



HAL
open science

Dimension reduction for the estimation of the conditional tail-index

Laurent Gardes, Alex Podgorny

► **To cite this version:**

Laurent Gardes, Alex Podgorny. Dimension reduction for the estimation of the conditional tail-index. 2024. hal-04589742

HAL Id: hal-04589742

<https://hal.science/hal-04589742>

Preprint submitted on 27 May 2024

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.

Dimension reduction for the estimation of the conditional tail-index

L. Gardes and A. Podgorny

Abstract

We are interested in the relationship between the large values of a real random variable Y and its associated covariate X that takes its values in a subset \mathcal{X} of \mathbb{R}^p when the conditional distribution of Y given $X = x$ is heavy-tailed with tail index $\gamma(x) > 0$. Estimating this index is a crucial step for the inference of the conditional distribution, but this task becomes more challenging as the dimension p increases. The objective of this work is to propose a dimension reduction method to obtain a more efficient estimator of the extreme value index. Specifically, we assume the existence of a subspace \mathcal{S}_0 of dimension $q < p$ with basis $B_0 \in \mathbb{R}^{p \times q}$ and a positive function $g(\cdot)$ such that for all $x \in \mathcal{X}$, $\gamma(x) = g(B_0^\top x)$. We propose a method to estimate the matrix B_0 and establish its consistency. We illustrate the advantages of this dimension reduction procedure for estimating the extreme value index through simulations and an application to real data.

Keywords— Tail-index, dimension reduction, heavy-tailed distributions

1 Introduction

Let (X, Y) be a random couple where Y is a \mathbb{R} -valued response variable associated to a covariate X taking its values in the set $\mathcal{X} \subset \mathbb{R}^p$ with $p \in \mathbb{N} \setminus \{0\}$. The purpose of nonparametric regression is the estimation of the regression function $\mathcal{X} \ni x \mapsto \mathbb{E}(Y | X = x)$. Given a sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of independent copies

of (X, Y) and a point $x \in \mathcal{X}$, two main approaches for the estimation of $\mathbb{E}(Y | X = x)$ can be found in the literature: the kernel method introduced independently by Nadaraya [13] and Watson [21] and the nearest-neighbors approach (see for instance Cover [6]). However, it is well known that, for a given level of accuracy of one of these estimators, the number n of observations grows exponentially with respect to the dimension p . This phenomenon is often referred to as the curse of dimensionality (see Bellman [2]). A solution to avoid this problem is to assume the existence of a mean dimension reduction subspace (see Cook and Li [5]). Recall that a q -dimensional linear subspace \mathcal{S} is a mean dimension reductions subspace if $\mathbb{E}(Y | X) = \mathbb{E}(Y | B^\top X)$ almost surely where $B \in \mathbb{R}^{p \times q}$ is a basis of \mathcal{S} . Hence, if B is known (or at least can be estimated), inference on $\mathbb{E}(Y | X)$ can be done by replacing X by the covariate $B^\top X$ of dimension $q \leq p$. When q is much smaller than p , the kernel method and the nearest-neighbors approaches provide more accurate estimates.

In some practical cases, there is a need to pay special attention to the right tail of the distribution, rather than to the central part of the distribution. This is the case in finance (see, e.g., Rockafellar and Uryasev [15]) or insurance (see for instance Brazauskas et al. [4] and Read and Vogel [14]). In these fields of application, conditional heavy-tailed distributions are often considered and will be the focus of this paper. Recall that the conditional distribution of Y given X is heavy-tailed if

$$S(y, x) := \mathbb{P}(Y > y | X = x) = y^{-1/\gamma(x)} \mathcal{L}(y, x),$$

for almost all $x \in \mathcal{X}$, where $\gamma(\cdot)$ is a positive function referred to as the conditional tail-index and $\mathcal{L}(\cdot, x)$ is a positive slowly varying function, namely $\mathcal{L}(ty, x)/\mathcal{L}(y, x) \rightarrow 1$ as $y \rightarrow \infty$ for all $t > 0$ and $x \in \mathcal{X}$. The conditional tail-index controls the tail behavior of the conditional distribution. The capability to estimate it accurately is therefore of primary importance. Of course, we are still confronted to the curse of dimensionality all the more so since only the largest observations are kept in the estimation process. Despite its practical interest, little work has been done on dimension reduction for extreme values. A first attempt can be found in Gardes [7] where it is assumed that there exists a q -dimensional subspace \mathcal{S} for which

$$\lim_{y \rightarrow \infty} \sup_{x \in \mathcal{X}} \left| \frac{S(y, x)}{S_B(y, B^\top x)} - 1 \right| = 0, \quad (1)$$

where $B \in \mathbb{R}^{p \times q}$ is a basis of \mathcal{S} and $S_B(\cdot, B^\top x)$ is the conditional survival function of Y given $B^\top X = B^\top x$. If such a tail dimension reduction subspace \mathcal{S} exists, inference on the conditional tail distribution can be achieved by replacing the covariate X by its reduced version $B^\top X$. Recently, a different way to define a tail dimension reduction subspace was introduced in Aghbalou et al. [1]. Their approach can be seen as an adaptation of the Sliced Inverse Regression method introduced by Li [12] and differs significantly from Gardes [7]. In particular, condition (1) and the specific condition used in Aghbalou et al. [1] are not equivalent. Moreover, the final objective in Gardes [7] is the estimation of extreme quantile while the one in Aghbalou et al. [1] is to predict the occurrence of tail event. Another attempt to reduce the dimension in the tail of the distribution can be found in Bousebata et al. [3]. The authors propose a specific inverse regression model and the dimension reduction is achieved by using a tail version of the partial least square method.

As mentioned above, an important step in inference on conditional heavy-tailed distributions is the estimation of the conditional tail index, but none of the aforementioned papers specifically address this issue. However, in some situations, a more accurate dimension reduction subspace can be found for the purpose of estimating $\gamma(\cdot)$. For instance, if for all $x \in \mathcal{X}$ there exists $B \in \mathbb{R}^{p \times q}$ with $q < p$ such that $S(y, x) = y^{-1/g(B^\top x)} c(x)$, where $g(\cdot)$ and $c(\cdot)$ are positive functions then the tail dimension reduction subspace defined in Gardes [7] is the whole set \mathbb{R}^p while in this case, the dimension reduction subspace of interest is obviously $\text{span}(B)$, the linear subspace of \mathbb{R}^p spanned by the columns of B .

The main contribution of this work is precisely to deal with the situation where the conditional tail-index is defined on a lower dimensional linear subspace. This subspace is referred to as the tail-index dimension reduction (TIDR) subspace. The framework and the main definitions, in particular the definition of the TIDR subspace \mathcal{S} , are given in Section 2. Section 3 is devoted to the estimation of a basis of \mathcal{S} . The consistency of the proposed estimator is also established. Section 4 presents the results of a simulation study, followed by an application to a real data set in Section 5. A procedure to estimate the dimension of the TIDR subspace is also presented and illustrated in Section 4. All the proofs are gathered in Section 6.

2 Tail-index dimension reduction subspace

2.1 Framework

Throughout this paper, we consider a $(\mathbb{R}^p \times \mathbb{R})$ -valued random couple (X, Y) defined on a given probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For any $q \leq p$, $B \in \mathbb{R}^{p \times q}$ and $y \in \mathbb{R}$, let

$$S_B(y, B^\top X) := \mathbb{P}(Y > y \mid B^\top X)$$

be the conditional survival function of Y given $B^\top X$. When B is the identity matrix (i.e., $q = p$) the conditional survival function of Y given that X is denoted $S(y, X)$. For $\omega \in \Omega$ such that $B^\top X(\omega) = z$, we write $S_B(y, B^\top X)(\omega) = S_B(y, z)$. The couple (X, Y) is distributed according to the following model.

- (M) The support \mathcal{X} of X is assumed to be compact with a non-empty interior and the support of Y is confined to the positive half-line. Moreover, for all $x \in \mathcal{X}$, the conditional distribution of Y given $X = x$ is heavy tailed with tail-index $\gamma(x)$ i.e., $S(y, x) = y^{-1/\gamma(x)} \mathcal{L}(y, x)$, for some slowly varying function $\mathcal{L}(\cdot, x)$.

In Section 2.2, we establish a result ensuring (under suitable conditions) that if the distribution of Y given X is heavy-tailed so is the conditional distribution of Y given $B^\top X$ for any $B \in \mathbb{R}^{p \times q}$. The definition of the tail-index dimension reduction subspace is given in Section 2.3

2.2 Preliminary result

For a given matrix $B \in \mathbb{R}^{p \times q}$ with $1 \leq q < p$, we can ask ourselves what happens if the covariate X is replaced by its projection $X_B := B(B^\top B)^{-1} B^\top X$ onto the linear subspace $\text{span}(B)$. Obviously, the knowledge of the conditional distribution of Y given X_B is equivalent to the one of Y given $B^\top X$. Since $S_B(y, B^\top X) = \mathbb{E}[S(y, X) \mid B^\top X]$, the conditional survival function of Y given $B^\top X$ can be seen as a mixture of the conditional survival function of Y given X . Hence, for all $x \in \mathcal{X}$, we are inclined to think that the decay of the conditional distribution of Y given $B^\top X = B^\top x$ is driven by the largest tail-index involved in the mixture i.e., by

$$\xi_B(B^\top x) := \max_{z: B^\top z = B^\top x} \gamma(z). \quad (2)$$

Note that $\xi_B(\cdot)$ is a measurable function corresponding to the right endpoint of the conditional distribution of $\gamma(X)$ given $B^\top X = \cdot$. Under additional assumptions on

the functions $\mathcal{L}(\cdot, \cdot)$ and $\gamma(\cdot)$, we prove hereafter that, as soon as model (M) holds, the conditional distribution of Y given $B^\top X$ is also heavy tailed with tail index $\xi_B(B^\top X)$. These additional conditions are given below. The first one is dedicated to the tail-index in model (M).

(C.1) The distribution of the random variable $\gamma(X)$ is absolutely continuous with respect to the Lebesgue measure. Moreover,

$$\underline{\gamma} := \inf_{x \in \mathcal{X}} \gamma(x) > 0 \text{ and } \bar{\gamma} := \sup_{x \in \mathcal{X}} \gamma(x) < \infty,$$

The second condition is related to the function $\mathcal{L}(\cdot, \cdot)$.

(C.2) There exists a positive measurable function $c(\cdot)$ for which

$$\limsup_{y \rightarrow \infty} \sup_{x \in \mathcal{X}} \left| \frac{\mathcal{L}(y, x)}{c(x)} - 1 \right| = 0, \text{ with } \inf_{x \in \mathcal{X}} c(x) > 0 \text{ and } \sup_{x \in \mathcal{X}} c(x) < \infty,$$

We are now in position to state a result that will be central to the definition of our dimension reduction subspace.

Proposition 1. *Assume that model (M) holds with conditions (C.1) and (C.2). For each $B \in \mathcal{B}_q$ with $1 \leq q < p$, the conditional distribution of Y given $B^\top X$ is heavy tailed with tail-index $\xi_B(B^\top X)$.*

Note that under (C.1), the tail-index of the conditional distribution of Y given $B^\top X$ belongs to the interval $[\underline{\gamma}, \bar{\gamma}] \subset (0, \infty)$.

2.3 Main definitions

We can now give the definition of the dimension reduction subspace that will be referred hereafter to as the TIDR subspace. Roughly speaking, if the tail-index in model (M) is such that $\gamma(\cdot) = g(B^\top \cdot)$ for some matrix $B \in \mathbb{R}^{p \times q}$ and some positive function $g(\cdot)$ then $\mathcal{S} = \text{span}(B)$ is a TIDR subspace. More precisely,

Definition 1. *Let $B \in \mathbb{R}^{p \times q}$ be a matrix with q linearly independent columns. The linear subspace $\mathcal{S} = \text{span}(B)$ is a tail-index dimension reduction (TIDR) subspace if $\xi_B(B^\top X) = \gamma(X)$ almost surely.*

The basis $B \in \mathbb{R}^{p \times q}$ such that $\mathcal{S} = \text{span}(B)$ is of course not unique. In all what follows, we use the canonical basis of \mathcal{S} which is defined hereafter.

Definition 2. Let $\tilde{B} \in \mathbb{R}^{p \times q}$ be any basis of \mathcal{S} and let \check{B}^\top be the matrix \tilde{B}^\top in its reduced row echelon form. The canonical basis B of \mathcal{S} is the orthogonal basis obtained by applying the Gram-Schmidt process on the columns of \check{B} .

It is well known that the reduced row echelon form of a given matrix is unique and thus so is the canonical basis. If $B \in \mathcal{B}_q$ then $B^\top B = I_q$, the identity matrix of dimension q . When $q = 1$, the elements of the set \mathcal{B}_1 are the vectors $\beta \in \mathbb{R}^p$ with a positive first nonzero entry for which $\beta^\top \beta = 1$. Note also that \mathcal{B}_q is a compact subset of $\mathbb{R}^{p \times q}$.

When $B = I_p$, it is readily seen that $\xi_B(B^\top x) = \gamma(x)$ for all $x \in \mathcal{X}$ and thus that $\mathcal{S} = \mathbb{R}^p$ is always a TIDR subspace. Of course, we are interested in finding the smallest subspace. More specifically, we are interested in finding the central tail-index subspace define as follows.

Definition 3. A linear subspace \mathcal{S}_0 is the central tail-index (CTI) subspace if \mathcal{S}_0 is a TIDR subspace such that $\mathcal{S}_0 \subset \mathcal{S}$ for all TIDR subspace \mathcal{S} .

This definition is a straightforward adaptation of the definition of the central mean subspace (see Cook and Li [5]). The CTI subspace does not necessarily exist but if it is the case, it is unique. We assume in the remainder of this paper that the random couple (X, Y) is such that the CTI subspace \mathcal{S}_0 exists. We also suppose that the dimension q of the CTI subspace is known.

In what follows, we denote by $B_0 \in \mathcal{B}_q$ the canonical basis of the CTI subspace \mathcal{S}_0 . As a consequence of Proposition 1, the conditional distribution of Y given $B_0^\top X$ (which is not necessarily equals to the one of Y given X) is heavy tailed with tail-index $\xi_{B_0}(B_0^\top X) = \gamma(X)$. Hence, assuming that B_0 is known, the conditional tail-index can be estimated using a sample of the random couple $(B_0^\top X, Y) \in \mathbb{R}^q \times \mathbb{R}$ instead of $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$. When $q \ll p$, this permits us to construct a more efficient estimator of $\gamma(\cdot)$. Of course in practice, B_0 is unknown. Its estimation is investigated in the next section.

3 Estimation of the CTI subspace

In all that section, we assume that the CTI subspace \mathcal{S}_0 exists and that its dimension q is known. The estimation of q will be investigated in Section 4.

3.1 General procedure

To estimate the canonical basis $B_0 \in \mathcal{B}_q$ of the CTI subspace, we start by remarking that for all $x \in \mathcal{X}$ and $B \in \mathcal{B}_q$, one has $\gamma(x) = \xi_{B_0}(B_0^\top x) \leq \xi_B(B^\top x)$, since $\gamma(x)$ belongs to the support of the conditional distribution of $\gamma(X)$ given $B^\top X = B^\top x$. The unicity of the CTI subspace entails that for all $B \in \mathcal{B}_q$ with $B \neq B_0$,

$$\mathbb{P} \left[\gamma(X) < \xi_B(B^\top X) \right] > 0, \quad (3)$$

leading to

$$\arg \min_{B \in \mathcal{B}_q} \mathbb{E} \left[\xi_B(B^\top X) \right] = \{B_0\}.$$

For technical reasons, the above expectation can be estimated consistently as soon as the density of X is bounded away from zero. We thus slightly modify the optimization problem by introducing a trimming factor $\mathbb{1}_{\mathcal{X}_0}(X)$ where \mathcal{X}_0 is a compact subset in the interior of \mathcal{X} . In other words, we now want to minimize the function $B \mapsto \mathbb{E} \left[\xi_B(B^\top X) \mathbb{1}_{\mathcal{X}_0}(X) \right]$ over the set \mathcal{B}_q . This modification was also adopted by Ichimura [11] for estimating the direction of a single-index model. To ensure that B_0 is still the only solution of the trimmed optimization problem, a slightly stronger condition than (3) is required.

