, for all  $\varepsilon \in (0, 1)$ , there exists  $\kappa > 0$  such that for all  $\delta \in (0, \kappa]$ ,

$$\mathbb{P} \left[ \left| \frac{\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | X, Z)}{\mathbb{P}(Y > \mathcal{Y}_\delta(Z) | Z)} - 1 \right| \leq \frac{2\varepsilon}{1 - \varepsilon} \right] = 1, \quad (22)$$

which concludes the proof. ■

**Proof of Proposition 1** – First recall that since  $\mathcal{S}(B_0)$  is a TDR subspace, there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^p)$  with  $\mathbb{P}(X \in \mathcal{A}) = 1$  such that for all  $x \in \mathcal{A}$ ,  $Q(0|X = x) = Q(0|B_0^\top X = B_0^\top x)$ . Hence, for all  $x \in \mathcal{A}$ , [20, Lemma 1.2.12] and the definition of the TDR subspace entail that the distribution function  $\mathbb{P}(Y \leq \cdot | B_0^\top X = B_0^\top x)$  belongs to the maximum domain of attraction of an extreme value distribution with extreme value index  $\gamma(x)$  or equivalently that for all  $u > 0$  and  $x \in \mathcal{A}$

$$\lim_{\alpha \rightarrow 0} \frac{Q(u\alpha|B_0^\top X = B_0^\top x) - Q(\alpha|B_0^\top X = B_0^\top x)}{a(\alpha^{-1}|x)} = L_{\gamma(x)}(1/u). \quad (23)$$

As a consequence, from [20, Theorem 1.2.6], for all  $x \in \mathcal{A}$  there exist positive functions  $c(\cdot|B_0^\top x)$  and  $d(\cdot|B_0^\top x)$  (depending on  $x$  only through  $B_0^\top x$ ) such that for all  $y \in (y_0, Q(0|B_0^\top X = B_0^\top x))$ ,

$$\mathbb{P}(Y > y | B_0^\top X = B_0^\top x) = c(y|B_0^\top x) \exp \left\{ - \int_{y_0}^y \frac{ds}{d(s|B_0^\top x)} \right\},$$

with  $c(y|B_0^\top x) \rightarrow c_0 > 0$  as  $y \rightarrow Q(0|B_0^\top X = B_0^\top x)$ . Now, let us introduce the differentiable function

$$S_0(y|B_0^\top x) := \min \left( 1, c_0 \exp \left\{ - \int_{y_0}^y \frac{ds}{d(s|B_0^\top x)} \right\} \right).$$

According to Remarks 1.2.7 and 1.2.8 in [20], the survival function  $S_0(\cdot|B_0^\top x)$  belongs to the maximum domain of attraction of an extreme value distribution with extreme value index  $\gamma(x)$ . More precisely, taking  $\tilde{a}(\alpha^{-1}|B_0^\top x) := d_0(S_0^{\leftarrow}(\alpha|B_0^\top x))$  where  $S_0^{\leftarrow}(\cdot|B_0^\top x)$  is the generalized inverse of  $S_0(\cdot|B_0^\top x)$  (recall that for any non-increasing  $\Phi(\cdot)$ , its generalized inverse is given by  $\Phi^{\leftarrow}(\cdot) = \inf\{y; \Phi(y) \leq \cdot\}$ ) and where  $d_0(\cdot) := -S_0(\cdot|B_0^\top x)/S_0'(\cdot|B_0^\top x)$ , one has for all  $u > 0$  and  $x \in \mathcal{A}$

$$\lim_{\alpha \rightarrow 0} \frac{S_0^{\leftarrow}(u\alpha|B_0^\top x) - S_0^{\leftarrow}(\alpha|B_0^\top x)}{\tilde{a}(\alpha^{-1}|B_0^\top x)} = L_{\gamma(x)}(1/u).$$

Since the left hand side in the previous limit only depends on  $x$  through  $B_0^\top x$ , same holds for the right hand side. Hence there exists a real-valued function  $\tilde{\gamma}(\cdot)$  such that  $\gamma(x) = \tilde{\gamma}(B_0^\top x)$  for all  $x \in \mathcal{A}$ . Finally, since  $\mathbb{P}(Y > y | B_0^\top X = B_0^\top x) / S_0(y|B_0^\top x) \rightarrow 1$  as  $y \rightarrow Q(0|B_0^\top X = B_0^\top x)$ , using again [20, Lemma 1.2.12] entails that for all  $u > 0$  and  $x \in \mathcal{A}$

$$\lim_{\alpha \rightarrow 0} \frac{Q(u\alpha|B_0^\top X = B_0^\top x) - Q(\alpha|B_0^\top X = B_0^\top x)}{\tilde{a}(\alpha^{-1}|B_0^\top x)} = L_{\tilde{\gamma}(B_0^\top x)}(1/u). \quad (24)$$

We conclude the proof by remarking that (23) and (24) lead to  $\tilde{a}(\alpha^{-1}|B_0^\top x)/a(\alpha^{-1}|x) \rightarrow 1$  as  $\alpha$  goes to 0.  $\blacksquare$

To prove Theorem 2, the following intermediate result is required. It deals with the asymptotic properties of the conditional survival function estimator given in (3).

**Proposition 4.** *Assume that there exists a full rank matrix  $B_0$  such that  $\mathcal{S}(B_0)$  is a TDR subspace and suppose conditions (9) and (10) hold. Let  $(\alpha_n)$  and  $(H_n)$  be two sequences such that  $\alpha_n \rightarrow 0$ ,  $n|H_n|\alpha_n \rightarrow \infty$  and  $\tau_n^{-1}\|H_n\|_\infty \rightarrow 0$  with  $\tau_n := (n|H_n|\alpha_n)^{-1/2}[\ln(n|H_n|\alpha_n)]^{1/2}$ . If there exists  $\xi \in (0, 1)$  such that for all  $x \in \text{supp}(X)$ ,*

$$\max \left\{ \sup_{(t, \beta) \in A_n} \left| \frac{\mathbb{P}(Y > Q(\beta|B_0^\top X = B_0^\top x) | B_0^\top X = t)}{\beta} - 1 \right| ; \bar{\eta}(\xi \delta_{\alpha_n}^{-1}(X = x)) \right\} = \mathcal{O}(\tau_n), \quad (25)$$

where  $A_n := D(B_0^\top x, H_n) \times [\xi \nu \alpha_n, \xi^{-1} \alpha_n]$ , and if  $\tau_n^{-1} \text{ERV}(\alpha_n, u | X = x) \rightarrow 0$  locally uniformly on  $u \in (0, \infty)$  then, for all  $\nu \in (0, 1)$ , there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^p)$  with  $\mathbb{P}(X \in \mathcal{A}) = 1$  such

that for all  $x \in \mathcal{A}$  and for all sequences  $(y_n(u|x))$  such that  $[y_n(u|x) - Q(u\alpha_n|X = x)]/a(\alpha_n^{-1}|X = x) \rightarrow 0$  locally uniformly on  $u \in (0, \infty)$ .

$$\sup_{u \in [\nu, 1]} \left| \frac{\widehat{S}_n(y_n(u|x)|B_0, x)}{\mathbb{P}(Y > y_n(u|x)|X = x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(\tau_n).$$

Before proving this result, remark that from [14, Lemma 5], Proposition 4 entails that

$$\frac{Q(\alpha_n|X = x)}{a(\alpha_n^{-1}|X = x)} \sup_{u \in [\nu, 1]} \left| \frac{\widehat{Q}_n(u\alpha_n|B_0, x)}{Q(u\alpha_n|X = x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(\tau_n), \quad (26)$$

Thus  $\widehat{Q}_n(\cdot|B_0, x)$  is a consistent estimator of  $Q(\cdot|X = x)$  but under a restrictive condition on the order  $\alpha_n$  (namely,  $n|H_n|\alpha_n \rightarrow \infty$ ).