(C.3) For all $B \in \mathcal{B}_q$ with $B \neq B_0$, $\mathbb{P} \left[\{ \gamma(X) < \xi_B(B^\top X) \} \cap \{ X \in \mathcal{X}_0 \} \right] > 0$,

Under (C.3), we have

$$\arg \min_{B \in \mathcal{B}_q} \mathbb{E} \left[\xi_B(B^\top X) \mathbb{1}_{\mathcal{X}_0}(X) \right] =: \arg \min_{B \in \mathcal{B}_q} \Psi(B, \mathcal{X}_0) = \{B_0\}.$$

Given a sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of independent copies of the random couple (X, Y) , the natural idea to estimate the canonical basis of the CTI subspace is to minimize over \mathcal{B}_q an estimation of the expectation $\Psi(B, \mathcal{X}_0)$. The empirical counterpart of $\Psi(B, \mathcal{X}_0)$ is given by

$$\tilde{\Psi}_n(B, \mathcal{X}_0) := \frac{1}{n} \sum_{i=1}^n \xi_B(B^\top X_i) \mathbb{1}_{\mathcal{X}_0}(X_i).$$

It can be shown that under model (M) where the tail-index satisfies condition (C.1) and is such that

(C.4) for all $x \in \mathcal{X}$, the function $B \mapsto \xi_B(B^\top x)$ is Lipschitz in $B \in \mathcal{B}_q$, that is to say there exists a measurable function $M : \mathcal{X} \rightarrow [0, \infty)$ such that $0 < \mathbb{E}(M(X)) < \infty$ and, for all $B_1 \in \mathcal{B}_q$, $B_2 \in \mathcal{B}_q$ and $x \in \mathcal{X}$,

$$\left| \xi_{B_1}(B_1^\top x) - \xi_{B_2}(B_2^\top x) \right| \leq M(x) \|B_1 - B_2\|,$$

then,

$$\sup_{B \in \mathcal{B}_q} \left| \tilde{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right| \xrightarrow{\mathbb{P}} 0,$$

see the proof of Lemma 2. Hence, given an estimator $\hat{\xi}_{n,B}(B^\top x)$ of $\xi_B(B^\top x)$, we propose to estimate $\Psi(B, \mathcal{X}_0)$ for all $B \in \mathcal{B}_q$ by the plug-in estimate

$$\hat{\Psi}_n(B, \mathcal{X}_0) := \frac{1}{n} \sum_{i=1}^n \hat{\xi}_{n,B}(B^\top X_i) \mathbb{1}_{\mathcal{X}_0}(X_i). \quad (4)$$

The definition of the proposed estimator of B_0 is given below.

Definition 4. Under model (M), for a sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of independent copies of the random couple (X, Y) , the estimator \hat{B}_n of the canonical basis of CTI subspace minimizes the function $B \mapsto \hat{\Psi}_n(B, \mathcal{X}_0)$.

3.2 Main result

The main result of this section establishes that if the conditional tail-index estimator is uniformly consistent i.e.,

$$(C.5) \quad \sup_{B \in \mathcal{B}_q} \sup_{x \in \mathcal{X}_0} \left| \hat{\xi}_{n,B}(B^\top x) - \xi_B(B^\top x) \right| \xrightarrow{\mathbb{P}} 0,$$

then the estimator \hat{B}_n in Definition 4 is consistent.

Theorem 1. Under model (M) with conditions (C.1) to (C.4), if the estimator $\hat{\Psi}_n(\cdot, \mathcal{X}_0)$ defined in (4) satisfies condition (C.5) then $\|\hat{B}_n - B_0\| \xrightarrow{\mathbb{P}} 0$ for any norm $\|\cdot\|$ in $\mathbb{R}^{p \times q}$.

In the next section, we propose an estimator of the conditional tail-index satisfying condition (C.5).

3.3 A local-Hill estimator of the CTI subspace

For $B \in \mathcal{B}_q$ and $x \in \mathcal{X}$, let us now propose an appropriate estimator of the conditional tail-index $\xi_B(B^\top x)$ of the conditional distribution of Y given $B^\top X = B^\top x$ (see Proposition 1) leading to an estimator of B_0 . To motivate the definition of our estimator, let us start with the unrealistic situation where we have at our disposal a sample $\{\check{Z}_i(B, x), i = 1, \dots, m\}$ of $m \in \mathbb{N} \setminus \{0\}$ independent random variables with common distribution the conditional distribution of Y given $B^\top X = B^\top x$. The associated order statistics are denoted $\check{Z}_{(1)}(B, x) < \dots < \check{Z}_{(n)}(B, x)$. The best known

estimator of the tail-index is the Hill estimator which is given by

$$\check{\xi}_B^{(H)}(B^\top x) := \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} \ln \frac{\check{Z}_{(m-i+1)}(B, x)}{\check{Z}_{(m-\lfloor \alpha m \rfloor)}(B, x)},$$

for some $\alpha \in (1/m, 1)$. Of course, this is not a proper estimator since the random variables $\{\check{Z}_i(B, x), i = 1, \dots, m\}$ are not observed. We therefore propose to replace them in the expression of $\check{\xi}_B^{(H)}(B^\top x)$ by a suitable set of observed random variables. More precisely, let us introduce the set,

$$\mathcal{T}(B, x, h) := \{z \in \mathcal{X} \mid \|B^\top z - B^\top x\| \leq h\},$$

where $h = h_n > 0$ and let

$$M^* = M(B, x, h) := \sum_{i=1}^n \mathbb{1}_{\mathcal{T}(B, x, h)}(X_i),$$

be the random number of covariates in the set $\mathcal{T}(B, x, h)$. We denote by $\{W_i^* = W_i(B, x, h), i = 1, \dots, M^*\}$ the set of covariates that belong to $\mathcal{T}(B, x, h)$. The associated response variables are denoted $\{Z_i^* = Z_i(B, x, h), i = 1, \dots, M^*\}$. Note that the presence of a star (\star) recalls the dependence on B, x and h . This shortcut will be used in the rest of that paper. Intuitively, when h is close enough to zero, the random variable Z_i^* is approximately distributed as $\check{Z}_i(B, x)$. For some sequence $(\alpha_n) = (\alpha) \in (0, 1)$, this leads us to introduce the estimator of the conditional tail-index defined below.

Definition 5. *Under model (M), for all $(B, x) \in \mathcal{B}_q \times \mathcal{X}$, the local-Hill estimator of the tail-index $\xi_B(B^\top x)$ is*

$$\widehat{\xi}_B^{(H)}(B^\top x) = \widehat{\xi}_B^{(H)}(B^\top x, \alpha, h) := \frac{1}{\lfloor \alpha M^* \rfloor} \sum_{i=1}^{\lfloor \alpha M^* \rfloor} \ln \frac{Z_{(M^*-i+1)}^*}{Z_{(M^*-\lfloor \alpha M^* \rfloor)}^*},$$

if $\alpha M^* > 1$ and $\widehat{\xi}_B^{(H)}(B^\top x, \alpha, h) = 0$ otherwise.

Note that this estimator is quite similar to the one introduced in Gardes and Stupfler [8]. Under additional assumptions given below, we show that this estimator is uniformly consistent over the set $\Theta := \mathcal{B}_q \times \mathcal{X}_0$. This set is endowed with the product distance δ defined for all $\theta = (B, x) \in \Theta$ and $\theta' = (B', x') \in \Theta$ by $\delta(\theta, \theta') := \|B - B'\| + \|x - x'\|$. Below, we give the assumptions required to establish the uniform consistency. Note that these assumptions are more or less the same that the ones used in [8].

(H.1) The distribution of X is absolutely continuous with respect to the Lebesgue measure.

Note that since $B \in \mathcal{B}_q$ is a full rank matrix, condition (H.1) entails that for all $B \in \mathcal{B}_q$, the distribution of $B^\top X$ is also absolutely continuous. Let us denote by $f_B(\cdot)$ the associated density function.

(H.2) For all $B \in \mathcal{B}_q$ and $x \in \mathcal{X}_0$, $f_B(B^\top x) > 0$. Moreover, there exists a positive constant K_f such that for all $(B, B') \in \mathcal{B}_q^2$ and $(x, x') \in \mathcal{X}^2$,

$$\sup_{B \in \mathcal{B}_q} \left| f_B(B^\top x) - f_{B'}([B']^\top x') \right| \leq K_f (\|B - B'\| + \|x - x'\|).$$

Since Θ is compact, this condition entails that

$$\sup_{(B,x) \in \Theta} f_B(B^\top x) < C_2 < \infty \text{ and } \inf_{(B,x) \in \Theta} f_B(B^\top x) > C_1 > 0.$$

(H.3) The function $(B, x) \in \mathcal{B}_q \times \mathcal{X} \mapsto \xi_B(B^\top x)$ is continuous.

For all $B \in \mathcal{B}_q$ and $z \in B^\top \mathcal{X}$, let $Q_B(\cdot, z)$ be the inverse function of the function $y \mapsto S_B(y, z)$. From Proposition 1, we have for almost all $z \in B^\top \mathcal{X}$ and $u \in (0, 1)$ that

$$Q_B(u, z) = u^{-\xi_B(z)} \ell_B(u^{-1}, z),$$

where for all $B \in \mathcal{B}_q$ and $z \in B^\top \mathcal{X}$, $\ell_B(\cdot, z)$ is a slowly varying function. The following condition on $\ell_B(\cdot, \cdot)$ is required. With the notation

$$\Delta_{B,x}(y) := \sup_{t>1} \ln \left(\frac{\ell_B(ty|B^\top x)}{\ell_B(y|B^\top x)} \right),$$

we assume that

(H.4) $\lim_{y \rightarrow \infty} \sup_{(B,x) \in \mathcal{B}_q \times \mathcal{X}} \Delta_{B,x}(y) = 0$.

Finally, for all $B \in \mathcal{B}_q$, we measure the oscillation of the function $x \mapsto Q_B(\cdot, B^\top x)$ by

$$\omega(u, v, B, x, h) = \omega^*(u, v) := \sup_{(\beta, x') \in \mathcal{D}_{u,v}^*} \left| \ln \left(\frac{Q_B(\beta, B^\top x)}{Q_B(\beta, B^\top x')} \right) \right|$$

where

$$\mathcal{D}_{u,v}^* = \mathcal{D}_{u,v}(B, x, h) = \left\{ (b, y) \in (0, 1) \times \mathcal{X}; b \in [u, v] \text{ and } \|B^\top(x - y)\| \leq h \right\}.$$

The oscillation is controlled by following assumption.

(H.5) There exists $\kappa > 0$ such that

$$\lim_{n \rightarrow \infty} \sup_{(B,x) \in \mathcal{B}_q \times \mathcal{X}} \omega^* \left(n^{-(1+\kappa)}, 1 - n^{-(1+\kappa)} \right) = 0.$$

We establish below the uniform consistency of the local-Hill estimator given in Definition 5.

Theorem 2. *Under model (M) and conditions (C.1), (H.1)-(H.5), if the sequences (α) and (h) are such that $nh^q \alpha / \ln(n) \rightarrow \infty$ and $\alpha \ln(nh^q) \rightarrow 0$ as $n \rightarrow \infty$ then,*

$$\sup_{(B,x) \in \Theta} \left| \widehat{\xi}_B^{(H)}(B^\top x) - \xi_B(B^\top x) \right| \xrightarrow{\mathbb{P}} 0.$$

Since the proof of Theorem 2 is quite long and technical, it is postponed to the Appendix. The scheme of the proof is similar to the one of [8, Theorem 1] with however some additional difficulties due to the necessity to prove uniformity on \mathcal{B}_q . Using the local-Hill estimator of the conditional tail-index we define the local-Hill CTI estimator of B_0 as follows.

Definition 6. *Under model (M), for a sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of independent copies of the random couple (X, Y) , the local Hill estimator $\widehat{B}_n^{(H)}$ of the canonical basis of CTI subspace minimizes the function*

$$B \mapsto \widehat{\Psi}_n^{(H)}(B, \mathcal{X}_0) := \frac{1}{n} \sum_{i=1}^n \widehat{\xi}_B^{(H)}(B^\top X_i) \mathbb{I}_{\mathcal{X}_0}(X_i).$$

The consistency of $\widehat{B}_n^{(H)}$ follows directly from Theorems 1 and 2.

Corollary 1. *Under model (M), conditions (C.1)-(C.4) and (H.1)-(H.5), if the sequences (α) and (h) are such that $nh^q \alpha / \ln(n) \rightarrow \infty$ and $\alpha \ln(nh^q) \rightarrow 0$ as $n \rightarrow \infty$ then, $\|\widehat{B}_n^{(H)} - B_0\| \xrightarrow{\mathbb{P}} 0$.*

4 Finite sample size behavior

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent copies of a random couple (X, Y) for which there exists a CTI subspace \mathcal{S}_0 with canonical basis B_0 . In this section, we are interested in the finite sample behavior of two estimators: the local-Hill CTI estimator $\widehat{B}_n^{(H)}$ of B_0 , see Definition 6, and the estimator $\widehat{\xi}_{\widehat{B}_n^{(H)}}^{(H)}([\widehat{B}_n^{(H)}]^\top x)$ of the conditional tail-index $\gamma(x)$. We start by assuming that the dimension q of \mathcal{S}_0 is known. Its estimation is investigated in Section 4.5. This Section is organized as follows. Section 4.1

is devoted to the choice of the tuning parameters required to compute these estimators. We explain in Section 4.2 the procedure we adopted to find the solution to the optimization problem involved in the definition of the local-Hill CTI estimator. Different models for the distribution of (X, Y) are given in Section 4.3 and the simulation results are commented in Section 4.4. Finally, a way to estimate the dimension q of the CTI subspace is presented in Section 4.5.

4.1 Selection of the tuning parameters

The local-Hill CTI estimator depends on the choice of a compact set \mathcal{X}_0 included in the unknown support \mathcal{X} , see Equation (4). We suggest to take

$$\mathcal{X}_0 := \{x \in \mathcal{X} \mid \text{HD}(x) \geq \eta\},$$

where $\eta \in (0, 1)$ and $\text{HD}(x)$ is the multivariate Tukey's depth function given for all $x \in \mathcal{X}$ by $\text{HD}(x) := \min\{\mathbb{P}_X(H_{\theta,x}) \mid \theta \in \mathbb{R}^p\}$, with $H_{\theta,x}$ the half-space in the direction of θ passing through the point x . The value $\text{HD}(x) \in [0, 1/2]$ can be interpreted as a distance between x and the center of the distribution: the point x is close to the center of the distribution when $\text{HD}(x)$ is close to $1/2$. In practice, the distribution \mathbb{P}_X is unknown. The subspace \mathcal{X}_0 is thus replaced by $\widehat{\mathcal{X}}_0 := \{x \in \mathcal{X} \mid \widehat{\text{HD}}_n(x) \geq \eta\}$, where $\widehat{\text{HD}}_n(x)$ is the empirical estimator of $\text{HD}(x)$. The level η is arbitrarily chosen such that 20% of the sample lies within $\widehat{\mathcal{X}}_0$.

Two other tuning parameters are required to compute the local-Hill CTI estimator. First of all, a bandwidth h must be chosen to select the observations closest to the point of interest x . According to Lemma 5 in the Appendix, the number of selected observations is asymptotically proportional to $nh^q\vartheta_q$, where ϑ_q is the volume of the unit q -ball. In all what follows, we use the uniform norm $\|\cdot\|_\infty$ for which $\vartheta_q = 2^q$. Denoting by $\text{diam}(\mathcal{X}) := \sup\{\|x - y\|_\infty \mid (x, y) \in \mathcal{X}^2\}$ the diameter of the support \mathcal{X} , we propose to take $h = \text{diam}(\mathcal{X}) \times (n^{-b}/\vartheta_q)^{1/q}$ for some $b > 0$. Note that the bandwidth h increases with the diameter of \mathcal{X} . Next, we have to choose a probability $\alpha \in (0, 1)$ controlling the number of largest observations used in the estimation procedure. Proposing a method to select the parameters h (or equivalently b) and α is beyond the scope of this paper and left for future work. However, based on numerous simulations, it seems that our estimation procedure is relatively robust to a non optimal choice. We have decided to take $b = 2/9$ and $\alpha = n^{-3/10}$ which provide good

results in most situations.

4.2 Finding the solution of the optimization problem

Recall that the estimator $\widehat{B}_n^{(H)}$ is solution of the constrained optimization problem

$$(P) \quad \min_{B \in \mathbb{R}^{p \times q}} \widehat{\Psi}_n^{(H)}(B, \mathcal{X}_0) \quad \text{subject to} \quad B \in \mathcal{B}_q.$$

Searching for the solution by moving along the feasible set is rather difficult. In particular, the solution strongly depends on the initialization of the algorithm. We thus decided to relax the constraint by considering the problem

$$(P') \quad \min_{B \in \mathbb{R}^{p \times q}} \widehat{\Psi}_n^{(H)}(B, \mathcal{X}_0) \quad \text{subject to} \quad B^\top B - I_q = 0.$$

Since \mathcal{B}_q is the set of all canonical q -dimensional basis, each element of \mathcal{B}_q corresponds to a different subspace of \mathbb{R}^p . Of course, this is no longer the case if we only impose B to be an orthogonal matrix but (P') is easier to solve than (P) . Moreover, if B (resp. B') is a solution of (P) (resp. (P')), then

$$\widehat{\Psi}_n^{(H)}(B', \mathcal{X}_0) \leq \widehat{\Psi}_n^{(H)}(B, \mathcal{X}_0).$$

To solve the constrained optimization problem (P') , we apply an algorithm in the class of penalty methods. More specifically, let $(P_j)_{j \in \mathbb{N}}$ be a sequence of penalized (unconstrained) problems with

$$(P_j) \quad \min_{B \in \mathbb{R}^{p \times q}} \widehat{\Psi}_n^{(H)}(B, \mathcal{X}_0) + c_j \times g(B),$$

where (c_j) is an increasing sequence converging to infinity and $g : \mathbb{R}^{p \times q} \rightarrow \mathbb{R}$ is a penalty function i.e., a function such that $g(B) = 0$ if B is feasible ($B^\top B = I_q$) and $g(B) > 0$ otherwise. For an arbitrarily chosen matrix $B_{(0)} \in \mathbb{R}^{p \times q}$, we optimize (P_0) by using the Particle Swarm Optimization (PSO) algorithm with its first particle initialized to $B_{(0)}$. Denoting by $B_{(1)}$ the obtained solution, we next optimize (P_1) with the PSO algorithm where the first particle is initialized to $B_{(1)}$. This process is repeated until convergence. More details on the PSO algorithm can be found in Shami et al. [17] and in Wang et al. [20]. In our simulation study, this procedure is implemented as follows. For the initialization, the matrix $B_{(0)}$ is randomly chosen by generating independently each of its components uniformly over $[-1, 1]$. For the penalty function we opt for the natural choice $g(B) = \|B^\top B - I_q\|_F^2$, where $\|\cdot\|_F$ is the Frobenius norm, with the penalty sequence $c_j = 2^j \times 10^{-2}$. The maximal number iterations for the PSO algorithm is 100 and the convergence is assumed to be achieved when $\|B_{(j)} - B_{(j+1)}\|_F \leq 10^{-3} \sqrt{pq}$ where $B_{(j+1)}$ is the solution of (P_j) .

4.3 Models

Four models are considered for the distribution of the random couple (X, Y) . The random vector X is uniformly distributed on the space $\mathcal{X} := [0, 1]^p$ with $p \in \{4, 16\}$. We have here $\text{diam}(\mathcal{X}) = 1$. Let $Q(\cdot, x)$ be the inverse of the function $y \mapsto S(y, x)$. The random variable Y is given by $Y := U^{-\gamma(x)} \ell(U^{-1}, x)$ where U is a standard uniform random variable. Let $B_1 = (0, 0, 5, 5, 0, \dots, 0)^\top \in \mathbb{R}^p$. Two different slowly varying functions are used in our simulation study:

$$\ell^{(1)}(u^{-1}, x) := [1 + \exp\{B_1^\top x - u^{-1}\}]^{-1} \text{ and } \ell^{(2)}(u^{-1}, x) := \exp(-u/2) B_1^\top x.$$

For the first three models, the dimension of the CTI subspace is $q = 1$ with canonical basis $B_0 = (2, 1, 0, \dots, 0)^\top / \sqrt{5} \in \mathbb{R}^p$. Two functions $\gamma(\cdot) = \xi_{B_0}(B_0^\top \cdot)$ are considered: for all $z \in \mathbb{R}$,

$$\xi_{B_0}^{(1)}(z) := \frac{1}{10} + \frac{9}{10} \left(\frac{\sqrt{5}z}{3} \right)^4 \text{ and } \xi_{B_0}^{(2)}(z) := \frac{1}{10} + \frac{9}{10} |\cos(2z)|.$$

- **Model 1** - For all $x \in \mathcal{X}$, $\gamma(x) = \xi_{B_0}^{(1)}(B_0^\top x)$ and $\ell(u^{-1}, x) = \ell^{(1)}(u^{-1}, x)$.

For this model, since $\ell^{(1)}(u^{-1}, x) \rightarrow 1$ as $u \rightarrow 0$, the basis B_0 also satisfies condition (1) that is to say that the CTI subspace corresponds to the dimension reduction subspace introduced in Gardes [7]. For the next model, we consider the slowly varying function $\ell^{(2)}(u^{-1}, x)$ which converges to $B_1^\top x$ as $u \rightarrow 0$. As a consequence, the CTI subspace is still $\text{span}(B_0)$ while the dimension reduction subspace in [7] is $\text{span}(B_0, B_1)$.

- **Model 2** - For all $x \in \mathcal{X}$, $\gamma(x) = \xi_{B_0}^{(1)}(B_0^\top x)$ and $\ell(u^{-1}, x) = \ell^{(2)}(u^{-1}, x)$.

In Models 1 and 2, the conditional tail index $\xi_{B_0}^{(1)}(\cdot)$ is strictly monotone. The function $\xi_{B_0}^{(2)}(\cdot)$, which is symmetric with respect to the vertical line $y = \pi/4$, is used in Model 3.

- **Model 3** - For all $x \in \mathcal{X}$, $\gamma(x) = \xi_{B_0}^{(2)}(B_0^\top x)$ and $\ell(u^{-1}, x) = \ell^{(1)}(u^{-1}, x)$.

For the last model, the dimension of the CTI linear subspace is $q = 2$ with canonical basis $B_0 = (e_1, e_2) \in \mathbb{R}^{p \times 2}$ where $e_1 = (1, 0, \dots, 0)^\top$ and $e_2 = (0, 1, 0, \dots, 0)^\top$.

- **Model 4** - For all $x \in \mathcal{X}$, $\ell(u^{-1}, x) := \ell^{(1)}(u^{-1}, x)$ and

$$\gamma(x) = \xi_{B_0}(B_0^\top x) := \frac{1}{10} + \frac{9}{5} \left[(e_1^\top x - 0.5)^2 + (e_2^\top x - 0.5)^2 \right].$$

Let us mention that for Model 4 with $p = 16$, the optimization procedure described in Section 4.2 is repeated 10 times for different initialization matrix $B_{(0)}$. We kept the initialization providing the smallest value of $\widehat{\Psi}_n^{(H)}(\widehat{B}_n, \mathcal{X}_0)$.

4.4 Simulation results

For each model, we generate $N = 100$ samples of size $n = 2000$. We first compare the local-Hill CTI estimator $\widehat{B}_n^{(H)}$, see Definition 6, with the TDR estimator proposed by Gardes [7]. This estimator is denoted $\widehat{B}_n^{(G)}$. Note that $\widehat{B}_n^{(H)}$ is consistent only if $\gamma(x) > 0$ for all $x \in \mathcal{X}$ while $\widehat{B}_n^{(G)}$ is consistent regardless the sign of $\gamma(x)$. As the execution time to obtain the estimator of Gardes [7] is important, we only make the comparison in the case $p = 4$. Also, we choose not to compare our estimator to those proposed in Aghbalou et al. [1] mainly because we cannot check if the condition used in the definition of their dimension reduction subspace is satisfied or not by the models of Section 4.3. The performance of a given estimator \check{B}_n of B_0 is measured by the distance

$$\frac{1}{\sqrt{pq}} \|\check{B}_n - B_0\|_F = \left(\frac{1}{pq} \sum_{i=1}^p \sum_{j=1}^q \widehat{\delta}_{i,j}^2 \right)^{1/2},$$

where $\widehat{\delta}_{i,j}$ is the element of row i and column j of the matrix $\check{B}_n - B_0$. The division by \sqrt{pq} ensures that the above distance does not necessarily increases with p and q . To implement the estimator $\widehat{B}_n^{(G)}$, we have to choose a bandwidth h to select the most pertinent observations and a number k of largest observations. We propose to select these tuning parameters as in Gardes [7], namely $h = n^{-2/9}$ and $\alpha = n^{-1/3}$. The TDR estimator in Gardes [7] is also the solution of a constrained optimization problem which is solved with the procedure described in Section 4.2.

The boxplots of the errors are given in Figure 1 for $p = 4$. For all the models, the local-Hill estimator provides the best results. For Model 2, where the CTI subspace is included in the dimension reduction subspace introduced in [7], the estimation of B_0 by $\widehat{B}_n^{(G)}$ is inaccurate. This is in line with the theory. For Model 3 where the function $\xi_{B_0}(\cdot)$ is symmetric with respect to the vertical line $y = \pi/4$, $\widehat{B}_n^{(G)}$ is clearly less efficient. This phenomenon is also observed with the SIR method which fails to recover the true direction in presence of a symmetric relationship. Note that Gardes' estimator failed to estimate the dimension reduction subspace in Model 4 for which $q = 2$.

The finite sample behavior of the local-Hill CTI estimator is also investigated for a larger dimension ($p = 16$). We can check in Table 1 that its accuracy seems not to be affected by the dimension of the covariate.

Model	1	2	3	4
$p = 4$	0.061 (0.024)	0.235 (0.066)	0.035 (0.017)	0.037 (0.014)
$p = 16$	0.055 (0.012)	0.138 (0.022)	0.068 (0.043)	0.083 (0.048)

Table 1: Average of the errors of the local-Hill estimator on $N = 100$ replications of a sample of size $n = 2000$ from Models 1 to 4 with $p = 4$ and $p = 16$. The standard deviation is given in parentheses.

We are now interested in the effect of the dimension reduction on the estimation of the tail index. For that purpose, we look at the finite sample behavior of the local-Hill estimator $\widehat{\xi}_{\check{B}_n}^{(H)}(\check{B}_n^\top x)$ where \check{B}_n is one of the two estimators of the CTI subspace considered above. To demonstrate the need of reducing the dimension, these tail index estimators are compared to the estimator $\widehat{\xi}_{I_p}^{(H)}(I_p x)$ which is the estimator of $\gamma(x)$ obtained without prior dimension reduction. Finally, we also look at the behavior of the tail index 'estimator' $\widehat{\xi}_{B_0}^{(H)}(B_0^\top x)$ that can only be computed in the unrealistic case where the true CTI subspace is known. All these estimators are compared according to their mean squared error

$$E_\gamma(\check{B}_n) := \mathbb{E} \left[\left(\widehat{\xi}_{\check{B}_n}^{(H)}(\check{B}_n^\top X) - \gamma(X) \right)^2 \right].$$

The expectation $E_\gamma(\check{B}_n)$ is approximated with a Monte Carlo method by generating 1000 uniform random variables X on \mathcal{X} .

The average and the standard deviation over the $N = 100$ replications are gathered in Tables 2 and 3. As expected from the previous results, the estimation of the extreme value index based on the local-Hill CTI estimator is better for all the models. Also, the error $E_\gamma(I_p)$ is always larger than $E_\gamma(\widehat{B}_n^{(H)})$ showing the gain of the dimension reduction procedure. As expected, this gain is more important when $p = 16$.

4.5 Estimation of the dimension of the CTI subspace

In the above simulations, the true dimension q of the CTI subspace was used. In practice, this dimension is, in most of the cases, unknown and need to be estimated.

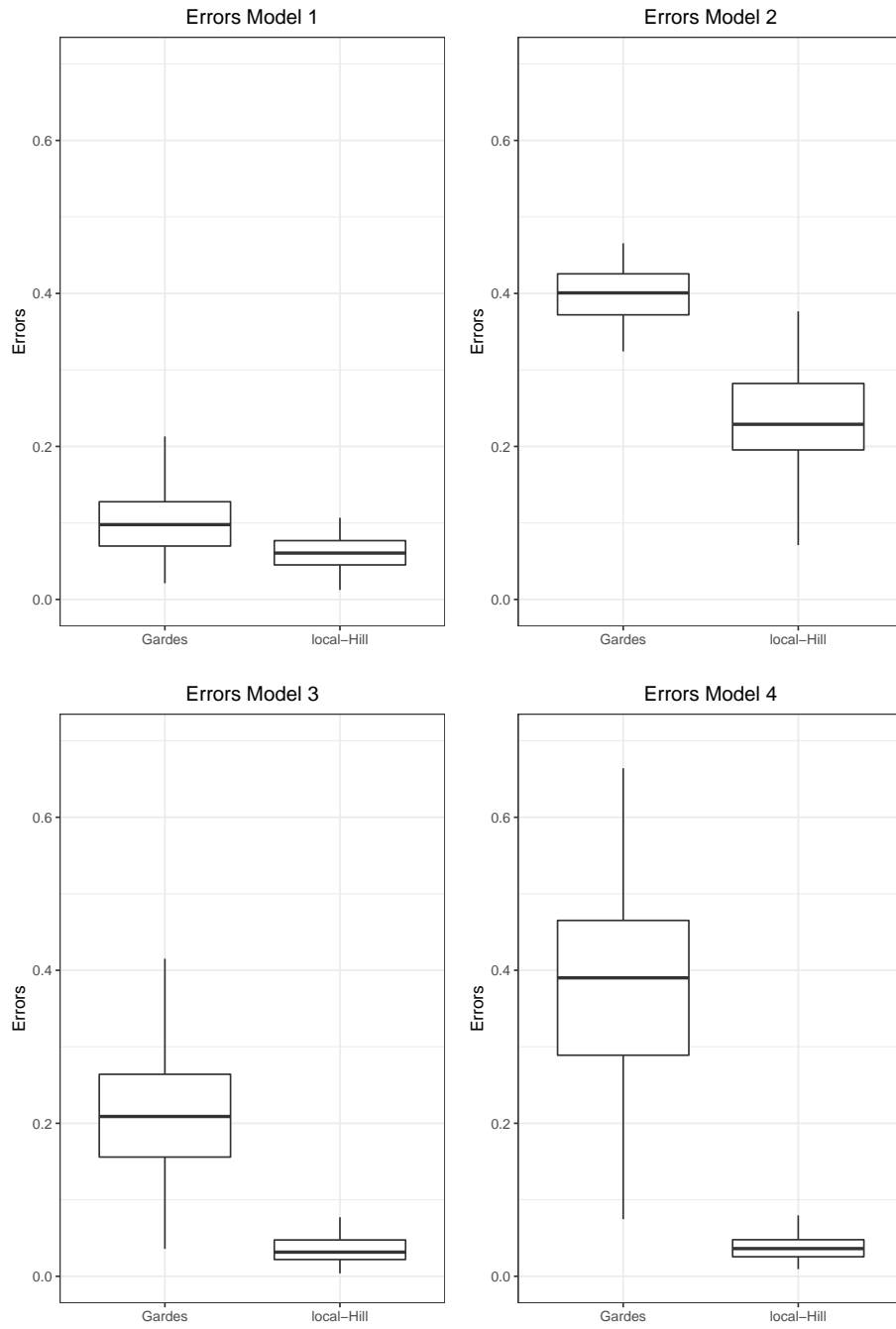


Figure 1: Comparison of the local-Hill estimator and the estimator $\widehat{B}_n^{(G)}$ on $N = 100$ replications of a sample of size $n = 2000$ from Models 1 to 4 with $p = 4$.

Model	$E_\gamma(\widehat{B}_n^{(H)})$	$E_\gamma(\widehat{B}_n^{(G)})$	$E_\gamma(B_0)$	$E_\gamma(I_p)$
1	0.015 (0.016)	0.021 (0.021)	0.008 (0.006)	0.100 (0.037)
2	0.018 (0.008)	0.027 (0.008)	0.011 (0.006)	0.026 (0.006)
3	0.064 (0.018)	0.122 (0.062)	0.031 (0.015)	0.221 (0.046)
4	0.048 (0.017)	0.133 (0.053)	0.049 (0.011)	0.168 (0.040)

Table 2: Average of the mean squared errors $E_\gamma(\check{B}_n)$ for $p = 4$. The standard deviation is given in parentheses.

Model	$E_\gamma(\widehat{B}_n^{(H)})$	$E_\gamma(B_0)$	$E_\gamma(I_p)$
1	0.016 (0.013)	0.008 (0.006)	2.888 (0.256)
2	0.021 (0.008)	0.011 (0.006)	0.230 (0.040)
3	0.065 (0.042)	0.031 (0.015)	3.585 (0.347)
4	0.059 (0.030)	0.049 (0.011)	3.151 (0.297)

Table 3: Average of the mean squared errors $E_\gamma(\check{B}_n)$ for $p = 16$. The standard deviation is given in parentheses.