**Proof of Proposition 4** – First remark that  $\widehat{S}_n(y_n(u|x)|B_0, x) = \widehat{\Phi}_n(u|x)/[n|H_n|\widehat{f}_{n, B_0^\top X}(B_0^\top x)]$ , where

$$\begin{aligned} \widehat{f}_{n, B_0^\top X}(B_0^\top x) &:= \frac{1}{n|H_n|} \sum_{i=1}^n K(H_n^{-1}B_0^\top(x - X_i)) \\ \text{and } \widehat{\Phi}_n(u|x) &:= \sum_{i=1}^n K(H_n^{-1}B_0^\top(x - X_i)) \mathbb{I}_{\{Y_i > y_n(u|x)\}}. \end{aligned}$$

Note that  $\widehat{f}_{n, B_0^\top X}(\cdot)$  is the classical kernel estimator of the probability distribution function  $f_{B_0^\top X}(\cdot)$ . Under (10) and since  $n|H_n| \rightarrow \infty$ , it can be shown (see for instance Parzen [27]) that

$$\frac{\widehat{f}_{n, B_0^\top X}(B_0^\top x)}{f_{B_0^\top X}(B_0^\top x)} - 1 = \mathcal{O}_{\mathbb{P}}(\|H_n\|_\infty) + \mathcal{O}_{\mathbb{P}}\left((n|H_n|)^{-1/2}\right) = o_{\mathbb{P}}(\tau_n). \quad (27)$$

Let us now focus on the statistics  $\widehat{\Phi}_n(u|x)$ . According to [14, Lemma 6], letting  $\mu_n(u|x) := \mathbb{E}(\widehat{\Phi}_n(u|x))$  and  $v_{n,x}^2 := \ln(\mu_n(1|x))/\mu_n(1|x)$ , if the following conditions are satisfied:

**(C.1)** There exists a positive constant  $C_X$  such that  $K(H_n^{-1}B_0^\top(x - X_i)) \mathbb{I}_{\{Y_i > y_n(u|x)\}} \leq C_X$  for  $n$  large enough and all  $u \in [\nu, 1]$ ,

**(C.2)**  $\mu_n(\nu|x) \rightarrow \infty$  as  $n \rightarrow \infty$  and there exists a positive constant  $C_\mu$  such that for  $n$  large enough  $\mu_n(\nu|x)/\mu_n(1|x) \geq C_\mu$ ,

**(C.3)**

$$\sup \left\{ \left| \frac{\mu_n(u|x)}{\mu_n(u'|x)} - 1 \right|, u \in [\nu, 1] \text{ with } |u - u'| \leq (\mu_n(1|x))^{-1/2} \right\} = \mathcal{O}(v_{n,x}),$$

then

$$\sup_{u \in [\nu, 1]} \left| \frac{\widehat{\Phi}_n(u|x)}{\mu_n(u|x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(v_{n,x}). \quad (28)$$

Since  $K(\cdot)$  is bounded, it is clear that **(C.1)** holds with  $C_X = \|K\|_\infty$ . To check **(C.2)** and **(C.3)**, the first step is the computation of  $\mu_n(u|x)$ . Since  $X$  is absolutely continuous with respect to the Lebesgue measure and since  $B_0$  is a full rank matrix, the random vector  $B_0^\top X$  is also absolutely continuous with support  $\text{supp}(B_0^\top X) = \{B_0^\top x, x \in \text{supp}(X)\}$ . Hence,

$$\mu_n(u|x) = n \int_{\text{supp}(B_0^\top X)} K(H_n^{-1}(B_0^\top x - t)) \mathbb{P}(Y > y_n(u|x)|B_0^\top X = t) f_{B_0^\top X}(t) dt.$$

Letting  $s = H_n^{-1}(B_0^\top x - t)$  leads to

$$\mu_n(u|x) = n|H_n| \int_{\mathcal{U}_q} K(s) \mathbb{P}(Y > y_n(u|x) | B_0^\top X = B_0^\top x - H_n s) f_{B_0^\top X}(B_0^\top x - H_n s) ds,$$

since, for  $n$  large enough,  $\mathcal{U}_q \subset \{H_n^{-1}(B_0^\top x - t), t \in \text{supp}(B_0^\top X)\}$  where  $\mathcal{U}_q$  is the unit ball of  $\mathbb{R}^q$ . To sum up, since  $f_{B_0^\top X}(B_0^\top x) > 0$ ,

$$\begin{aligned} & \frac{\mu_n(u|x)}{n|H_n|f_{B_0^\top X}(B_0^\top x)\mathbb{P}(Y > y_n(u|x) | B_0^\top X = B_0^\top x)} \\ &= \int_{\mathcal{U}_q} K(s) \frac{\mathbb{P}(Y > y_n(u|x) | B_0^\top X = B_0^\top x - H_n s)}{\mathbb{P}(Y > y_n(u|x) | B_0^\top X = B_0^\top x)} \frac{f_{B_0^\top X}(B_0^\top x - H_n s)}{f_{B_0^\top X}(B_0^\top x)} ds. \end{aligned}$$

Remarking that  $B_0^\top x - H_n s = B_0^\top [x - B_0(B_0^\top B_0)^{-1}H_n s] \in \text{supp}(B_0^\top X)$  since  $\text{supp}(X)$  is an open set, condition (10) entails that  $f_{B_0^\top X}(B_0^\top x - H_n s)/f_{B_0^\top X}(B_0^\top x) - 1 = \mathcal{O}(\|H_n\|_\infty) = o(\tau_n)$  uniformly on  $s \in \mathcal{U}_q$  and  $x \in \text{supp}(X)$ . Moreover, using the first statement of Lemma 2, condition (12) entails that

$$\frac{\mathbb{P}(Y > y_n(u|x) | B_0^\top X = B_0^\top x - H_n s)}{\mathbb{P}(Y > y_n(u|x) | B_0^\top X = B_0^\top x)} = 1 + \mathcal{O}(\tau_n),$$

uniformly on  $s \in \mathcal{U}_q$ . As a first consequence,

$$\mu_n(u|x) \sim n|H_n|f_{B_0^\top X}(B_0^\top x)\mathbb{P}(Y > y_n(u|x) | B_0^\top X = B_0^\top x).$$

First part of Lemma 2 entails that for all  $\nu \in (0, 1)$ ,  $\mu_n(\nu|x)$  is proportional to  $n|H_n|\alpha_n \rightarrow \infty$  and that  $\mu_n(\nu|x)/\mu_n(1|x) \geq \nu/2$  for  $n$  large enough. Hence, condition **(C.2)** is satisfied. It also appears in turn that  $v_{n,x}$  is asymptotically proportional to  $\tau_n$  for all  $x \in \text{supp}(X)$ . It remains to show **(C.3)**. Let  $\delta_n(u|x) := Q(0|X = x) - y_n(u|x)$  if  $Q(0|X = x) < \infty$  and  $\delta_n(u|x) := [y_n(u|x)]^{-1}$  if  $Q(0|X = x) = +\infty$ . From the second statement of Lemma 2,  $\delta_n(u|x) \rightarrow 0$  uniformly on  $u \in [\nu, 1]$ . Hence, there exists  $N_x \in \mathbb{N}$  such that for all  $n \geq N_x$ ,  $\delta_n(u|x) \in (0, \kappa]$  where  $\kappa$  is such that (9) holds. Using (21) with  $Z = B_0^\top X$  and under (9), one has for  $n \geq N_x$

$$\begin{aligned} \left| \frac{\mathbb{P}(Y > y_n(u|x) | B_0^\top X = B_0^\top x)}{\mathbb{P}(Y > y_n(u|x) | X = x)} - 1 \right| &= \left| \frac{\mathbb{P}(Y > \mathcal{Y}_{\delta_n(u|x)}(B_0^\top x) | B_0^\top X = B_0^\top x)}{\mathbb{P}(Y > \mathcal{Y}_{\delta_n(u|x)}(B_0^\top x) | X = x)} - 1 \right| \\ &= \left| \frac{1 + \mathbb{E}[\eta_{\delta_n(u|x)}(X) | B_0^\top X = B_0^\top x]}{1 + \eta_{\delta_n(u|x)}(X = x)} - 1 \right| \\ &\leq \frac{2\bar{\eta}(\delta_n^{-1}(u|x))}{1 - \bar{\eta}(\delta_n^{-1}(u|x))}. \end{aligned}$$

Since  $\bar{\eta}(\cdot)$  is a decreasing function, the second statement of Lemma 2 leads to  $\bar{\eta}(\delta_n^{-1}(u|x)) \leq \bar{\eta}(\xi\delta_n^{-1}(X = x)) = \mathcal{O}(\tau_n)$  for all  $x \in \mathcal{A}$ . Hence, locally uniformly on  $u \in (0, \infty)$ ,

$$\frac{\mu_n(u|x)}{n|H_n|f_{B_0^\top X}(B_0^\top x)\mathbb{P}(Y > y_n(u|x) | X = x)} = 1 + \mathcal{O}(\tau_n). \quad (29)$$

Since  $\tau_n^{-1}\text{ERV}(\alpha_n, u|X = x) \rightarrow 0$  locally uniformly, one can use [14, Lemma 3] entailing that for  $u \in [\nu, 1]$  and  $u' \in [\nu, 1]$  with  $|u - u'| \leq (\mu_n(1|x))^{-1/2} = o(\tau_n)$

$$\begin{aligned} & \alpha_n^{-1} [\mathbb{P}(Y > y_n(u|x) | X = x) - \mathbb{P}(Y > y_n(u'|x) | X = x)] \\ &= \frac{1}{L_{\gamma(x)}^\leftarrow[L_{\gamma(x)}(1/u) + o(1)]} - \frac{1}{L_{\gamma(x)}^\leftarrow[L_{\gamma(x)}(1/u') + o(1)]} + o(\tau_n) \end{aligned} \quad (30)$$

where the terms in little-o do not depend on  $u$  and  $u'$  and  $L_{\gamma(x)}^{\leftarrow}(v) := (1 + \gamma(x)v)^{1/\gamma(x)}$ . Since the derivatives of  $1/L_{\gamma(x)}^{\leftarrow}(\cdot)$  and  $L_{\gamma(x)}(1/\cdot)$  are bounded, it is easy to check that (30) is a big-o of  $\tau_n$  and thus, from (29), condition **(C.3)** is satisfied. Thus, the consistency result (28) is true. Using (29) and since  $v_{n,x}$  is asymptotically proportional to  $\tau_n$  for all  $x \in \text{supp}(X)$ , (28) can be rewritten as

$$\sup_{u \in [\nu, 1]} \left| \frac{\hat{\Phi}_n(u|x)}{n|H_n|f_{B_0^\top X}(B_0^\top x)\mathbb{P}(Y > y_n(u|x)|X = x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(\tau_n). \quad (31)$$

Collecting (27) and (31) concludes the proof.  $\blacksquare$

**Proof of Theorem 2** – First, the consistency of  $\hat{\gamma}_n(B_0, x)$  is a direct consequence of [14, Theorem 1]. Indeed, under the assumptions of Theorem 2, Proposition 4 holds and thus assumptions of [14, Theorem 1] are satisfied leading to

$$\hat{\gamma}_n(B_0, x) - \gamma(x) = \mathcal{O}_{\mathbb{P}}(\tau_n) \text{ and } \hat{\gamma}_{n,-}(B_0, x) - \gamma_{-}(x) = \mathcal{O}_{\mathbb{P}}(\tau_n), \quad (32)$$

almost everywhere for  $x \in \text{supp}(X)$ . We are now interested in showing the consistency of the estimator  $\hat{a}_n(B_0, x)$ . We have

$$\frac{\hat{a}_n(B_0, x)}{a(\alpha_n^{-1}|x)} = \frac{\hat{Q}_n(\alpha_n|B_0, x)}{a(\alpha_n^{-1}|x)} \mathcal{T}_{\alpha_n}^{(1)}(\hat{Q}_n(\cdot|B_0, x)) \int_{\nu}^1 \varphi(u) L_0(1/u) du \Big/ \int_{\nu}^1 \varphi(u) L_{\hat{\gamma}_{n,-}(B_0, x)}(1/u) du.$$

Using the inequality  $|1 - \exp(x)| \leq |x| + x^2$  that holds for all  $x < \ln(2)$ , we have for all  $u \in (\nu, 1)$ ,

$$\begin{aligned} |L_{\hat{\gamma}_{n,-}(B_0, x)}(1/u) - L_{\gamma_{-}(x)}(1/u)| &\leq \int_1^{1/\nu} v^{\gamma_{-}(x)-1} |\exp[(\hat{\gamma}_{n,-}(B_0, x) - \gamma_{-}(x)) \ln(v)] - 1| dv \\ &\leq |\hat{\gamma}_{n,-}(B_0, x) - \gamma_{-}(x)| \tilde{L}_{\gamma_{-}(x)}(1/\nu) (1 + \ln^2(\nu) |\hat{\gamma}_{n,-}(B_0, x) - \gamma_{-}(x)|) = \mathcal{O}_{\mathbb{P}}(\tau_n), \end{aligned} \quad (33)$$

from (32). It is then straightforward to check that

$$\int_{\nu}^1 \varphi(u) L_{\hat{\gamma}_{n,-}(B_0, x)}(1/u) du \Big/ \int_{\nu}^1 \varphi(u) L_{\gamma_{n,-}(x)}(1/u) du = 1 + \mathcal{O}_{\mathbb{P}}(\tau_n).$$

As a consequence, since under the assumptions of Theorem 2,  $\hat{Q}_n(\alpha_n|B_0, x)/Q(\alpha_n|X = x) = 1 + \mathcal{O}_{\mathbb{P}}(\tau_n)$ , the estimator of  $a(\cdot|x)$  is such that

$$\frac{\hat{a}_n(B_0, x)}{a(\alpha_n^{-1}|x)} = \frac{Q(\alpha_n|X = x)}{a(\alpha_n^{-1}|x)} \mathcal{T}_{\alpha_n}^{(1)}(\hat{Q}_n(\cdot|B_0, x)) \frac{\int_{\nu}^1 \varphi(u) L_0(1/u) du}{\int_{\nu}^1 \varphi(u) L_{\gamma_{n,-}(x)}(1/u) du} (1 + \mathcal{O}_{\mathbb{P}}(\tau_n)).$$

A direct consequence of the result established in [14, eq. (30)] leads to

$$\hat{a}_n(B_0, x) = a(\alpha_n^{-1}|x)(1 + \mathcal{O}_{\mathbb{P}}(\tau_n)). \quad (34)$$