To do so, for all $d \in \{1, \dots, p\}$, let $B_{0,d} \in \mathcal{B}_d$ be a matrix such that $\Psi(B_{0,d}, \mathcal{X}_0) \leq \Psi(B, \mathcal{X}_0)$ for all $B \in \mathcal{B}_d$ and let us introduce the quantity,

$$c(d, \mathcal{X}_0) := \Psi(B_{0,d}, \mathcal{X}_0).$$

We have the following result.

Proposition 2. *Assume that (C.3) holds. For all $d \in \{1, \dots, q-1\}$, one has $c(d+1, \mathcal{X}_0) \leq c(d, \mathcal{X}_0)$ and for all $d \in \{q, \dots, p\}$, $c(d, \mathcal{X}_0) = \mathbb{E}[\gamma(X)\mathbb{I}_{\mathcal{X}_0}(X)]$.*

Unfortunately, if there exists $d \in \{1, \dots, q-1\}$ such that $c(d+1, \mathcal{X}_0) = c(d, \mathcal{X}_0)$, Proposition 2 does not permit us to characterize the dimension q . The proof of Proposition 2 mainly consists in showing that there exists a non empty set \mathcal{B}_{d+1}^* such that $\xi_B(B^\top x) \leq \xi_{B_{0,d}}(B_{0,d}^\top x)$ for all $d \in \{1, \dots, p-1\}$, $B \in \mathcal{B}_{d+1}^*$ and $x \in \mathcal{X}$. The natural condition ensuring that the function $d \mapsto c(d, \mathcal{X}_0)$ is decreasing for $d < q$ is given below.

(A) For all $d \in \{1, \dots, q-1\}$, there exists $B \in \mathcal{B}_{d+1}^*$ such that

$$\mathbb{P} [\{\xi_B(B^\top X) < \xi_{B_{0,d}}(B_{0,d}^\top X)\} \cap \{X \in \mathcal{X}_0\}] > 0.$$

Since under (A) and (C.3), the function $d \mapsto c(d, \mathcal{X}_0)$ is decreasing for $d < q$ and constant for $d \geq q$ we get that $q = \min\{d \in \{1, \dots, p\}; c(d, \mathcal{X}_0) \leq c(d+1, \mathcal{X}_0)\}$. We thus naturally propose to estimate the unknown dimension by

$$\hat{q}_n := \min\{d \in \{1, \dots, p\}; \hat{c}_n(d, \mathcal{X}_0) \leq \hat{c}_n(d+1, \mathcal{X}_0)\},$$

where

$$\hat{c}_n(d, \mathcal{X}_0) = \min_{B \in \mathcal{B}_d} \hat{\Psi}_n^{(H)}(B, \mathcal{X}_0).$$

As shown in Table 4, for Models 1, 3 and 4 with $p = 4$, the $N = 100$ replications of \hat{q}_n are always equal to the true dimension q of the CTI subspace. The worst result is obtained for Model 1 with $p = 16$ (see Table 5) where our estimator failed to recover the true dimension in 14% of the replications.

We also look at the finite sample behavior of the estimator $\hat{\xi}_{\hat{B}_n^{(d,H)}}([\hat{B}_n^{(d,H)}]^\top x)$ of $\gamma(x)$ where

$$\hat{B}_n^{(d,H)} = \arg \min_{B \in \mathcal{B}_d} \hat{\Psi}_n^{(H)}(B, \mathcal{X}_0).$$

As expected, the smallest mean squared error $E_\gamma(\hat{B}_n^{(d,H)})$ is observed when d is the true dimension but for all the consider models, using the estimated dimension \hat{q}_n instead of the true dimension provides very similar results.

Model	Estimation of q			$E_\gamma(\widehat{B}_n^{(d,H)})$			
	$\widehat{q}_n = 1$	$\widehat{q}_n = 2$	$\widehat{q}_n = 3$	$d = 1$	$d = 2$	$d = 3$	$d = \widehat{q}_n$
1 ($q = 1$)	100 %	0 %	0 %	0.015	0.035	0.053	0.015
2 ($q = 1$)	97 %	3 %	0 %	0.018	0.018	0.020	0.018
3 ($q = 1$)	100 %	0 %	0 %	0.035	0.105	0.152	0.035
4 ($q = 2$)	0 %	100 %	0 %	0.067	0.048	0.097	0.048

Table 4: Left part of the table: estimation of the CTI subspace dimension over $N = 100$ replications of a sample of size $n = 2000$ generated from Models 1 to 4 with $p = 4$. Right part of the table: mean squared error of the local-Hill tail index estimator for different values of d .

Model	Estimation of q			$E_\gamma(\widehat{B}_n^{(d,H)})$			
	$\widehat{q}_n = 1$	$\widehat{q}_n = 2$	$\widehat{q}_n = 3$	$d = 1$	$d = 2$	$d = 3$	$d = \widehat{q}_n$
1 ($q = 1$)	86 %	14 %	0 %	0.026	0.048	0.066	0.031
2 ($q = 1$)	89 %	11 %	0 %	0.022	0.027	0.032	0.022
3 ($q = 1$)	94 %	5 %	1 %	0.065	0.125	0.176	0.063
4 ($q = 2$)	0 %	98 %	2 %	0.101	0.059	0.103	0.059

Table 5: Left part of the table: estimation of the CTI subspace dimension over $N = 100$ replications of a sample of size $n = 2000$ generated from Models 1 to 4 with $p = 16$. Right part of the table: mean squared error of the local-Hill tail index estimator for different values of d .

5 Application to a real data set

As the world's economies and industries develop, people are increasingly exposed to pollution, especially in urban areas. Some pollutants, such as ozone, can cause serious health problems. It forms under certain meteorological conditions (temperature, humidity, etc.) from primary pollutants emitted directly into the atmosphere by human activities. The aim here is to identify the primary pollutants that can lead to extreme ozone concentrations.

Our study is based on the record of $n = 4841$ daily concentration pollutants in Chicago from 1987 to 2000. These data were obtained few years ago from the Internet-based Health and Air Pollution Surveillance System (iHAPSS) website but unfortunately they are no longer available. The data set provides n observations of the maximum daily concentration of ozone (O_3) and of several primary pollutants including the maximum daily concentration of nitrogen dioxide (NO_2), sulfur dioxide (SO_2), particulate matter with a diameter smaller than 10 micrometers (PM_{10}) and carbon monoxide (CO). This data set was considered by many authors to illustrate the effect of dimension reduction (see, for example, [7], [16] and [22]).

The response variable is given by $Y = \exp(O_3/\sigma_{O_3})$ where $\sigma_{O_3}^2$ is the variance of O_3 . In practice this variance is replaced by its empirical counterpart. The covariate $X = (X_1, \dots, X_p)^\top$ with $p = 4$ is the centered and normalized (empirical) version of the random vector (NO_2, SO_2, PM_{10}, CO). The exponential function is applied to the maximum daily concentration of ozone in order to obtain a heavy-tailed distributed random variable. By looking at the Pareto quantile plot $\{(\ln(k/i), \ln(Y_{n-i,n}/Y_{n-k,n})); i \in \{1, \dots, k = 200\}\}$, see Figure 2, we can see that the points are approximately located on a straight line which is what it is expected for a heavy-tailed distribution. It should be noted that we should check that the conditional distribution of Y given $X = x$ is heavy-tailed for all x in the support of X . This task cannot be done in practice, but the fact that Y is heavy-tailed allows us to hope that the same is true for conditional distributions.

We apply the dimension reduction procedure proposed in this paper to this data set. The set $\hat{\mathcal{X}}_0$ is obtained as in the simulation study. Applying the procedure described in Section 4.5, we find that $\hat{q}_n = 1$ (with $\hat{c}_n(1, \hat{\mathcal{X}}_0) = 0.531$ and $\hat{c}_n(2, \hat{\mathcal{X}}_0) = 0.594$). The tuning parameters are set to $\alpha = n^{-3/10}$ and $h = \text{diam}(\mathcal{X}) \times (n^{-b}/2^{\hat{q}_n})^{1/\hat{q}_n}$ with $b = 2/9$ and $\text{diam}(\mathcal{X}) := \max\{\|x_i - x_j\|_\infty \mid (i, j) \in \{1, \dots, n\}^2\} \approx 15$, where x_i is

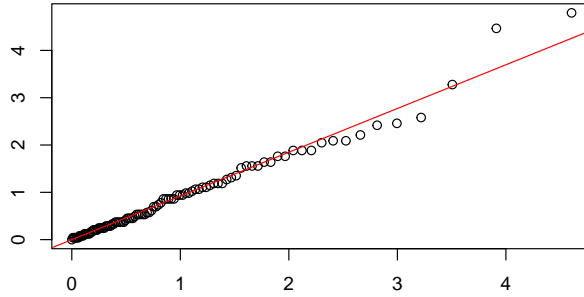


Figure 2: Pareto quantile plot (x-axis: $\ln(k/i)$, y-axis: $\ln(Y_{n-i,n}/Y_{n-k,n})$ for $i \in \{0, \dots, k = 200\}$. The fitted line is depicted in red.)

the observation of the covariate X for day i . By applying the minimization procedure described in Section 4.2, the estimated basis of the CTI subspace is given by

$$\widehat{B}_n = (0.757, 0.032, 0.608, 0.239)^\top.$$

Note that the two largest components of \widehat{B}_n correspond to NO_2 and PM_{10} . In Figure 3, the $(100 \times \alpha)\%$ largest observations of O_3 are represented versus the projection of the covariate unto the estimated CTI subspace. It appears that an increase in the extreme values of ozone is associated with increasing values of the 1-dimensional reduced covariate that is to say with increasing values of NO_2 and PM_{10} .

We can now estimate the function $x \in \mathcal{X} \mapsto \gamma(x)$ of conditional extreme value indices by $\widehat{\gamma}_n(x) := \widehat{\xi}_{\widehat{B}_n}^{(H)}(\widehat{B}_n^\top x)$. The estimated values $\widehat{\gamma}_n(x_i)$ for $x_i \in \widehat{\mathcal{X}}_0$ as a function of $\widehat{B}_n^\top x_i$ are depicted in Figure 4. This figure shows a positive correlation between the tail-index and the CTI direction. As a conclusion, it seems that the magnitude of extreme ozone concentrations increases with the concentrations of NO_2 and PM_{10} . Note that a study of this data set was also conducted by Scrucca [16]. The author concluded that ozone concentration (not necessarily extreme) mostly depends on the level of NO_2 . Our study shows that PM_{10} becomes another important factor for extreme ozone concentration.

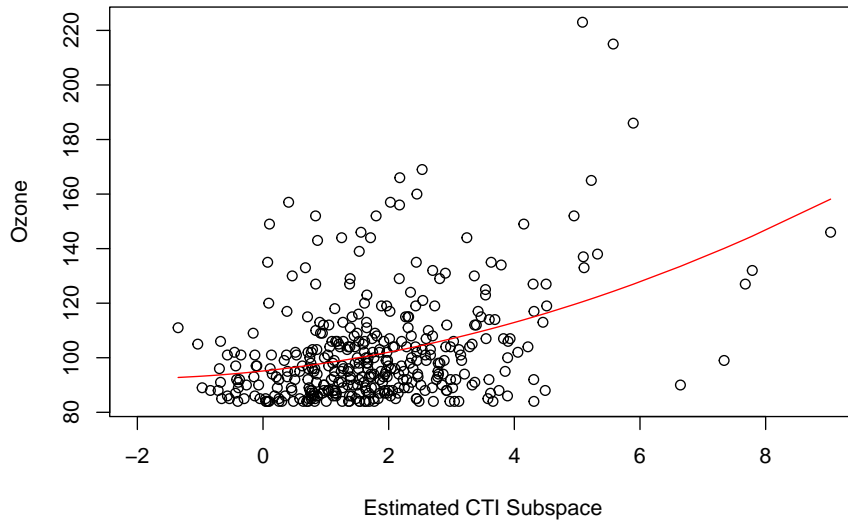


Figure 3: Largest observations of O_3 versus the projected covariates on the linear subspace spanned by \hat{B}_n . The smooth curve is obtained by fitting a polynomial of order 2.

Conclusion

When the tail-index function is positive and defined on a linear subspace of lower dimension (referred to as the CTI subspace), an estimation procedure of a basis of this subspace is proposed in this paper. Its weak consistency is established and a simulation study is conducted showing that our method is more efficient than the one introduced in Gardes [7]. In addition, we propose a method for estimating the unknown dimension of the dimension reduction subspace which gives satisfactory results on synthetic data. The implementation of the estimation procedure described in this paper in an **R** package is a work in progress.

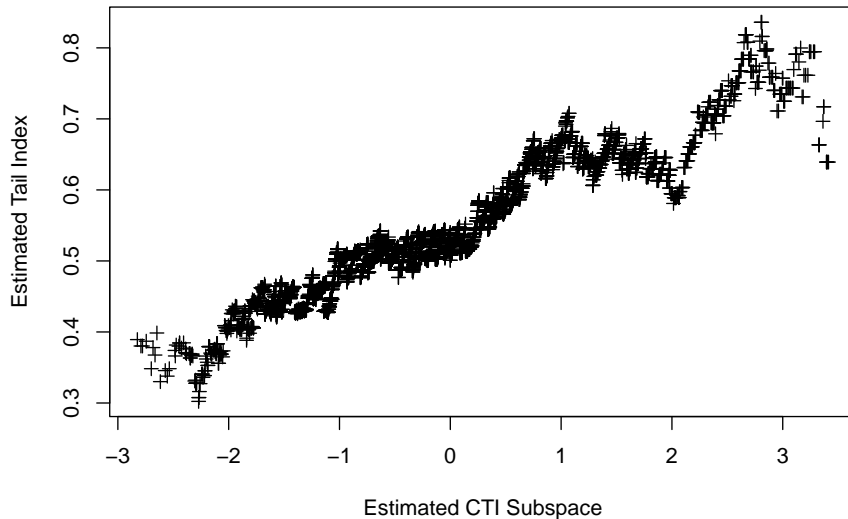


Figure 4: Estimated tail-index versus the projected covariates on the linear subspace spanned by \hat{B}_n .

6 Proofs

6.1 Preliminary results

Lemma 1. *Assume that model (M) holds with the conditions (C.1) and (C.2). For $B \in \mathcal{B}_q$, if the conditional tail-index $\gamma(\cdot)$ and the positive measurable function $c(\cdot)$ involved in (C.2) are such that the function*

$$y \in (0, \infty) \mapsto \mathbb{E} \left[c(X) y^{-1/\gamma(X)} \mid B^\top X \right],$$

is almost surely regularly varying with index $-1/\xi_B(B^\top X) < 0$, then the conditional distribution of Y given $B^\top X$ is heavy-tailed with tail-index $\xi_B(B^\top X)$.