We are now in position to study the asymptotic behavior of  $\check{Q}_n(\cdot|B_0, x)$ . We start with

$$\begin{aligned} &\frac{\tau_n^{-1}}{a(\alpha_n^{-1}|X = x) \tilde{L}_{\gamma(x)}(\alpha_n/\beta_n)} |\check{Q}_n(\beta_n|B_0, x) - Q(\beta_n|X = x)| \\ &\leq \frac{1}{\tilde{L}_{\gamma(x)}(\alpha_n/\beta_n)} \frac{|\hat{Q}_n(\alpha_n|B_0, x) - Q(\alpha_n|X = x)|}{\tau_n a(\alpha_n^{-1}|x)} \\ &+ \tau_n^{-1} \frac{\hat{a}_n(B_0, x)}{a(\alpha_n^{-1}|x)} \frac{|L_{\hat{\gamma}_n(B_0, x)}(1/u) - L_{\gamma(x)}(1/u)|}{\tilde{L}_{\gamma(x)}(\alpha_n/\beta_n)} \\ &+ \tau_n^{-1} \frac{L_{\gamma(x)}(\alpha_n/\beta_n)}{\tilde{L}_{\gamma(x)}(\alpha_n/\beta_n)} \left| \frac{\hat{a}_n(B_0, x)}{a(\alpha_n^{-1}|x)} - 1 \right| + \frac{\tau_n^{-1}}{\tilde{L}_{\gamma(x)}(\alpha_n/\beta_n)} \text{ERV}(\alpha_n, \beta_n/\alpha_n|X = x) \\ &=: T_{1,n}(x) + T_{2,n}(x) + T_{3,n}(x) + T_{4,n}(x). \end{aligned}$$

As a direct consequence of Proposition 4 (see (26)),  $T_{1,n}(x) = \mathcal{O}_{\mathbb{P}}(1/\tilde{L}_{\gamma(x)}(\alpha_n/\beta_n)) = \mathcal{O}_{\mathbb{P}}(1)$  since, as  $t \rightarrow \infty$ ,

$$\tilde{L}_s(t) \sim \begin{cases} t^s \ln(t)/s & \text{if } s > 0, \\ \ln^2(t)/2 & \text{if } s = 0, \\ 1/s^2 & \text{if } s < 0. \end{cases}$$

Mimicking the proof of (33), since  $\tau_n \ln^2(\alpha_n/\beta_n) \rightarrow 0$  and using (34) lead to  $T_{2,n}(x) = (1 + \mathcal{O}_{\mathbb{P}}(\tau_n))\mathcal{O}_{\mathbb{P}}[\tau_n^{-1}(\hat{\gamma}_n(B_0, x) - \gamma(x))] = \mathcal{O}_{\mathbb{P}}(1)$ . Now, since for all  $s \in \mathbb{R}$ ,  $L_s(t)/\tilde{L}_s(t) = \mathcal{O}(1)$  as  $t \rightarrow \infty$ , equation (34) entails that  $T_{3,n}(x) = \mathcal{O}_{\mathbb{P}}(1)$ . Finally, since by assumption,  $\tau_n^{-1}\text{ERV}(\alpha_n, \alpha_n/\beta_n|X = x)/\tilde{L}_{\gamma(x)}(\alpha_n/\beta_n) \rightarrow 0$ , it is clear that  $T_{4,n}(x) = \mathcal{O}_{\mathbb{P}}(1)$  and the proof is complete.  $\blacksquare$

**Proof of Proposition 3** – Let  $B_0 \in \mathcal{B}$  and  $B_1 \in \mathcal{B}$  two  $p \times q$  matrices of rank  $q$  with  $B_0 \neq B_1$ . Assume that  $\mathcal{S}(B_0)$  and  $\mathcal{S}(B_1)$  are minimum TDR subspaces. Let us first introduce the linear space  $E := \mathcal{S}(B_0) \cap \mathcal{S}(B_1)$ . We assume that  $E = \mathcal{S}(C)$  where  $C$  is a  $p \times r$  matrix with  $r \in \{0, \dots, q\}$ . Note that if  $r = q$  then  $B_0 = B_1$  and the result is proved. From now on, we assume that  $r < q$ . There thus exist two  $p \times (q - r)$  matrices  $D_0$  and  $D_1$  such that  $\mathcal{S}(B_0) = \mathcal{S}(D_0, C)$  and  $\mathcal{S}(B_1) = \mathcal{S}(D_1, C)$ . Since  $\mathcal{S}(B_0)$  and  $\mathcal{S}(B_1)$  are TDR subspaces, there exists a Borel set  $\mathcal{A}$  of  $\mathbb{R}^p$  with  $\mathbb{P}(X \in \mathcal{A}) = 1$  and such that for all  $x \in \mathcal{A}$ ,

$$\lim_{\alpha \rightarrow 0} \frac{Q(\alpha|B_0^\top X = B_0^\top x)}{Q(\alpha|B_1^\top X = B_1^\top x)} = 1.$$

Now, for a given  $x^* \in \mathcal{A}$ , let  $E^*$  be the linear space given by  $\{x \in \mathbb{R}^p | B_0^\top x = B_0^\top x^*\}$ . It is clear that for all  $x \in E^*$ ,  $C^\top x = C^\top x^*$  and thus that

$$\lim_{\alpha \rightarrow 0} \frac{Q(\alpha|B_0^\top X = D_0^\top x^* + C^\top x^*)}{Q(\alpha|B_1^\top X = D_1^\top x + C^\top x^*)} = 1,$$

for all  $x \in E^*$ . Since  $\mathbb{P}(X \in \{D_1^\top x; x \in E^*\}) = 1$ , it appears that, almost surely, the conditional quantile  $Q(\alpha|B_1^\top X = D_1^\top x + C^\top x^*)$  is constant in  $D_1^\top x$ . As a consequence, for all  $x \in \mathcal{A}$

$$\lim_{\alpha \rightarrow 0} \frac{Q(\alpha|C^\top X = C^\top x)}{Q(\alpha|X = x)} = 1,$$

which is in contradiction with the fact that  $\mathcal{S}(B_0)$  is a minimum TDR subspace (since  $C$  is a matrix of rank  $r < q$ ).  $\blacksquare$

**Proof of Theorem 3** – First remark that for all  $x \in \text{supp}(X)$ ,

$$\alpha^{-1}\mathbb{P}[Y > Q(\alpha|B_0^\top X = B_0^\top x)|X = x] = \frac{\mathbb{P}[Y > \mathcal{Y}_{\delta_{\alpha,x}}(B_0^\top X = B_0^\top x)|X = x]}{\mathbb{P}[Y > \mathcal{Y}_{\delta_{\alpha,x}}(B_0^\top X = B_0^\top x)|B_0^\top X = B_0^\top x]},$$

where, in order to not overload the notations,  $\delta_{\alpha,x}$  stands for  $\delta_\alpha(B_0^\top X = B_0^\top x)$ . By assumption, there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^p)$  with  $\mathbb{P}(X \in \mathcal{A}) = 1$  such that for all  $x \in \mathcal{A}$  and  $\kappa > 0$ , there exists  $\alpha_0 \in (0, 1)$  such that for all  $\alpha \in (0, \alpha_0)$ ,  $\delta_{\alpha,x} < \kappa$ . Hence, since  $\mathcal{S}(B_0)$  is a TDR subspace, for all  $\varepsilon > 0$ , there exists  $\alpha_0 \in (0, 1)$  such that for  $\alpha \in (0, \alpha_0)$ ,

$$\mathbb{P} [|\alpha^{-1}\mathbb{P}[Y > Q(\alpha|B_0^\top X)|X] - 1| < \varepsilon] = 1. \quad (35)$$

Mimicking the proof of the second statement of Theorem 1 and introducing for all  $B \in \mathcal{B}$  and  $j \in \{1, \dots, J\}$  the quantity

$$\Delta^{(j)}(\alpha, B) := \frac{\mathbb{P}[\{Y > Q(\alpha|B^\top X)\} \cap \{X \in \Pi_B^{(j)}(B^\top X)\}|B^\top X]}{\alpha\mathbb{P}[X \in \Pi_B^{(j)}(B^\top X)|B^\top X]} - 1,$$

one has that almost surely,

$$\Delta^{(j)}(\alpha, B_0) = \mathbb{E} \left[ \left( \frac{\mathbb{P}[Y > Q(\alpha|B_0^\top X)|X]}{\alpha} - 1 \right) \frac{\mathbb{I}_{\{X \in \Pi_{B_0}^{(j)}(B_0^\top X)\}}}{\mathbb{P}[X \in \Pi_{B_0}^{(j)}(B_0^\top X)|B_0^\top X]} \middle| B_0^\top X \right]. \quad (36)$$

Collecting (35) and (36) show that for all  $\varepsilon \in (0, 1)$ , there exists  $\alpha_0 \in (0, 1)$  such that for  $\alpha \in (0, \alpha_0]$ ,

$$T(\alpha, B_0) = \sum_{j=1}^J \left\{ \mathbb{E}[\Delta^{(j)}(\alpha, B_0)] \right\}^2 \leq J\varepsilon^2. \quad (37)$$

As a first conclusion, we have proved that  $T(\alpha, B_0) \rightarrow 0$  as  $\alpha \rightarrow 0$ . Now, since under the assumptions of Theorem 3, a minimum TDR subspace is unique (see Proposition 3), and since  $|\alpha^{-1}\mathbb{P}(Y > Q(\alpha|B^\top X)|X) - 1|$  converges to a limit in  $[0, \infty]$  uniformly on  $B \in \mathcal{B}$  as  $\alpha \rightarrow 0$ , it is easy to check starting from (36) that for all  $\varepsilon > 0$  and  $\eta > 0$ , there exists  $\alpha_{\varepsilon, \eta} \in (0, 1)$  such that for all  $\alpha \in (0, \alpha_{\varepsilon, \eta}]$ ,

$$\sup_{B \in \mathcal{B}, \|B - B_0\| \geq \varepsilon} T(\alpha, B) > \eta.$$

Assume that  $\|\tilde{B}_0(\alpha) - B_0\|$  does not converges to 0 as  $\alpha \rightarrow 0$ . One can thus find  $\varepsilon > 0$  such that for all  $\alpha_0 \in (0, 1)$ , there exists  $\alpha \in (0, \alpha_0]$  such that  $\|\tilde{B}_0(\alpha) - B_0\| > \varepsilon$ . Hence, for all  $\eta > 0$  and for all  $\alpha_0 \in (0, \alpha_{\varepsilon, \eta}]$ , there exists  $\alpha \in (0, \alpha_0]$  such that  $T(\alpha, \tilde{B}_0(\alpha)) > \eta$ . This point is obviously in contradiction with the fact that  $T(\alpha, \tilde{B}_0(\alpha)) \leq T(\alpha, B_0) \rightarrow 0$ , as  $\alpha \rightarrow 0$ . ■

## References

- [1] Araújo Santos, P., Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks Over Random Threshold methodology for tail index and high quantile estimation, *REVSTAT*, **4(3)**, 227–247.
- [2] Basu, D. and C.A.B. Pereira (1983). Conditional independence in statistics, *Sankhyā*, **45(3)**, Series A, 324–337.
- [3] Chaudhuri, P. (1991). Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation, *Annals of Statistics*, **19(2)**, 760–777.
- [4] Cook, R.D. (1994). On the interpretation of regression plots. *Journal of the American Statistical Association*, **89(425)**, 177–189.
- [5] Cook, R.D. and Forzani, L. (2009). Likelihood-based sufficient dimension reduction. *Journal of the American Statistical Association*, **104**, 197–208.
- [6] Cook, R.D. and Weisberg, S. (1991). Discussion of "Sliced Inverse Regression for dimension reduction". *Journal of the American Statistical Association*, **86**, 328–332.
- [7] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, **19(5B)**, 2557–2589.
- [8] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, **20(2)**, 311–333.
- [9] Davison, A.C. and Ramesh, N.I. (2000). Local likelihood smoothing of sample extremes, *Journal of the Royal Statistical Society, series B*, **62**, 191–208.

- [10] Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *The Annals of Statistics*, **17**(4), 1833–1855.
- [11] Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.
- [12] Fraga Alves, M. I., Gomes, M. I., de Haan, L. and Neves, C. (2007). A note on second order conditions in extreme value theory: linking general and heavy tail conditions, *REVSTAT - Statistical Journal*, **5**(3), 285–304.
- [13] Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2004). Sliced inverse regression in reference curves estimation, *Computational Statistics & Data Analysis*, **46**(1), 103–122.
- [14] Gardes, L. (2015). A general estimator for the extreme value index: applications to conditional and heteroscedastic extremes, *Extremes*, **18**(3), 479–510.
- [15] Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels, *Extremes*, **13**(2), 177–204.
- [16] Gardes, L. and Girard, S. (2012). Functional kernel estimators of large conditional quantiles, *Electronic Journal of Statistics*, **6**, 1715–1744.
- [17] Gardes, L. and Girard, S. (2016). On the estimation of the functional Weibull tail-coefficient, *Journal of Multivariate Analysis*, **146**, 29–45.
- [18] Gasser, T. and Müller, H.G. (1984). Estimating regression functions and their derivatives by the kernel method, *Scandinavian Journal of Statistics*, **11**, 171–185.
- [19] Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d’une série aléatoire, *Annals of Mathematics*, **44**, 423–453.
- [20] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer.
- [21] Hooke, R. and Jeeves, T.A. (1961). Direct search solution of numerical and statistical problems, *Journal of the ACM*, **8**(2), 212–229.
- [22] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of econometrics*, **58**, 71–120.
- [23] Koenker, R. and Basset, G.S. (1978). Regression quantiles, *Econometrica*, **46**, 33–50.
- [24] Li, K.C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86**, 316–327.
- [25] Li, K.C. (1992). On principal hessian directions for data visualization and dimension reduction: another application of Stein’s lemma. *Journal of the American Statistical Association*, **87**, 1025–1039.
- [26] Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and its Application*, **9**(1), 141–142.
- [27] Parzen, E. (1962). On the estimation of a probability density function and mode, *Annals of Mathematical Statistics*, **33**, 1065–1076.

- [28] Scruc a, L. (2011). Model-based SIR for dimension reduction, *Computational Statistics and Data Analysis*, **55**, 3010–3026.
- [29] Wang, E. and Tsai, C. (2009). Tail Index Regression, *Journal of the American Statistical Association*, **104(487)**, 1233–1240.
- [30] Watson, G. S. (1964). Smooth regression analysis, *Sankhya*, **26(15)**, 175–184.
- [31] Weisberg, W. (2002). Dimension reduction regression in R, *Journal of Statistical Software*, **7**, 1–22.
- [32] Wu, T.Z., Yu, K. and Yu, Y. (2010). Single-index quantile regression, *Journal of Multivariate Analysis*, **101**, 1607–1621.
- [33] Xia, Y. (2009). Model checking in regression via dimension reduction, *Biometrika*, **96(1)**, 133–148.
- [34] Yu, K. and Jones, M.C. (1998). Local Linear Quantile Regression, *Journal of the American Statistical Association*, **93(441)**, 228–237.