Proof – We have to prove that the conditional survival function $S_B(\cdot, B^\top X)$ is almost surely regularly varying with tail-index $-1/\xi_B(B^\top X)$. In the rest of the proof, all the equalities and inequalities are assumed to hold almost surely. We start with the

following equalities.

$$\begin{aligned} S_B(y, B^\top X) &= \mathbb{E} \left[S(y, X) \mid B^\top X \right] = \int_0^1 \mathbb{P} \left(S(y, X) > z \mid B^\top X \right) dz \\ &= \int_0^1 \mathbb{P} \left(\frac{1}{\gamma(X)} < \frac{\ln(\mathcal{L}(y, X))}{\ln(y)} - \frac{\ln(z)}{\ln(y)} \mid B^\top X \right) dz. \end{aligned}$$

Set $s = -\ln(z)/\ln(y)$. We get

$$S_B(y \mid B^\top X) = \ln(y) \int_0^\infty y^{-s} \mathbb{P} \left(\frac{1}{\gamma(X)} < s + \frac{\ln(\mathcal{L}(y, X))}{\ln(y)} \mid B^\top X \right) ds$$

By assumption (C.2), for all $\varepsilon > 0$, there exists $\mathcal{Y}_0 > 0$ such that for all $y \geq \mathcal{Y}_0$,

$$S_B(y \mid B^\top X) \leq \ln(y) \int_0^\infty y^{-s} \mathbb{P} \left(\frac{1}{\gamma(X)} - \frac{\ln(c(X))}{\ln(y)} < s + \frac{\varepsilon}{\ln(y)} \mid B^\top X \right) ds.$$

Hence, let $t = s + \varepsilon/\ln(y)$,

$$S_B(y \mid B^\top X) \leq e^\varepsilon \ln(y) \int_{\varepsilon/\ln(y)}^\infty y^{-t} \mathbb{P} \left(\frac{1}{\gamma(X)} - \frac{\ln(c(X))}{\ln(y)} < t \mid B^\top X \right) dt.$$

Since for y large enough,

$$\sup_{x \in \mathcal{X}} \left(\frac{\varepsilon}{\ln(y)} + \frac{\ln(c(x))}{\ln(y)} \right) < \inf_{x \in \mathcal{X}} \frac{1}{\gamma(x)},$$

we have, for y large enough, that

$$\mathbb{P} \left(\frac{1}{\gamma(X)} - \frac{\ln(c(X))}{\ln(y)} < \frac{\varepsilon}{\ln(y)} \mid B^\top X \right) = 0.$$

As a consequence,

$$\int_0^{\varepsilon/\ln(y)} y^{-s} \mathbb{P} \left(\frac{1}{\gamma(X)} - \frac{\ln(c(X))}{\ln(y)} < s \mid B^\top X \right) ds = 0,$$

leading to

$$S_B(y \mid B^\top X) \leq e^\varepsilon \ln(y) \int_0^\infty y^{-t} \mathbb{P} \left(\frac{1}{\gamma(X)} - \frac{\ln(c(X))}{\ln(y)} < t \mid B^\top X \right) dt.$$

In the same way, we have for all $\varepsilon > 0$ that

$$S_B(y \mid B^\top X) \geq e^{-\varepsilon} \ln(y) \int_0^\infty y^{-t} \mathbb{P} \left(\frac{1}{\gamma(X)} - \frac{\ln(c(X))}{\ln(y)} < t \mid B^\top X \right) dt.$$

Remarking that

$$\ln(y) \int_0^\infty y^{-t} \mathbb{P} \left(\frac{1}{\gamma(X)} - \frac{\ln(c(X))}{\ln(y)} < t \mid B^\top X \right) dt = \mathbb{E} \left[c(X) y^{-1/\gamma(X)} \mid B^\top X \right],$$

we have shown that for all $\varepsilon > 0$, we have for y large enough, that

$$e^{-\varepsilon} \mathbb{E} \left[c(X) y^{-1/\gamma(X)} \mid B^\top X \right] \leq S(y \mid B^\top X) \leq e^\varepsilon \mathbb{E} \left[c(X) y^{-1/\gamma(X)} \mid B^\top X \right],$$

which is the desired result. ■

The next result establishes the uniform consistency of the general estimator $\widehat{\Psi}_n(B, \mathcal{X}_0)$ of $\Psi(B, \mathcal{X}_0)$.

Lemma 2. *Assume that model (M) holds together with conditions (C.4) and (C.5). Then,*

$$\sup_{B \in \mathcal{B}_q} \left| \widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right| \xrightarrow{\mathbb{P}} 0.$$

Proof – We start by the inequality

$$\sup_{B \in \mathcal{B}_q} \left| \widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right| \leq T_{n,1} + T_{n,2},$$

$$T_{n,1} := \sup_{B \in \mathcal{B}_q} \left| \widetilde{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right|$$

$$\text{and } T_{n,2} := \sup_{B \in \mathcal{B}_q} \sup_{x \in \mathcal{X}_0} \left| \widehat{\xi}_{n,B}(B^\top x) - \xi_B(B^\top x) \right|.$$

Condition (C.5) ensures that $T_{n,2} \xrightarrow{\mathbb{P}} 0$. For the term $T_{n,1}$, we must prove a uniform law of large numbers for the set of parametric functions $\mathcal{F}_q := \{x \in \mathcal{X} \mapsto \xi_B(B^\top x) \mathbb{1}_{\mathcal{X}_0}(x); B \in \mathcal{B}_q\}$. This result is established by compiling different results that can be found for instance in the monograph of van der Vaart and Wellner [19]. Note that under (C.1), $\mathcal{F}_q \subset L_1(X)$ where $L_1(X)$ is the set of all functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that $\mathbb{E}(|f(X)|) < \infty$. Let $\|\cdot\|_1$ be the $L_1(X)$ -norm defined by $\|f\|_1 = \mathbb{E}(|f(X)|)$. According to [19, Theorem 2.4.1], the result of Lemma 2 is true if the bracketing number $N_{[]}(\varepsilon, \mathcal{F}_q, \|\cdot\|_1)$ is finite for all $\varepsilon > 0$ (see [19, Definition 2.1.6] for a definition of bracketing numbers). We now use [19, Theorem 2.7.11] ensuring that under (C.4), one has for all $\varepsilon > 0$,

$$N_{[]}(\varepsilon, \mathcal{F}_q, \|\cdot\|_1) \leq N\left(\frac{\varepsilon}{2\mathbb{E}(M(X))}, \mathcal{B}_q, \|\cdot\|\right),$$

where for all $\eta > 0$ and $\|\cdot\|$ any norm in $\mathbb{R}^{p \times q}$, $N(\eta, \mathcal{B}_q, \|\cdot\|)$ is the covering number of the set \mathcal{B}_q (see [19, Definition 2.1.5]). We conclude the proof by using the compacity of \mathcal{B}_q and the fact that the covering number of a compact subset is finite. ■

6.2 Proofs of main results

Proof of Proposition 1 – Thanks to Lemma 1, we only need to prove that for all $B \in \mathcal{B}_q$, the function

$$y \in (0, \infty) \mapsto \mathbb{E} \left[c(X) y^{1/\gamma(X)} \mid B^\top X \right], \quad (5)$$

is almost surely regularly varying with index $-1/\xi_B(B^\top X) < 0$. According to [9, Theorem 1.2.1 and Remark 1.2.3], for any positive random variable Z , the limit

$$\lim_{y \rightarrow \infty} \frac{\mathbb{E}[\ln(Z/y)\mathbb{I}_{(y,\infty)}(Z) | B^\top X]}{\mathbb{P}(Z > y | B^\top X)} = \xi_B(B^\top X), \quad (6)$$

holds almost surely if and only if the function

$$y \in (0, \infty) \mapsto \mathbb{P}(Z > y | B^\top X) = \mathbb{E} \left[\mathbb{P}(Z > y | X) | B^\top X \right],$$

is almost surely regularly varying with index $-1/\xi_B(B^\top X) < 0$. Taking for Z a positive random variable such that for all $x \in \mathcal{X}$ and $y > [c(x)]^{\gamma(x)}$,

$$\mathbb{P}(Z > y | X = x) = c(x)y^{-1/\gamma(x)},$$

proving (5) is thus equivalent to prove (6). In the rest of the proof, all equalities, inequalities, limits, etc. are assumed to hold almost surely. Using the tower property of the conditional expectation, we get

$$\mathbb{E}[\ln(Z/y)\mathbb{I}_{(y,\infty)}(Z) | B^\top X] = \mathbb{E} \left[\mathbb{E}[\ln(Z/y)\mathbb{I}_{(y,\infty)}(Z) | X] | B^\top X \right].$$

Moreover

$$\begin{aligned} \mathbb{E}[\ln(Z/y)\mathbb{I}_{(y,\infty)}(Z) | X] &= \int_0^\infty \mathbb{P}(Z > ye^z | X) dz \\ &= \gamma(X)c(X)y^{-1/\gamma(X)} = \gamma(X)\mathbb{P}(Z > y | X). \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}[\ln(Z/y)\mathbb{I}_{(y,\infty)}(Z) | B^\top X] &= \mathbb{E} \left[\gamma(X)\mathbb{P}(Z > y | X) | B^\top X \right] \\ &= \mathbb{E} \left[\mathbb{E}[\gamma(X)\mathbb{I}_{(y,\infty)}(Z) | X] | B^\top X \right] = \mathbb{E}[\gamma(X)\mathbb{I}_{(y,\infty)}(Z) | B^\top X] \\ &= \int_0^{\xi_B(B^\top X)} \mathbb{P} \left(\{\gamma(X) > t\} \cap \{Z > y\} | B^\top X \right) dt. \end{aligned}$$

As a consequence, prove (6) is equivalent to prove that

$$\lim_{y \rightarrow \infty} \int_0^{\xi_B(B^\top X)} \frac{\mathbb{P}(\{\gamma(X) > t\} \cap \{Z > y\} | B^\top X)}{\mathbb{P}(Z > y | B^\top X)} dt = \xi_B(B^\top X).$$

Since the integrand is positive and smaller than 1, one can use the dominated convergence theorem to exchange limit and integral. It thus finally remains to prove that for all $t \in (0, \xi_B(B^\top X))$,

$$\lim_{y \rightarrow \infty} \frac{\mathbb{P}(\{\gamma(X) > t\} \cap \{Z > y\} | B^\top X)}{\mathbb{P}(Z > y | B^\top X)} = 1.$$

Under (C.1), the distribution of the random variable $\gamma(X)$ is absolutely continuous and thus $\{\gamma(X) > t\} = \{\gamma(X) \geq t\}$. This leads to the equality

$$\frac{\mathbb{P}(\{\gamma(X) > t\} \cap \{Z > y\} | B^\top X)}{\mathbb{P}(Z > y | B^\top X)} = 1 - \frac{\mathbb{P}(\{\gamma(X) < t\} \cap \{Z > y\} | B^\top X)}{\mathbb{P}(Z > y | B^\top X)}.$$

The rest of the proof consists in showing that for all $t \in (0, \xi_B(B^\top X))$,

$$\lim_{y \rightarrow \infty} y^{1/t} \mathbb{P}(\{\gamma(X) < t\} \cap \{Z > y\} | B^\top X) = 0, \quad (7)$$

and

$$\lim_{y \rightarrow \infty} y^{1/t} \mathbb{P}(Z > y | B^\top X) = \infty. \quad (8)$$

Let us start by proving (7). We have

$$\begin{aligned} y^{1/t} \mathbb{P}(\{\gamma(X) > t\} \cap \{Z > y\} | B^\top X) &= \mathbb{E} \left[y^{1/t} \mathbb{I}_{(0,t)}(\gamma(X)) \mathbb{P}(Z > y | X) \Big| B^\top X \right] \\ &= \mathbb{E} \left[c(X) y^{1/t-1/\gamma(X)} \mathbb{I}_{(0,t)}(\gamma(X)) \Big| B^\top X \right]. \end{aligned}$$

Since for $y > 1$, one has $c(X) y^{1/t-1/\gamma(X)} \mathbb{I}_{(0,t)}(\gamma(X)) \leq c(X)$ and since by assumption $c(\cdot)$ is bounded, we can apply the dominated convergence theorem to get

$$\begin{aligned} &\lim_{y \rightarrow \infty} y^{1/t} \mathbb{P}(\{\gamma(X) > t\} \cap \{Z > y\} | B^\top X) \\ &= \mathbb{E} \left[c(X) \mathbb{I}_{(0,t)}(\gamma(X)) \lim_{y \rightarrow \infty} y^{1/t-1/\gamma(X)} \Big| B^\top X \right] = 0, \end{aligned}$$

proving (7). Let us now turn to the proof of (8). First, note that

$$\mathbb{P}(Z > y | B^\top X) = \mathbb{E} \left[\mathbb{P}(Z > y | X) \Big| B^\top X \right] = \mathbb{E} \left[c(X) y^{-1/\gamma(X)} \Big| B^\top X \right].$$

Moreover, for $t \in (0, \xi_B(B^\top X))$, let $\varepsilon_t(B^\top X) := (t^{-1} - \xi_B^{-1}(B^\top X))/2$. Since $\xi_B(B^\top X)$ is the right endpoint of the conditional distribution of $\gamma(X)$ given $B^\top X$, we have

$$\mathbb{P}(\gamma^{-1}(X) < t^{-1} - \varepsilon_t(B^\top X) | B^\top X) > 0.$$

Hence, for all $y > 0$,

$$\begin{aligned} y^{1/t} \mathbb{P}(Z > y | B^\top X) &\geq \mathbb{E} \left[c(X) y^{t^{-1} - \gamma^{-1}(X)} \mathbb{I}_{[0, t^{-1} - \varepsilon_t(B^\top X)]}(\gamma^{-1}(X)) \Big| B^\top X \right] \\ &\geq \inf_{x \in \mathcal{X}} c(x) \mathbb{P}(\gamma^{-1}(X) < t^{-1} - \varepsilon_t(B^\top X) | B^\top X) y^{\varepsilon_t(B^\top X)}, \end{aligned}$$

which converges to infinity as $y \rightarrow \infty$ since under (C.2), $\inf_{x \in \mathcal{X}} c(x) > 0$. ■

Proof of Theorem 1 – The scheme of this proof is similar to the one used in the proof of [11, Theorem 5.1]. From the definition (4) of \widehat{B}_n , we have

$$\mathbb{P}\left(\widehat{\Psi}_n(\widehat{B}_n, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right) = 1.$$

For a norm $\|\cdot\|$ in $\mathbb{R}^{p \times q}$, let $U(B_0, \rho)$ be the open ball in $\mathbb{R}^{p \times q}$ of center B_0 and radius $\rho > 0$. For all $\rho > 0$,

$$\begin{aligned} 1 &= \mathbb{P}\left(\widehat{\Psi}_n(\widehat{B}_n, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right) \\ &= \mathbb{P}\left(\left\{\widehat{\Psi}_n(\widehat{B}_n, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right\} \cap \{\widehat{B}_n \in U(B_0, \rho)\}\right) \\ &\quad + \mathbb{P}\left(\left\{\widehat{\Psi}_n(\widehat{B}_n, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right\} \cap \{\widehat{B}_n \notin U(B_0, \rho)\}\right). \end{aligned}$$

Hence, since

$$\mathbb{P}\left(\left\{\widehat{\Psi}_n(\widehat{B}_n, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right\} \cap \{\widehat{B}_n \in U(B_0, \rho)\}\right) \leq \mathbb{P}\left(\widehat{B}_n \in U(B_0, \rho)\right),$$

and

$$\begin{aligned} &\mathbb{P}\left(\left\{\widehat{\Psi}_n(\widehat{B}_n, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right\} \cap \{\widehat{B}_n \notin U(B_0, \rho)\}\right) \\ &\leq \mathbb{P}\left(\inf_{B \notin U(B_0, \rho)} \widehat{\Psi}_n(B, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right), \end{aligned}$$

we have that

$$\mathbb{P}\left(\widehat{B}_n \in U(B_0, \rho)\right) + \mathbb{P}\left(\inf_{B \notin U(B_0, \rho)} \widehat{\Psi}_n(B, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right) \geq 1.$$

As a consequence, if we show that

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\inf_{B \notin U(B_0, \rho)} \widehat{\Psi}_n(B, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right) = 0, \quad (9)$$

then, for all $\rho > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\widehat{B}_n \in U(B_0, \rho)\right) = 1,$$

which is the desired result. To prove (9), let us remark first that

$$\inf_{B \notin U(B_0, \rho)} \left(\widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0)\right) + \inf_{B \notin U(B_0, \rho)} \Psi(B, \mathcal{X}_0) \leq \inf_{B \notin U(B_0, \rho)} \widehat{\Psi}_n(B, \mathcal{X}_0).$$

Hence,

$$\begin{aligned} p_n &:= \mathbb{P}\left(\inf_{B \notin U(B_0, \rho)} \widehat{\Psi}_n(B, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right) \\ &\leq \mathbb{P}\left(\inf_{B \notin U(B_0, \rho)} \left(\widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0)\right) \right. \\ &\quad \left. + \Psi(B_0, \mathcal{X}_0) - \widehat{\Psi}_n(B_0, \mathcal{X}_0) \leq \Psi(B_0, \mathcal{X}_0) - \inf_{B \notin U(B_0, \rho)} \Psi(B, \mathcal{X}_0)\right) \end{aligned}$$

Now, since for all $(x, y) \in \mathbb{R}^2$, if $x \leq y$ then $|x| \geq -y$, we have

$$\begin{aligned} p_n \leq & \mathbb{P} \left(\left| \inf_{B \notin U(B_0, \rho)} \left(\widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right) \right. \right. \\ & \left. \left. + \Psi(B_0, \mathcal{X}_0) - \widehat{\Psi}_n(B_0, \mathcal{X}_0) \right| \geq \inf_{B \notin U(B_0, \rho)} \Psi(B, \mathcal{X}_0) - \Psi(B_0, \mathcal{X}_0) \right) \end{aligned}$$

Under condition (C.3), for all $\rho > 0$, there exists $\varepsilon > 0$ such that

$$\inf_{B \notin U(B_0, \rho)} \Psi(B, \mathcal{X}_0) - \Psi(B_0, \mathcal{X}_0) > \varepsilon.$$

Furthermore,

$$\begin{aligned} & \left| \inf_{B \notin U(B_0, \rho)} \left(\widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right) + \Psi(B_0, \mathcal{X}_0) - \widehat{\Psi}_n(B_0, \mathcal{X}_0) \right| \\ & \leq \left| \inf_{B \notin U(B_0, \rho)} \left(\widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right) \right| + \left| \Psi(B_0, \mathcal{X}_0) - \widehat{\Psi}_n(B_0, \mathcal{X}_0) \right| \\ & \leq \sup_{B \notin U(B_0, \rho)} \left| \widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right| + \left| \Psi(B_0, \mathcal{X}_0) - \widehat{\Psi}_n(B_0, \mathcal{X}_0) \right| \\ & \leq \sup_{B \in \mathcal{B}_q} \left| \widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right| + \left| \Psi(B_0, \mathcal{X}_0) - \widehat{\Psi}_n(B_0, \mathcal{X}_0) \right|. \end{aligned}$$

Hence,

$$p_n \leq \mathbb{P} \left(\sup_{B \in \mathcal{B}_q} \left| \widehat{\Psi}_n(B, \mathcal{X}_0) - \Psi(B, \mathcal{X}_0) \right| + \left| \Psi(B_0, \mathcal{X}_0) - \widehat{\Psi}_n(B_0, \mathcal{X}_0) \right| \geq \varepsilon \right).$$

We conclude the proof by using Lemma 2. \blacksquare

Proof of Proposition 2 – Let us start with some notations. For all matrix $B \in \mathcal{B}^d$ with $d \in \{1, \dots, p\}$ and for any $x \in \mathcal{X}$, let $\mathcal{A}_x(B) := \{z \in \mathbb{R}^p; B^\top z = B^\top x\}$. Note that, according to (2), $\xi_B(B^\top x) = \max \{\gamma(z); z \in \mathcal{A}_x(B)\}$. Let us introduce the set \mathcal{B}_{d+1}^* gathering the matrices of the form $[B_{0,d}, u] \in \mathbb{R}^{p \times (d+1)}$ where $u \in \mathbb{R}^p$ is a vector linearly independent to the columns of $B_{0,d}$. For all $C_{d+1} \in \mathcal{B}_{d+1}^*$, let $\tilde{C}_{d+1} \in \mathcal{B}_{d+1}$ be the canonical basis of the subspace spanned by the columns of C_{d+1} . It is readily seen that $\mathcal{A}_x(\tilde{C}_{d+1}) = \mathcal{A}_x(C_{d+1})$ and thus that for all $x \in \mathcal{X}$, $\xi_{\tilde{C}_{d+1}}(\tilde{C}_{d+1}^\top x) = \xi_{C_{d+1}}(C_{d+1}^\top x)$. Since $\mathcal{A}_x(C_{d+1}) \subset \mathcal{A}_x(B_{0,d})$, for all $x \in \mathcal{X}$, we have $\xi_{\tilde{C}_{d+1}}(\tilde{C}_{d+1}^\top x) \leq \xi_{B_{0,d}}(B_{0,d}^\top x)$, leading to

$$\begin{aligned} \min_{B \in \mathcal{B}_{d+1}^*} \Psi(B, \mathcal{X}_0) &= \min_{B \in \mathcal{B}_{d+1}^*} \mathbb{E} \left[\xi_B(B^\top X) \mathbb{I}_{\mathcal{X}_0}(X) \right] \leq \mathbb{E} \left[\xi_{\tilde{C}_{d+1}}(\tilde{C}_{d+1}^\top X) \mathbb{I}_{\mathcal{X}_0}(X) \right] \\ &\leq \mathbb{E} \left[\xi_{B_{0,d}}(B_{0,d}^\top X) \mathbb{I}_{\mathcal{X}_0}(X) \right] = \min_{B \in \mathcal{B}_d} \Psi(B, \mathcal{X}_0), \end{aligned} \quad (10)$$

for all $d \in \{1, \dots, p\}$. Moreover, under (C.3),

$$\min_{B \in \mathcal{B}_q} \Psi(B, \mathcal{X}_0) = \Psi(B_0, \mathcal{X}_0) = \mathbb{E} [\gamma(X) \mathbb{I}_{\mathcal{X}_0}(X)].$$

Since for all $B \in \mathcal{B}_d$ and $x \in \mathcal{X}$, $\xi_B(B^\top x) = \max\{\gamma(z); B^\top z = B^\top x\} \geq \gamma(x)$, we have, for all $d \geq q$,

$$\mathbb{E}[\gamma(X)\mathbb{I}_{\mathcal{X}_0}(X)] \leq \min_{B \in \mathcal{B}_d} \Psi(B, \mathcal{X}_0). \quad (11)$$

From (10), we also have that for all $d \geq q$,

$$\min_{B \in \mathcal{B}_d} \Psi(B, \mathcal{X}_0) \leq \min_{B \in \mathcal{B}_q} \Psi(B, \mathcal{X}_0) = \mathbb{E}[\gamma(X)\mathbb{I}_{\mathcal{X}_0}(X)]. \quad (12)$$

Gathering (11) and (12) entails that for all $d \geq q$,

$$\min_{B \in \mathcal{B}_d} \Psi(B, \mathcal{X}_0) = \mathbb{E}[\gamma(X)\mathbb{I}_{\mathcal{X}_0}(X)],$$

and the proof is complete. ■

References

- [1] Aghbalou, A., Portier, F., Sabourin, A. and Zhou, C. (2024). Tail Inverse Regression: dimension reduction for prediction of extremes, *Bernoulli*, **30(1)**, 503–533.
- [2] Bellman, R.E. (1961). *Adaptative Control Process: A Guided Tour*, Princeton University Press.
- [3] Bousebata, M., Enjolras, G., and Girard, S. (2021). Extreme Partial Least-Squares regression, *Journal of Multivariate Analysis*, To appear.
- [4] Brazauskas, V., Jones, B.L., Puri, M.L. and Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view, *Journal of Statistical Planning and Inference*, **138**, 3590–3604.
- [5] Cook, R.D. and Li, B. (2002). Dimension reduction for conditional mean in regression, *Annals of Statistics*, **30(2)**, 455–474.
- [6] Cover, T. (1968). Estimation by the nearest neighbor rule, *IEEE Transactions on Information Theory*, **14**, 50–55.
- [7] Gardes, L. (2018). Tail Dimension Reduction for extreme quantile estimation, *Extremes*, **21(1)**, 57–95.
- [8] Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator, *Extremes*, **17(1)**, 45–75.
- [9] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.

- [10] Härdle, W. and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation, *The Annals of Statistics*, **13(4)**, 1465–1481.
- [11] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics*, **58**, 71–120.
- [12] Li, K-C (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86(414)**, 316–327.
- [13] Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and its Application*, **9(1)**, 141–142.
- [14] Read, L.K. and Vogel, R.M. (2015). Reliability, return periods, and risk under nonstationarity, *Water Resources Research*, **51(8)**, 6381–6398.
- [15] Rockafellar, R.T. and Uryasev, S. (2002). Conditional Value-at-Risk for general loss distributions, *Journal of Banking and Finance*, **26**, 1443–1471.
- [16] Scrucca, L. (2011). Model-based SIR for dimension reduction, *Computational Statistics and Data Analysis*, **55**, 3010–3026.
- [17] Shami, T.M., El-Saleh, A.A., Alswaitti, M., Al-Tashi, Q. and Mirjalili, S. (2022). Particle Swarm Optimization: A Comprehensive Survey, *IEEE Access*, **10**, 10031–10061.
- [18] Tukey, J. W. (1975). Mathematics and the picturing of data, *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, **5(2)**, 523–531.
- [19] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer New York.
- [20] Wang, D., Tan, D. and Liu, L. (2018). Particle swarm optimization algorithm: an overview, *Soft computing*, **22**, 387–408.
- [21] Watson, G. S. (1964). Smooth regression analysis, *Sankhyā*, **26(15)**, 175–184.
- [22] Xia, Y. (2009). Model checking in regression via dimension reduction, *Biometrika*, **96(1)**, 133–148.

Appendix – Proof of Theorem 2

We shall prove that the estimator $\widehat{\xi}_B^{(H)}(B^\top x; \alpha, h) = \widehat{\xi}_B^{(H)}(B^\top x)$ satisfies condition (C.5) i.e., that for all $\varepsilon > 0$, the probability

$$p_n := \mathbb{P} \left(\sup_{(B,x) \in \Theta} \left| \widehat{\xi}_B^{(H)}(B^\top x) - \xi_B(B^\top x) \right| > \varepsilon \right),$$

converges to 0 as n goes to infinity. The proof consists in showing the uniform consistency on a sequence of “sufficiently large” subsets Θ_n of Θ and deal with the oscillation of the estimator. We use the same scheme as the one used for the proof of [8, Theorem 1] but here we have to deal with the uniformity on $\mathcal{B}_q \times \mathcal{X}$ and not only on \mathcal{X} . This gives rise to technical difficulties.

Since Θ is a compact subset, for a fixed $\eta > q$ and every $n \in \mathbb{N} \setminus \{0\}$, there exists a finite subset Θ_n of Θ such that for all $\theta := (B, x) \in \Theta$, one can find $\tilde{\theta}_n(\theta) = (\tilde{B}_n(\theta), \tilde{x}_n(\theta)) \in \Theta_n$ satisfying $\delta(\tilde{\theta}_n(\theta), \theta) < n^{-\eta}$. Note that Θ_n can be chosen such that for some $c = c(\eta) > 0$, $\text{card}(\Theta_n) = \mathcal{O}(n^c)$ for all $n \in \mathbb{N} \setminus \{0\}$, see [10, Proof of Lemma 1]. The triangular inequality yields:

$$\begin{aligned} p_n \leq & \mathbb{I} \left\{ \sup_{\theta=(B,x) \in \Theta} \left| \xi_B(B^\top x) - \xi_{\tilde{B}_n(\theta)} \left([\tilde{B}_n(\theta)]^\top \tilde{x}_n(\theta) \right) \right| > \varepsilon/3 \right\} \\ & + \mathbb{P} \left(\sup_{\tilde{\theta}=(\tilde{B},\tilde{x}) \in \Theta_n} \left| \widehat{\xi}_{\tilde{B}}^{(H)} \left(\tilde{B}^\top \tilde{x} \right) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| > \varepsilon/3 \right) \\ & + \mathbb{P} \left(\sup_{\theta=(B,x) \in \Theta} \left| \widehat{\xi}_{\tilde{B}_n(\theta)}^{(H)} \left([\tilde{B}_n(\theta)]^\top \tilde{x}_n(\theta) \right) - \widehat{\xi}_B^{(H)} \left(B^\top x \right) \right| > \varepsilon/3 \right). \end{aligned}$$

We must prove that the three terms in the above inequality converge to 0 as n goes to infinity. This is carried out in Propositions 3, 4 and 5 given below.

Proposition 3. *Under condition (H.3), for n large enough and for all $\varepsilon > 0$,*

$$\sup_{\theta=(B,x) \in \Theta} \left| \xi_B(B^\top x) - \xi_{\tilde{B}_n(\theta)} \left([\tilde{B}_n(\theta)]^\top \tilde{x}_n(\theta) \right) \right| > \varepsilon.$$

Proposition 4. *Under conditions (C.1), (H.1), (H.2), (H.4) and (H.5), if the sequences (α) and (h) are such that $nh^q \alpha / \ln(n) \rightarrow \infty$ and $\alpha \ln(nh^q) \rightarrow 0$ as $n \rightarrow \infty$ then, for all $\varepsilon > 0$,*

$$\mathbb{P} \left(\sup_{\tilde{\theta}=(\tilde{B},\tilde{x}) \in \Theta_n} \left| \widehat{\xi}_{\tilde{B}}^{(H)} \left(\tilde{B}^\top \tilde{x} \right) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| > \varepsilon \right) \rightarrow 0.$$

Proposition 5. *Under conditions (C.1), (H.1), (H.2), (H.4) and (H.5), if the sequences (α) and (h) are such that $nh^q\alpha/\ln(n) \rightarrow \infty$ and $\alpha\ln(nh^q) \rightarrow 0$ as $n \rightarrow \infty$ then, for all $\varepsilon > 0$,*

$$\mathbb{P} \left(\sup_{\theta=(B,x) \in \Theta} \left| \widehat{\xi}_{\tilde{B}_n(\theta)}^{(H)} \left([\tilde{B}_n(\theta)]^\top \tilde{x}_n(\theta) \right) - \widehat{\xi}_B^{(H)} \left(B^\top x \right) \right| > \varepsilon \right) \rightarrow 0.$$

Proof of Proposition 3

From (H.3), the function $(B, x) \in \Theta \mapsto \xi_B(B^\top x)$ is continuous. Since Θ is compact, it is also uniformly continuous which shows the result since $\delta(\tilde{\theta}_n(\theta), \theta) < n^{-\eta} \rightarrow 0$ as n goes to infinity. \blacksquare

Proof of Proposition 4

Notations

For all $i \in \{1, \dots, n\}$ and $B \in \mathcal{B}_q$, let

$$V_i(B) := S_B(Y_i, B^\top X_i) \quad \text{and} \quad U_i^* = U_i(B, x, h) = S_B(Z_i^*, B^\top W_i^*), \quad (13)$$

where the random variables Z_i^* and W_i^* have been introduced in Section 3.3. Recall that the star (\star) is here to recall the dependence on B , x and h .

Let $\mathbf{m} := (m(\tilde{B}, \tilde{x}); (\tilde{B}, \tilde{x}) \in \Theta_n)$ be a list of $\text{card}(\Theta_n)$ integers. We introduce the event

$$\mathfrak{B}_n(\mathbf{m}) := \bigcap_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left\{ M(\tilde{B}, \tilde{x}, h) = m(\tilde{B}, \tilde{x}) \right\}.$$

Finally, denoting by ϑ_q the volume of the unit q -ball and by

$$\mathbf{L}_n = \left\{ \mathbf{m}; m(\tilde{B}, \tilde{x}) \in \left\{ \left\lceil \frac{C_1 \vartheta_q}{2} nh^q \right\rceil, \dots, \left\lceil \frac{3C_2 \vartheta_q}{2} nh^q \right\rceil \right\} \text{ for all } (\tilde{B}, \tilde{x}) \in \Theta_n \right\},$$

we introduce the event

$$\mathfrak{A}_n := \bigcup_{\mathbf{m} \in \mathbf{L}_n} \mathfrak{B}_n(\mathbf{m}).$$

Note that the events $\{\mathfrak{B}_n(\mathbf{m}); \mathbf{m} \in \mathbf{L}_n\}$ are disjoint.

Preliminary results

The following result of real analysis will be useful in the proof of Lemma 6. A proof can be found in [8, Lemma 1].

Lemma 3. Let $\{a_1, \dots, a_n\}$ and $\{b_1, \dots, b_n\}$ be two sets of distincts real numbers such that for all $i = 1, \dots, n$, $a_i \leq b_i$. Denoting by $\{a_{(1)}, \dots, a_{(n)}\}$ and $\{b_{(1)}, \dots, b_{(n)}\}$ the ordered sets, we have for all $i = 1, \dots, n$ that $a_{(i)} \leq b_{(i)}$.

The next result is dedicated to the random variables $V_i(B)$ and U_i^* .

Lemma 4. i) The random variables $V_1(B), \dots, V_n(B)$ are independent standard uniform random variables which are independent from X_1, \dots, X_n .

ii) For all $(\tilde{B}, \tilde{x}) \in \Theta_n$ and conditionally to $\mathfrak{B}_n(\mathbf{m})$ with $\mathbf{m} := (m(\tilde{B}, \tilde{x}); (\tilde{B}, \tilde{x}) \in \Theta_n)$, the random variables $U_1^*, \dots, U_{m(\tilde{B}, \tilde{x})}^*$ are independent standard uniform random variables.

iii) For all $(\tilde{B}, \tilde{x}) \in \Theta_n$ and conditionally to $\mathfrak{B}_n(\mathbf{m})$, the random variables

$$\left\{ E_i(\tilde{B}, \tilde{x}, h) = E_i^* := i \ln \left(\frac{U_{(i)}^*}{U_{(i+1)}^*} \right), i = 1, \dots, m(\tilde{B}, \tilde{x}) \right\},$$

are independent standard exponential random variables.