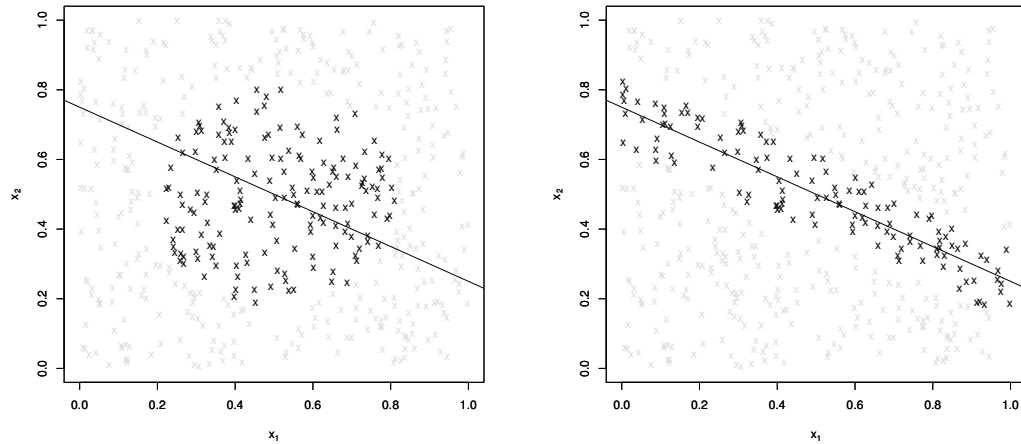


Figure 1: Representation of  $n = 500$  realizations of the random vector  $X = (X_1, X_2)^\top$ . The straight line is the set  $\{x = (x_1, x_2)^\top, B_0^\top x = B_0^\top x_0\}$ . The black points are the observations used to compute the estimators  $\check{Q}_n(\cdot|x_0)$  (left) and  $\check{Q}_n^{(DR)}(\cdot|B_0, x_0)$  (right).

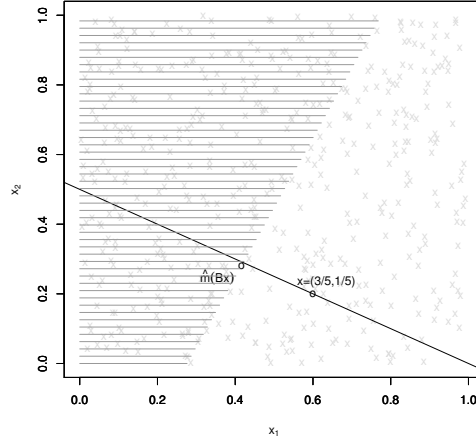


Figure 2: Representation of  $n = 500$  realizations of the random vector  $X = (X_1, X_2)^\top$ . The straight line is the set  $\{s \in \mathbb{R}^2, B^\top s = B^\top x = 1\}$  with  $B = (1, 2)^\top / \sqrt{5}$ . The set  $\Pi_1(B^\top X = B^\top x)$  is the hatched area, the rest is the set  $\Pi_2(B^\top X = B^\top x)$ .

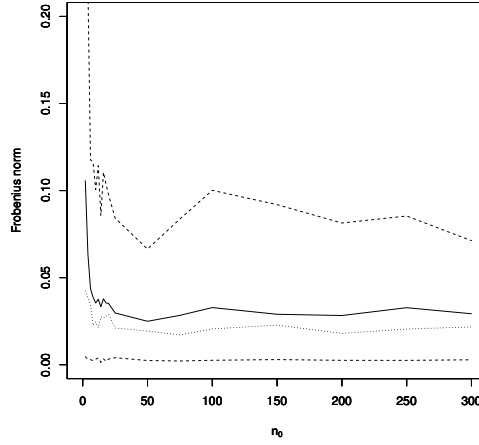


Figure 3: Under model 1 with  $\sigma = 1/5$ , the empirical quantiles and the empirical mean of the  $N = 100$  obtained values of  $\|\hat{B}_{0,n}(\alpha_n) - B_0\|_F$  are represented as a function of  $n_0$ . The two dashed lines are the empirical quantiles of order 0.05 and 0.95, the dotted line is the median and the full line is the empirical mean.

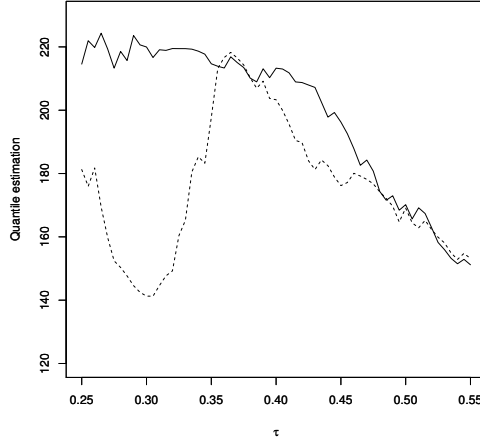


Figure 4: For the air pollution dataset, estimators  $\check{Q}(1/n|\hat{B}_{0,n}(\alpha_n), x_\tau)$  (full line)  $\check{Q}(1/n|\hat{B}_{0,n}^{\text{SIR}}(\alpha_n), x_\tau)$  (dashed line) of the conditional quantile of order  $1/n$  of  $Y$  given  $X = x_\tau$  as a function of  $\tau$ .

		$\ \hat{B}_{0,n} - B_0\ _F$	$\ \hat{B}_{0,n}^{\text{SIR}} - B_0\ _F$	$E_Q(\hat{B}_{0,n})$	$E_Q(B_0)$	$E_Q(I_p)$
Model 1	(1)	0.031 (0.100)	1.759 (0.048)	0.966 (1.787)	0.571 (0.411)	6.966 (2.801)
	(2)	0.029 (0.025)	1.751 (0.058)	0.841 (0.491)	0.672 (0.540)	4.353 (0.900)
	(3)	0.061 (0.053)	1.734 (0.082)	1.977 (1.304)	1.631 (1.333)	5.253 (0.377)
Model 2	(1)	0.082 (0.318)	1.511 (0.130)	0.232 (0.151)	0.199 (0.004)	0.252 (0.020)
	(2)	0.006 (0.014)	1.375 (0.148)	0.070 (0.016)	0.063 (0.004)	0.847 (0.058)
	(3)	0.005 (0.004)	1.167 (0.208)	0.059 (0.005)	0.057 (0.005)	0.461 (0.043)
Model 3	(1)	0.026 (0.089)	1.351 (0.050)	0.123 (0.303)	0.058 (0.043)	1.638 (0.396)
	(2)	0.028 (0.030)	1.376 (0.063)	0.072 (0.039)	0.051 (0.031)	0.647 (0.170)
	(3)	0.057 (0.050)	1.395 (0.101)	0.109 (0.049)	0.084 (0.053)	0.428 (0.070)

Table 1: Estimation of the TDR direction and large conditional quantile of order  $\beta_n = 2/n$  under models 1 to 3 with (1):  $\sigma = 1/3$ , (2):  $\sigma = 1/5$  and (3):  $\sigma = 1/8$ . The given values are the empirical means over the  $N = 100$  replications (the standard deviation is between brackets).