Proof – Points i) and ii) are proved in [8, Lemma 4]. For the point iii), let $m = m(\tilde{B}, \tilde{x})$ and for all $i = 1, \dots, m$, let $F_i^* := -\ln(U_i^*)$. Using the result of ii), we get that conditionally to $\mathfrak{B}_n(\mathbf{m})$, F_1^*, \dots, F_m^* are independent standard exponential random variables. Hence, if $F_{(1)} < \dots < F_{(m)}$ are the associated order statistics,

$$\{E_i^*, i = 1, \dots, m\} = \{i (F_{(m-i+1)}^* - F_{(m-i)}^*), i = 1, \dots, m\}.$$

We conclude the proof by using Rényi representation theorem. ■

The next result is dedicated to the number of observations in the set $\mathcal{T}(B, x, h)$.

Lemma 5. Under conditions (H.1) and (H.2), if $nh^q / \ln(n) \rightarrow \infty$ as $n \rightarrow \infty$ then,

$$\frac{1}{nh^q} \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left| M(\tilde{B}, \tilde{x}, h) - \vartheta_q nh^q f_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| \xrightarrow{\mathbb{P}} 0.$$

Lemma 5 means that, uniformly on $(\tilde{B}, \tilde{x}) \in \Theta_n$,

$$\frac{M(\tilde{B}, \tilde{x}, h)}{nh^q} \xrightarrow{\mathbb{P}} \vartheta_q f_{\tilde{B}}(\tilde{B}^\top \tilde{x}).$$

As a consequence, since

$$\mathfrak{A}_n := \bigcap_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left\{ \frac{M(\tilde{B}, \tilde{x}, h)}{nh^q} \in \left[\frac{C_1 \vartheta_q}{2}, \frac{3C_2 \vartheta_q}{2} \right] \right\},$$

and since, from (H.2), $0 < C_1 < f_{\tilde{B}}(\tilde{B}^\top \tilde{x}) < C_2$ for all $(\tilde{B}, \tilde{x}) \in \Theta_n$, it is readily seen that $\mathbb{P}(\mathfrak{A}_n) \rightarrow 1$.

Proof Lemma 5 – For all $(\tilde{B}, \tilde{x}) \in \Theta_n$, we have the inequality

$$\left| M(\tilde{B}, \tilde{x}, h) - \vartheta_q n h^q f_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| \leq T_{1,n}(\tilde{B}, \tilde{x}, h) + T_{2,n}(\tilde{B}, \tilde{x}, h),$$

where

$$T_{1,n}(\tilde{B}, \tilde{x}, h) := \left| n \mathbb{P} \left(X \in \mathcal{T}(\tilde{B}, \tilde{x}, h) \right) - \vartheta_q n h^q f_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right|,$$

and

$$T_{2,n}(\tilde{B}, \tilde{x}, h) := \left| M(\tilde{B}, \tilde{x}, h) - n \mathbb{P} \left(X \in \mathcal{T}(\tilde{B}, \tilde{x}, h) \right) \right|.$$

We thus have to prove that, as $n \rightarrow \infty$,

$$\frac{1}{n h^q} \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} T_{1,n}(\tilde{B}, \tilde{x}, h) \rightarrow 0 \quad \text{and} \quad \frac{1}{n h^q} \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} T_{2,n}(\tilde{B}, \tilde{x}, h) \xrightarrow{\mathbb{P}} 0. \quad (14)$$

Denoting by \mathcal{U}_q the unit q -ball, we get

$$\begin{aligned} \mathbb{P} \left(X \in \mathcal{T}(\tilde{B}, \tilde{x}, h) \right) &= h^q \int_{\mathcal{U}_q} f_{\tilde{B}}(\tilde{B}^\top \tilde{x} + h u) \, du \\ &= \vartheta_q h^q f_{\tilde{B}}(\tilde{B}^\top \tilde{x}) + h^q \int_{\mathcal{U}_q} \left[f_{\tilde{B}}(\tilde{B}^\top \tilde{x} + h u) - f_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right] \, du. \end{aligned} \quad (15)$$

Since $\tilde{B} \in \mathcal{B}_q$ is an orthogonal matrix, we can write

$$f_{\tilde{B}}(\tilde{B}^\top \tilde{x} + h u) = f_{\tilde{B}}(\tilde{B}^\top (\tilde{x} + h \tilde{B}(\tilde{B}^\top \tilde{B})^{-1} u)).$$

By construction of the set \mathcal{B}_q , we have $\|\tilde{B}(\tilde{B}^\top \tilde{B})^{-1}\| = \|\tilde{B}\| < \infty$. Hence, for all $u \in \mathcal{U}_q$, the vector $\tilde{x} + h \tilde{B}(\tilde{B}^\top \tilde{B})^{-1} u$ belongs to the ball of center \tilde{x} and radius $h \|\tilde{B}\|$. Since $h \rightarrow 0$ as $n \rightarrow \infty$, using [8, Lemma 2] we get that for n large enough and uniformly on $(\tilde{B}, \tilde{x}) \in \Theta_n$, the vector $\tilde{x} + h \tilde{B} u$ belongs to \mathcal{X} . Hence, from condition (H.2), for n large enough, we have that for all $u \in \mathcal{U}_q$, there exists $K_f > 0$ such that

$$\left| f_{\tilde{B}}(\tilde{B}^\top \tilde{x} + h u) - f_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| \leq K_f h \|\tilde{B} u\| \leq K h. \quad (16)$$

Gathering (15) and (16), we obtain that for all $(\tilde{B}, \tilde{x}) \in \Theta_n$, $T_{1,n}(\tilde{B}, \tilde{x}, h) \leq K_f \vartheta_q n h^{q+1}$, proving the first part of (14). Let us now focus on the second part of (14). For all $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P} \left(\frac{1}{n h^q} \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} T_{2,n}(\tilde{B}, \tilde{x}, h) > \varepsilon \right) &= \mathbb{P} \left(\bigcup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left\{ \frac{1}{n h^q} T_{2,n}(\tilde{B}, \tilde{x}, h) > \varepsilon \right\} \right) \\ &\leq \text{card}(\Theta_n) \times \mathbb{P} \left(\frac{1}{n h^q} T_{2,n}(\tilde{B}, \tilde{x}, h) > \varepsilon \right) \end{aligned}$$

We now use Bernstein's inequality to get

$$\mathbb{P}\left(\frac{1}{nh^q}T_{2,n}(\tilde{B}, \tilde{x}, h) > \varepsilon\right) \leq 2 \exp\left(-\frac{\varepsilon nh^q}{2[\lambda_n(\tilde{B}, \tilde{x}, h) - 1/3]}\right),$$

where

$$\lambda_n(\tilde{B}, \tilde{x}, h) := \frac{1}{\varepsilon h^q} \mathbb{P}\left(X \in \mathcal{T}(\tilde{B}, \tilde{x}, h)\right) \mathbb{P}\left(X \notin \mathcal{T}(\tilde{B}, \tilde{x}, h)\right).$$

Using (15) and (16), we have

$$\begin{aligned} \mathbb{P}\left(X \in \mathcal{T}(\tilde{B}, \tilde{x}, h)\right) \mathbb{P}\left(X \notin \mathcal{T}(\tilde{B}, \tilde{x}, h)\right) &\leq \mathbb{P}\left(X \in \mathcal{T}(\tilde{B}, \tilde{x}, h)\right) \\ &\leq \vartheta_q h^q \left(f_{\tilde{B}}(\tilde{B}^\top \tilde{x}) + R_n\right), \end{aligned}$$

where $|R_n| \leq K \vartheta_q h \rightarrow 0$. Hence, from (H.2), there exists $C_2 > 0$ such that $\lambda_n(\tilde{B}, \tilde{x}, h) \leq 2\vartheta_q C_2/\varepsilon$. Recall that $\text{card}(\Theta_n) = \mathcal{O}(n^c)$, we finally get that

$$\mathbb{P}\left(\frac{1}{nh^q} \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} T_{2,n}(\tilde{B}, \tilde{x}, h) > \varepsilon\right) = \mathcal{O}\left(n^c \exp\left(-\frac{\varepsilon nh^q}{2[2\vartheta_q C_2/\varepsilon - 1/3]}\right)\right),$$

which converges to 0 since $nh^q/\ln(n) \rightarrow \infty$. \blacksquare

Let us now provide an expansion of the conditional tail-index estimator. We use the notations introduced in Lemma 4.

Lemma 6. *Assume that $nh^q\alpha \rightarrow \infty$ as $n \rightarrow \infty$. Uniformly on $(\tilde{B}, \tilde{x}) \in \Theta_n$, for n large enough, we have conditionally to $\mathfrak{B}_n(\mathbf{m})$ with $\mathbf{m} := (m = m(\tilde{B}, \tilde{x}); (\tilde{B}, \tilde{x}) \in \Theta_n)$, that*

$$\begin{aligned} \left| \hat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \frac{1}{[\alpha m]} \sum_{i=1}^{[\alpha m]} E_i^* \right| &\leq 2\omega^* \left(V_{(1)}(\tilde{B}), V_{(n)}(\tilde{B})\right) \\ &\quad + \Delta_{\tilde{B}, \tilde{x}}(1/U_{([\alpha M^*]+1)}^*). \end{aligned}$$

Proof – Since $nh^q\alpha \rightarrow \infty$, we have conditionally to $\mathfrak{B}_n(\mathbf{m})$ and uniformly on $(\tilde{B}, \tilde{x}) \in \Theta_n$ that for n large enough,

$$\hat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) = \frac{1}{[\alpha m]} \sum_{i=1}^{[\alpha m]} \ln\left(\frac{Z_{(m-i+1)}^*}{Z_{(m-[\alpha m])}^*}\right),$$

where we use the shortcuts $m = m(\tilde{B}, \tilde{x})$ and $Z_i^* = Z_i(\tilde{B}, \tilde{x}, h)$. For all $i \in \{1, \dots, [\alpha m]\}$, we have

$$\ln Z_i^* = \ln\left(Q_{\tilde{B}}(U_i^*, \tilde{B}^\top W_i^*)\right) \leq \ln\left(Q_{\tilde{B}}(U_i^*, \tilde{B}^\top \tilde{x})\right) + \omega^*(U_{(1)}^*, U_{(m)}^*),$$

where $U_i^* = U_i(\tilde{B}, \tilde{x}, h)$ and $\omega^*(u, v) = \omega(u, v, \tilde{B}, \tilde{x}, h)$. Using Lemma 3, we get that for all $i = 1, \dots, m$

$$\ln Z_{(m-i+1)}^* \leq \ln \left(Q_{\tilde{B}}(U_{(i)}^*, \tilde{B}^\top \tilde{x}) \right) + \omega^*(U_{(1)}^*, U_{(m)}^*).$$

Similarly,

$$\ln Z_{(m-i+1)}^* \geq \ln \left(Q_{\tilde{B}}(U_{(i)}^*, \tilde{B}^\top \tilde{x}) \right) - \omega^*(U_{(1)}^*, U_{(m)}^*).$$

We deduce from these two last inequalities that

$$\left| \ln \left(\frac{Z_{(m-i+1)}^*}{Z_{(m-\lfloor \alpha m \rfloor)}^*} \right) - \ln \left(\frac{Q_{\tilde{B}}(U_{(i)}^*, \tilde{B}^\top \tilde{x})}{Q_{\tilde{B}}(U_{(\lfloor \alpha m \rfloor + 1)}^*, \tilde{B}^\top \tilde{x})} \right) \right| \leq 2\omega^*(V_{(1)}(\tilde{B}), V_{(m)}(\tilde{B})),$$

since $V_{(1)}(\tilde{B}) \leq U_{(1)}^*$ and $V_{(m)}(\tilde{B}) \geq U_{(m)}^*$. As a consequence, conditionally to $\mathfrak{B}_n(\mathbf{m})$ and uniformly on $(\tilde{B}, \tilde{x}) \in \Theta_n$ we have for n large enough,

$$\left| \hat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) - \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} \ln \left(\frac{Q_{\tilde{B}}(U_{(i)}^*, \tilde{B}^\top \tilde{x})}{Q_{\tilde{B}}(U_{(\lfloor \alpha m \rfloor + 1)}^*, \tilde{B}^\top \tilde{x})} \right) \right| \leq 2\omega^*(V_{(1)}(\tilde{B}), V_{(m)}(\tilde{B})).$$

Now remark that from Lemma 4, iii) and since $\ln(Q_{\tilde{B}}(u, \tilde{B}^\top \tilde{x})) = \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \ln(u^{-1}) + \ln(\ell_{\tilde{B}}(u^{-1}, \tilde{B}^\top \tilde{x}))$,

$$\begin{aligned} \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} \ln \left(\frac{Q_{\tilde{B}}(U_{(i)}^*, \tilde{B}^\top \tilde{x})}{Q_{\tilde{B}}(U_{(\lfloor \alpha m \rfloor + 1)}^*, \tilde{B}^\top \tilde{x})} \right) &= \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} i \ln \left(\frac{Q_{\tilde{B}}(U_{(i)}^*, \tilde{B}^\top \tilde{x})}{Q_{\tilde{B}}(U_{(i+1)}^*, \tilde{B}^\top \tilde{x})} \right) \\ &= \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} E_i^* + \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} i \ln \left(\frac{\ell_{\tilde{B}}(1/U_{(i)}^*, \tilde{B}^\top \tilde{x})}{\ell_{\tilde{B}}(1/U_{(i+1)}^*, \tilde{B}^\top \tilde{x})} \right). \end{aligned}$$

We conclude the proof by remarking that

$$\begin{aligned} &\left| \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} i \ln \left(\frac{\ell_{\tilde{B}}(1/U_{(i)}^*, \tilde{B}^\top \tilde{x})}{\ell_{\tilde{B}}(1/U_{(i+1)}^*, \tilde{B}^\top \tilde{x})} \right) \right| \\ &= \left| \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} \ln \left(\frac{\ell_{\tilde{B}}(1/U_{(i)}^*, \tilde{B}^\top \tilde{x})}{\ell_{\tilde{B}}(1/U_{(\lfloor \alpha m \rfloor + 1)}^*, \tilde{B}^\top \tilde{x})} \right) \right| \leq \Delta_{\tilde{B}, \tilde{x}}(1/U_{(\lfloor \alpha m \rfloor + 1)}^*). \end{aligned}$$

■

We are now in position to prove Proposition 4.