## Appendix – Additional results

The aim of the next result is to rephrase condition (12) appearing in Theorem 2 in term of the conditional distribution of  $Y$  given  $X$ .

**Lemma 3.** *Assume that there exists a full rank matrix  $B_0$  such that  $\mathcal{S}(B_0)$  is a TDR subspace (i.e. for all  $\delta > 0$ ,  $\mathbb{P}(Y > \mathcal{Y}_\delta(B_0^\top X)|X) = s_\delta(B_0^\top X)(1 + \eta_\delta(X))$  a.s. where  $s_\delta(B_0^\top X = \cdot)$  and  $\eta_\delta(X = \cdot)$  are measurable functions) and that condition (5) hold. For  $(x, t) \in \text{supp}(X) \times \mathbb{R}^q$  and  $\zeta \in (0, 1)$ , let*

$$\tilde{\delta}_\zeta(t, x) := \begin{cases} Q(0|X = B_0(B_0^\top B_0)^{-1}t) - Q(\zeta|X = x) & \text{if } Q(0|X = x) < \infty, \\ 1/Q(\zeta|X = x) & \text{if } Q(0|X = x) = \infty. \end{cases}$$

If  $(\alpha_n)$  and  $(H_n)$  are two sequences converging to 0 with  $n|H_n|\alpha_n \rightarrow \infty$ , if there exists  $\xi \in (0, 1)$  such that

$$\max \left\{ \sup_{(t, \zeta) \in A_n} \left| \frac{s_{\check{\delta}_\zeta(t, x)}(B_0^\top X = t)}{s_{\check{\delta}_\zeta(B_0^\top x, x)}(B_0^\top X = B_0^\top x)} - 1 \right| ; \bar{\eta}(\xi \delta_{\alpha_n}^{-1}(X = x)) \right\} = \mathcal{O}(\tau_n),$$

where  $A_n = D(B_0^\top x, H_n) \times [\xi \nu \alpha_n, \xi^{-1} \alpha_n]$  and if for all  $\varepsilon > 0$ , there exists  $N \in \mathbb{N}$  such that for  $n \geq N$ ,

$$\begin{cases} \inf_{t \in D(B_0^\top x, H_n)} Q(0|X = B_0(B_0^\top B_0)^{-1}t) = \infty & \text{if } Q(0|X = x) = \infty, \\ \sup_{t \in D(B_0^\top x, H_n)} |Q(0|X = B_0(B_0^\top B_0)^{-1}t) - Q(0|X = x)| / a(\alpha_n^{-1}|x) < \varepsilon & \text{if } Q(0|X = x) < \infty, \end{cases} \quad (38)$$

then, there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^p)$  with  $\mathbb{P}(X \in \mathcal{A}) = 1$  such that for all  $x \in \mathcal{A}$ ,

$$\sup_{(t, \zeta) \in A_n} \left| \frac{\mathbb{P}(Y > Q(\zeta|B_0^\top X = B_0^\top x)|B_0^\top X = t)}{\zeta} - 1 \right| = \mathcal{O}(\tau_n).$$

**Proof** – We start with the fact that for all  $\delta > 0$  and  $t \in \text{supp}(B_0^\top X)$ ,

$$\mathbb{P}(Y > \mathcal{Y}_\delta(t)|B_0^\top X = t) = s_\delta(B_0^\top X = t) \{1 + \mathbb{E}[\eta_\delta(X)|B_0^\top X = t]\}.$$

For  $\zeta \in (0, 1)$ , let

$$\check{\delta}_{0, \zeta}(t, x) := \begin{cases} Q(0|X = B_0(B_0^\top B_0)^{-1}t) - Q(\zeta|B_0^\top X = B_0^\top x) & \text{if } Q(0|X = x) < \infty, \\ 1/Q(\zeta|B_0^\top X = B_0^\top x) & \text{if } Q(0|X = x) = \infty. \end{cases}$$

Since  $\mathcal{S}(B_0)$  is a TDR subspace one has that  $Q(0|B_0(B_0^\top B_0)^{-1}t) = Q(0|B_0^\top X = B_0^\top x)$  almost surely and thus,

$$\begin{aligned} \frac{\mathbb{P}(Y > Q(\zeta|B_0^\top X = B_0^\top x)|B_0^\top X = t)}{\zeta} &= \frac{s_{\check{\delta}_{0, \zeta}(t, x)}(B_0^\top X = t)}{s_{\check{\delta}_{0, \zeta}(B_0^\top x, x)}(B_0^\top X = B_0^\top x)} \\ &\times \frac{1 + \mathbb{E}[\eta_{\check{\delta}_{0, \zeta}(t, x)}(X)|B_0^\top X = t]}{1 + \mathbb{E}[\eta_{\check{\delta}_{0, \zeta}(B_0^\top x, x)}(X)|B_0^\top X = B_0^\top x]}. \end{aligned} \quad (39)$$

Let us focus on the first factor of (39). Under (5) and since  $\mathcal{S}(B_0)$  is a TDR subspace, using [20, Lemma 1.2.12], it is easy to check that there exists a Borel set  $\mathcal{A} \in \mathcal{B}(\mathbb{R}^p)$  with  $\mathbb{P}(X \in \mathcal{A}) = 1$  such that for all  $x \in \mathcal{A}$  and uniformly on  $\zeta \in [\xi \nu \alpha_n, \xi^{-1} \alpha_n]$ ,

$$\begin{aligned} Q(\zeta|B_0^\top X = B_0^\top x) &= Q(\zeta|X = x) + a(\alpha_n^{-1}|x) \left( \frac{Q(\zeta|B_0^\top X = B_0^\top x) - Q(\zeta|X = x)}{a(\alpha_n^{-1}|x)} \right) \\ &= Q(\zeta|X = x) + o(a(\alpha_n^{-1}|x)). \end{aligned}$$

Hence, mimicking the proof of the first statement of Lemma 2, we know that there exists  $\tilde{\xi} \in (0, 1)$  such that  $Q(\zeta|B_0^\top X = B_0^\top x) = Q(\tilde{\zeta}|X = x)$  where  $\tilde{\zeta} \in [\tilde{\xi} \zeta, \tilde{\xi}^{-1} \zeta]$ . Hence  $\check{\delta}_{0, \zeta}(t, x) = \tilde{\delta}_{\tilde{\zeta}}(t, x)$  and, by assumption, one can find  $\xi \in (0, 1)$  such that

$$\sup_{(t, \zeta) \in A_n} \left| \frac{s_{\check{\delta}_{0, \zeta}(t, x)}(B_0^\top X = t)}{s_{\check{\delta}_{0, \zeta}(B_0^\top x, x)}(B_0^\top X = B_0^\top x)} - 1 \right| = \mathcal{O}(\tau_n), \quad (40)$$

for all  $x \in \mathcal{A}$ .

Let us now consider the second factor of (39). We have shown before that  $\check{\delta}_{0, \zeta}(t, x) = \tilde{\delta}_{\tilde{\zeta}}(t, x)$ .

Hence, if  $Q(0|X=x) < \infty$ , under (38), one has for all  $t \in D(B_0^\top x, H_n)$  that  $\check{\delta}_{0,\zeta}(t, x) = Q(0|X=x) - [Q(\check{\zeta}|X=x) - o(a_n(\alpha_n^{-1}|x))]$  and if  $Q(0|X=x) = \infty$ ,  $\check{\delta}_{0,\zeta}(t, x) = 1/Q(\check{\zeta}|X=x)$ . Using the second statement of Lemma 2, one can find  $\xi \in (0, 1)$  such that  $\check{\delta}_{0,\zeta}(t, x) \leq \xi^{-1}\delta_{\alpha_n}(X=x)$  uniformly on  $t \in D(B_0^\top x, H_n)$ . Applying condition (9) and since  $\bar{\eta}(\xi\delta_{\alpha_n}^{-1}(X=x)) = \mathcal{O}(\tau_n)$ , leads to

$$\frac{1 + \mathbb{E}[\eta_{\check{\delta}_{0,\beta}(t,x)}(X)|B_0^\top X = t]}{1 + \mathbb{E}[\eta_{\check{\delta}_{0,\beta}(B_0^\top x, x)}(X)|B_0^\top X = B_0^\top x]} = \mathcal{O}(\tau_n). \quad (41)$$

Collecting (39), (40) and (41) concludes the proof.  $\blacksquare$

The aim of the following lemma is to study the distributions introduced in paragraph 5.2.

**Lemma 4.** *Using same notations as in paragraph 5.2, one has for all  $x \in \mathbb{R}^p$  and  $\delta > 0$ ,*  
*(i)  $S_1(\delta^{-1}|x) = s_{\delta,1}(B_0^\top X = B_0^\top x)[1 + \eta_{\delta,1}(X=x)]$  with  $s_{\delta,1}(B_0^\top X = B_0^\top x) = \delta^{1/g_0(B_0^\top x)}$  and for  $\delta \in [0, 1]$ ,  $|\eta_{\delta,1}(X=x)| \leq \bar{\eta}_1(\delta^{-1})$  where  $\bar{\eta}_1(\cdot)$  is a decreasing function converging to 0 at infinity and defined for  $y > 1$  by  $\bar{\eta}_1(y) = 1 - y^{-1/3}[\ln(1/(1 - \delta^{-1/3}))]^{-1}[1 + \exp(10 - y^{1/3})]^{-3}$ .*  
*(ii)  $S_2(g_2(B_0^\top x) - \delta|x) = S_1(\delta^{-1}|x)$ .*  
*(iii)  $S_3(\delta^{-1}|x) = \exp[-\delta^{-1/g_0(B_0^\top x)}L_3(\delta^{-1}|x)]$  where  $L_3(\cdot|x)$  converges to 1 at infinity. Furthermore,  $S_3(\delta^{-1}|x) = s_{\delta,3}(B_0^\top X = B_0^\top x)[1 + \eta_{\delta,3}(X=x)]$  with  $s_{\delta,3}(B_0^\top X = B_0^\top x) = \exp[-\delta^{-1/g_0(B_0^\top x)}]$  and for  $\delta \in [0, 1]$ ,  $|\eta_{\delta,3}(X=x)| \leq \bar{\eta}_3(\delta^{-1})$  where  $\bar{\eta}_3(\cdot)$  is a decreasing function converging to 0 at infinity and defined for  $y > 1$  by  $\bar{\eta}_3(y) = 1 - \exp\{y^3[1 - (1 + \exp(10 - \exp(y^{1/3}))^3]\}$ .*

**Proof** – (i) First, remark that for  $\alpha \in [0, 1]$  and  $x \in \mathbb{R}^p$ ,  $Q_1(\alpha|X=x) = \alpha^{-g_0(B_0^\top x)}\tilde{L}_1(\alpha|x)$ , where  $\tilde{L}_1(\alpha|x) = \alpha^{g_0(B_0^\top x)}[\ln(1/(1-\alpha))]^{-g_0(B_0^\top x)}[1 + g_1(B_0^\top x)\exp(-\alpha^{-1})]^{-1}$ . Since the distribution function of  $Y$  given  $X=x$  is continuous,  $Q_1(S_1(\delta^{-1}|x)|X=x) = \delta^{-1}$  for all  $\delta > 0$  and thus,

$$S_1(\delta^{-1}|x) = \delta^{1/g_0(B_0^\top x)}[\tilde{L}_1(S_1(\delta^{-1}|x)|x)]^{1/g_0(B_0^\top x)}. \quad (42)$$

It is quite easy to check that on  $[0, 1]$ ,  $\tilde{L}_1(\cdot|x)$  is a decreasing function with  $\tilde{L}_1(\alpha|x) \rightarrow 1$  as  $\alpha \rightarrow 0$ . Hence, (42) entails that for  $\delta \in [0, 1]$ ,  $S_1(\delta^{-1}|x) \leq \delta^{1/g_0(B_0^\top x)} \leq \delta^{1/3}$  since  $g_0(B_0^\top x) \in [1/3, 3]$ . As a consequence, for  $\delta \in [0, 1]$ ,

$$|\eta_{\delta,1}(X=x)| = 1 - [\tilde{L}_1(S_1(\delta^{-1}|x)|x)]^{1/g_0(B_0^\top x)} \leq 1 - [\tilde{L}_1(\delta^{1/3}|x)]^3 \leq \bar{\eta}_1(\delta^{-1}),$$

since  $g_0(B_0^\top x) \in [1/3, 3]$  and  $g_1(B_0^\top x) \in [1, \exp(10)]$ .

(ii) This is a direct consequence of the equality  $Q_2(\alpha|X=x) = g_2(B_0^\top x) - [Q_1(\alpha|X=x)]^{-1}$ .  
 (iii) Since the distribution function of  $Y$  given  $X=x$  is continuous,  $Q_3(S_3(\delta^{-1}|x)|X=x) = \delta^{-1}$  for all  $\delta > 0$  and thus,  $S_3(\delta^{-1}|x) = \exp[-\delta^{-1/g_0(B_0^\top x)}\tilde{L}_3(S_3(\delta^{-1}|x)|x)]$  with  $\tilde{L}_3(\alpha|x) = [1 + g_1(B_1^\top x)\exp(-\alpha^{-1})]^{1/g_0(B_0^\top x)}$ . Obviously,  $L_3(\cdot|x) := \tilde{L}_3(S_3(\cdot|x)|x)$  converges to 1 at infinity. Furthermore,  $S_3(\delta^{-1}|x) = \exp[-\delta^{-1/g_0(B_0^\top x)}][1 + \eta_{\delta,3}(X=x)]$  with

$$\eta_{\delta,3}(X=x) = \exp\left\{\delta^{-1/g_0(B_0^\top x)}\left[1 - \tilde{L}_3(S_3(\delta^{-1}|x)|x)\right]\right\} - 1.$$

It is easy to check that on  $[0, 1]$ ,  $\tilde{L}_3(\cdot|x)$  is an increasing function larger than 1. Thus, for all  $\delta \in [0, 1]$ ,  $S_3(\delta^{-1}|x) \leq \exp(-\delta^{-1/g_0(B_0^\top x)}) \leq \exp(-\delta^{-1/3})$  and hence,

$$\begin{aligned} |\eta_{\delta,3}(X=x)| &= 1 - \exp\left\{\delta^{-1/g_0(B_0^\top x)}\left[1 - \tilde{L}_3(S_3(\delta^{-1}|x)|x)\right]\right\} \\ &\leq 1 - \exp\left\{\delta^{-1/g_0(B_0^\top x)}\left[1 - \tilde{L}_3(\exp(-\delta^{-1/3})|x)\right]\right\} \leq \bar{\eta}_3(\delta^{-1}), \end{aligned}$$

using the facts that  $g_0(B_0^\top x) \in [1/3, 3]$  and  $g_1(B_1^\top x) \in [1, \exp(10)]$ .  $\blacksquare$