Proof of Proposition 4

For all $\varepsilon > 0$, we have

$$\begin{aligned} &\mathbb{P} \left(\sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left| \hat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| > \varepsilon \right) \\ &\leq [1 - \mathbb{P}(\mathfrak{A}_n)] + \mathbb{P} \left(\left\{ \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left| \hat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| > \varepsilon \right\} \cap \mathfrak{A}_n \right). \end{aligned}$$

Since Lemma 5 entails that $\mathbb{P}(\mathfrak{A}_n) \rightarrow 1$, we only need to focus on the second term. We have

$$\begin{aligned} & \mathbb{P} \left(\left\{ \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left| \widehat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| > \varepsilon \right\} \cap \mathfrak{A}_n \right) \\ & \leq \sum_{\mathbf{m} \in \mathbf{L}_n} \mathbb{P} \left(\left\{ \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left| \widehat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| > \varepsilon \right\} \middle| \mathfrak{B}_n(\mathbf{m}) \right) \mathbb{P}(\mathfrak{B}_n(\mathbf{m})). \end{aligned}$$

Since the events $\{\mathfrak{B}_n(\mathbf{m}), \mathbf{m} \in \mathbf{L}_n\}$ are disjoint, we obtain

$$\begin{aligned} & \mathbb{P} \left(\left\{ \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left| \widehat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| > \varepsilon \right\} \cap \mathfrak{A}_n \right) \\ & \leq \sup_{\mathbf{m} \in \mathbf{L}_n} \mathbb{P} \left(\left\{ \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left| \widehat{\xi}_{\tilde{B}}^{(H)}(\tilde{B}^\top \tilde{x}) - \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \right| > \varepsilon \right\} \middle| \mathfrak{B}_n(\mathbf{m}) \right) =: \sup_{\mathbf{m} \in \mathbf{L}_n} T(\mathbf{m}). \end{aligned}$$

Using Lemma 6 and keeping the same notations, we get that $T(\mathbf{m}) \leq T_1(\mathbf{m}) + T_2(\mathbf{m}) + T_3(\mathbf{m})$, where

$$T_1(\mathbf{m}) := \mathbb{P} \left(\sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \omega^* \left(V_{(1)}(\tilde{B}), V_{(n)}(\tilde{B}) \right) > \varepsilon/8 \middle| \mathfrak{B}_n(\mathbf{m}) \right),$$

$$T_2(\mathbf{m}) := \mathbb{P} \left(\sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \xi_{\tilde{B}}(\tilde{B}^\top \tilde{x}) \left| \frac{1}{\lfloor \alpha m \rfloor} \sum_{i=1}^{\lfloor \alpha m \rfloor} E_i^* - 1 \right| > \varepsilon/2 \middle| \mathfrak{B}_n(\mathbf{m}) \right),$$

and

$$T_3(\mathbf{m}) := \mathbb{P} \left(\sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \Delta_{\tilde{B}, \tilde{x}}(1/U_{\lfloor \alpha m \rfloor + 1}^*) > \varepsilon/4 \middle| \mathfrak{B}_n(\mathbf{m}) \right).$$

Let us first focus on the term $T_1(\mathbf{m})$. Recall that there exists some $c > 0$ such that $\text{card}(\Theta_n) = \mathcal{O}(n^c)$. Let $\kappa > c$ and let us introduce the set

$$\mathcal{G}_n := \bigcap_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left\{ V_{(1)}(\tilde{B}) > n^{-(1+\kappa)} \right\} \cap \left\{ V_{(n)}(\tilde{B}) < 1 - n^{-(1+\kappa)} \right\}.$$

We have,

$$\begin{aligned} 1 - \mathbb{P}(\mathcal{G}_n) &= \mathbb{P} \left(\bigcup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \left\{ V_{(1)}(\tilde{B}) \leq n^{-(1+\kappa)} \right\} \cup \left\{ V_{(n)}(\tilde{B}) \geq 1 - n^{-(1+\kappa)} \right\} \right) \\ &\leq 2 \text{card}(\Theta_n) \left(1 - [1 - n^{-(1+\kappa)}]^n \right) = \mathcal{O}(n^{c-\kappa}) \rightarrow 0. \end{aligned}$$

Hence, for all $\mathbf{m} \in \mathbf{L}_n$, we have

$$T_1(\mathbf{m}) \leq \mathbb{I} \left\{ \omega^* \left(n^{-(1+\kappa)}, 1 - n^{-(1+\kappa)} \right) > \varepsilon/8 \right\} + [1 - \mathbb{P}(\mathcal{G}_n | \mathfrak{B}_n(\mathbf{m}))].$$

From Lemma 4 *i*), we know that the events \mathcal{G}_n and $\mathfrak{B}_n(\mathbf{m})$ are independent. Hence, $\mathbb{P}(G_n \mid \mathfrak{B}_n(\mathbf{m})) = \mathbb{P}(G_n) \rightarrow 1$ and using condition (H.5),

$$\sup_{\mathbf{m} \in \mathbf{L}_n} T_1(\mathbf{m}) \leq \mathbb{I} \left\{ \omega^* \left(n^{-(1+\kappa)}, 1 - n^{-(1+\kappa)} \right) > \varepsilon/8 \right\} + [1 - \mathbb{P}(\mathcal{G}_n)] \rightarrow 0. \quad (17)$$

For the term $T_2(\mathbf{m})$, we have under (C.1)

$$T_2(\mathbf{m}) \leq \text{card}(\Theta_n) \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \mathbb{P} \left(\left| \frac{1}{[\alpha m]} \sum_{i=1}^{[\alpha m]} E_i^* - 1 \right| > \frac{\varepsilon}{2\bar{\gamma}} \mid \mathfrak{B}_n(\mathbf{m}) \right).$$

Using Chernoff's inequality for exponential random variables, there exists a constant $C_3 > 0$ such that for all $\mathbf{m} \in \mathbf{L}_n$,

$$T_2(\mathbf{m}) \leq \text{card}(\Theta_n) \exp(-C_3 [\alpha m]) \leq \text{card}(\Theta_n) \exp \left(-\frac{3C_2 C_3 \vartheta_q}{2} n h^q \alpha \right).$$

Since $\text{card}(\Theta_n) = \mathcal{O}(n^c)$ and $n h^q \alpha / \ln(n) \rightarrow \infty$, we get that

$$\sup_{\mathbf{m} \in \mathbf{L}_n} T_2(\mathbf{m}) \rightarrow 0. \quad (18)$$

Finally, for the term $T_3(\mathbf{m})$, we start with the inequality

$$T_3(\mathbf{m}) \leq \text{card}(\Theta_n) \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \mathbb{P} \left(\Delta_{\tilde{B}, \tilde{x}} (1/U_{([\alpha m]+1)}^*) > \varepsilon/4 \mid \mathfrak{B}_n(\mathbf{m}) \right).$$

For all sequence (ε_n) such that $\varepsilon_n \rightarrow 0$, condition (H.4) entails that for n large enough and uniformly on $(\tilde{B}, \tilde{x}) \in \Theta_n$, if the event $\{U_{([\alpha m]+1)}^* \leq \varepsilon_n\}$ holds then $\Delta_{\tilde{B}, \tilde{x}} (1/U_{([\alpha m]+1)}^*) \leq \varepsilon/4$. As a consequence,

$$T_3(\mathbf{m}) \leq \text{card}(\Theta_n) \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \mathbb{P} \left(U_{([\alpha m]+1)}^* > \varepsilon_n \mid \mathfrak{B}_n(\mathbf{m}) \right).$$

From the second part of Lemma 4, conditionally to $\mathfrak{B}_n(\mathbf{m})$, the random variable $U_{([\alpha m]+1)}^*$ follows a Beta distribution with parameters $[\alpha m]+1$ and $m - [\alpha m]$. Hence,

$$\begin{aligned} & \mathbb{P} \left(U_{([\alpha m]+1)}^* > \varepsilon_n \mid \mathfrak{B}_n(\mathbf{m}) \right) \\ &= \frac{m!}{[\alpha m]!(m - [\alpha m] - 1)!} \int_{\varepsilon_n}^1 x^{[\alpha m]} (1-x)^{m - [\alpha m] - 1} dx \\ &\leq m^{[\alpha m]+1} (1 - \varepsilon_n)^{m - [\alpha m] - 1} \end{aligned}$$

Since $\ln(1-u) \leq -u$ for all $u \in (0, 1)$, we have for all $\mathbf{m} \in \mathbf{L}_n$,

$$\begin{aligned} & \mathbb{P} \left(U_{([\alpha m]+1)}^* > \varepsilon_n \mid \mathfrak{B}_n(\mathbf{m}) \right) \\ &\leq \exp \left[-m\varepsilon_n \left(\frac{m - [\alpha m] + 1}{2m} - ([\alpha m] + 1) \frac{\ln(m)}{m\varepsilon_n} \right) \right]. \end{aligned}$$

Since $m = \mathcal{O}(nh^q)$ uniformly on $(\tilde{B}, \tilde{x}) \in \Theta_n$, it is readily seen that for all sequence ε_n such that $\alpha \ln(nh^q)/\varepsilon_n \rightarrow 0$, for n large enough, there exists a constant $C_4 > 0$ such that

$$\sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \mathbb{P}(U_{(\lfloor \alpha m \rfloor + 1)}^* > \varepsilon_n \mid \mathfrak{B}_n(\mathbf{m})) \leq \exp(-C_4 nh^q \varepsilon_n).$$

As a consequence, since $\text{card}(\Theta_n) = \mathcal{O}(n^c)$, for a sequence (ε_n) such that $\alpha \ln(nh^q)/\varepsilon_n \rightarrow 0$ and $nh^q \varepsilon_n / \ln(n) \rightarrow \infty$, we get that

$$\sup_{\mathbf{m} \in \mathbf{L}_n} T_3(\mathbf{m}) \rightarrow 0. \quad (19)$$

For instance, one can take $\varepsilon_n = (\alpha \ln(nh^q))^{1/2}$. Since by assumption, $\alpha \ln(nh^q) \rightarrow 0$, we have $\alpha \ln(nh^q)/\varepsilon_n \rightarrow 0$. Moreover,

$$\frac{nh^q \varepsilon_n}{\ln(n)} = \frac{nh^q \alpha}{\ln(n)} \left(\frac{\ln(nh^q)}{\alpha} \right)^{1/2} \rightarrow \infty,$$

since by assumption $nh^q \alpha / \ln(n) \rightarrow \infty$.

The conclusion of the proof is then straightforward by gathering (17), (18) and (19). ■

Proof of Proposition 5

Notations

Recall that for all $\theta = (B, x) \in \Theta$ we have $\delta(\tilde{\theta}_n(\theta), \theta) < n^{-\eta}$ with $\eta > q$. For all $\tilde{C} > 0$ and $(\tilde{B}, \tilde{x}) \in \Theta_n$, let

$$\begin{aligned} \tilde{\mathcal{T}}_{n, \tilde{C}}(\tilde{B}, \tilde{x}, h) &:= \left\{ y \in \mathcal{X}; h - \tilde{C}n^{-\eta} < \left\| \tilde{B}^\top(\tilde{x} - y) \right\| \leq h + \tilde{C}n^{-\eta} \right\} \\ &= \mathcal{T}(\tilde{B}, \tilde{x}, h + \tilde{C}n^{-\eta}) \setminus \mathcal{T}(\tilde{B}, \tilde{x}, h - \tilde{C}n^{-\eta}) \end{aligned}$$

Note that since $\eta > q$, for n large enough, $\tilde{C}n^{-\eta} < h$. Let also

$$\tilde{D}_n(\tilde{B}, \tilde{x}) := \sum_{i=1}^n \mathbb{I} \left\{ X_i \in \tilde{\mathcal{T}}_{n, \tilde{C}}(\tilde{B}, \tilde{x}, h) \right\}.$$

Preliminary results

Lemma 7. *There exists a constant $\tilde{C} > 0$ such that for all $\theta = (B, x) \in \Theta$,*

$$\mathcal{T}(\tilde{B}_n(\theta), \tilde{x}_n(\theta), h - \tilde{C}n^{-\eta}) \subset \mathcal{T}(B, x, h) \subset \mathcal{T}(\tilde{B}_n(\theta), \tilde{x}_n(\theta), h + \tilde{C}n^{-\eta})$$

Proof of Lemma 7 – From the reverse triangular inequality, $\|B^\top(y-x)\| \geq |T_{n,1}(\theta) - T_{n,2}(\theta)|$, where

$$T_{n,1}(y, \theta) := \left\| [\tilde{B}_n(\theta)]^\top (\tilde{x}_n(\theta) - y) \right\|$$

and $T_{n,2}(y, \theta) := \left\| (B - \tilde{B}_n(\theta))^\top (y - x) + [\tilde{B}_n(\theta)]^\top (x - \tilde{x}_n(\theta)) \right\|.$

Now, remark that there exists a constant $K_0 > 0$ such that $\|(B - \tilde{B}_n(\theta))^\top\| \leq K_0 \|B - \tilde{B}_n(\theta)\|$. In particular, $\|[\tilde{B}_n(\theta)]^\top\| \leq K_0$.

Furthermore, since x and y belong to the compact \mathcal{X} , we have $\|x - y\| \leq 2N(\mathcal{X})$ where $N(\mathcal{X}) = \sup_{x \in \mathcal{X}} \|x\| < \infty$. As a consequence, letting $\tilde{C} := K_0 \max(2N(\mathcal{X}), 1)$, we get

$$T_{n,2}(y, \theta) \leq \tilde{C} \delta(\tilde{\theta}_n(\theta), \theta) \leq \tilde{C} n^{-\eta}.$$

Hence, if $T_{n,1}(y, \theta) > h + \tilde{C} n^{-\eta}$ then $\|B^\top(y-x)\| > h$. In other word, if $y \notin \mathcal{T}(\tilde{B}_n(\theta), \tilde{x}_n(\theta), h + \tilde{C} n^{-\eta})$ then $y \notin \mathcal{T}(B, x, h)$ proving the second inclusion. Next, if $T_{n,1}(y, \theta) \leq h - \tilde{C} n^{-\eta}$ then $\|B^\top(y-x)\| \leq \min(h - \tilde{C} n^{-\eta}, \tilde{C} n^{-\eta}) < h$. This entails that if $y \in \mathcal{T}(\tilde{B}_n(\theta), \tilde{x}_n(\theta), h - \tilde{C} n^{-\eta})$ then $y \in \mathcal{T}(B, x, h)$ proving the first inclusion and completing the proof. \blacksquare

Lemma 8. *Assume that conditions (H.1) and (H.2) holds. If $nh^q / \ln(n) \rightarrow \infty$ as $n \rightarrow \infty$ then for all integer K such that $K > \lfloor c/(\eta - 1) \rfloor$,*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \tilde{D}_n(\tilde{B}, \tilde{x}) \geq K \right) = 0.$$

Proof of Lemma 8 – It suffices to prove that for all $K > c/(\eta - 1)$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \tilde{D}_n(\tilde{B}, \tilde{x}) \geq K \right) &\leq \lim_{n \rightarrow \infty} \sum_{(\tilde{B}, \tilde{x}) \in \Theta_n} \mathbb{P} \left(\tilde{D}_n(\tilde{B}, \tilde{x}) \geq K \right) \\ &\leq \lim_{n \rightarrow \infty} \text{card}(\Theta_n) \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \mathbb{P} \left(\tilde{D}_n(\tilde{B}, \tilde{x}) \geq K \right) = 0. \end{aligned}$$

Let $h_\pm := h \pm \tilde{C} n^{-\eta}$ and let I be the set of all list $\mathbf{i} := \{i_1, \dots, i_K\}$ of K distinct indices in $\{1, \dots, n\}^K$.

$$\begin{aligned} \mathbb{P} \left(\tilde{D}_n(\tilde{B}, \tilde{x}) \geq K \right) &= \mathbb{P} \left(\bigcup_{\mathbf{i} \in I} \bigcap_{s=1}^K \{h_- < \|\tilde{B}(X_{i_s} - \tilde{x})\| \leq h_+\} \right) \\ &= \binom{n}{K} \left[\mathbb{P} \left(h_- < \|\tilde{B}(X - \tilde{x})\| \leq h_+ \right) \right]^K. \end{aligned}$$

From (15) in the proof of Lemma 5,

$$\begin{aligned} \mathbb{P}\left(h_- < \|\tilde{B}(X - \tilde{x})\| \leq h_+\right) &= \mathbb{P}\left(\|\tilde{B}(X - \tilde{x})\| \leq h_+\right) - \mathbb{P}\left(\|\tilde{B}(X - \tilde{x})\| \leq h_-\right) \\ &= \vartheta_q f_{\tilde{B}}\left(\tilde{B}^\top \tilde{x}\right) [(h_+)^q - (h_-)^q] + (h_+)^q R_+ - (h_-)^q R_-, \end{aligned}$$

where

$$R_{\pm} := \int_{\mathcal{U}_q} \left[f_{\tilde{B}}\left(\tilde{B}^\top \tilde{x} + h_{\pm} u\right) - f_{\tilde{B}}\left(\tilde{B}^\top \tilde{x}\right) \right] du.$$

First remark that

$$(h_+)^q - (h_-)^q = h^q \left[\left(1 + \tilde{C} \frac{n^{-\eta}}{h}\right)^q - \left(1 - \tilde{C} \frac{n^{-\eta}}{h}\right)^q \right].$$

Since $\eta > q$, we know that $n^{-\eta}/h \rightarrow 0$ and thus,

$$(h_+)^q - (h_-)^q = \mathcal{O}(h^{q-1} n^{-\eta}). \quad (20)$$

Furthermore, $(h_+)^q R_+ - (h_-)^q R_- = (h_+)^q R_{1,+} - (h_-)^q R_{1,-} + [(h_+)^q - (h_-)^q] R_2$,

where

$$\begin{aligned} R_{1,\pm} &:= \int_{\mathcal{U}_q} \left[f_{\tilde{B}}\left(\tilde{B}^\top \tilde{x} + h_{\pm} u\right) - f_{\tilde{B}}\left(\tilde{B}^\top \tilde{x} + hu\right) \right] du \\ \text{and } R_2 &:= \int_{\mathcal{U}_q} \left[f_{\tilde{B}}\left(\tilde{B}^\top \tilde{x} + hu\right) - f_{\tilde{B}}\left(\tilde{B}^\top \tilde{x}\right) \right] du. \end{aligned}$$

Mimicking the proof of (16) in the proof of Lemma 5, we have

$$|R_{1,\pm}| \leq \vartheta_q K_f \tilde{C} n^{-\eta} \quad \text{and} \quad |R_2| \leq \vartheta_q K_f h.$$

From (20), $(h_+)^q R_+ - (h_-)^q R_- = \mathcal{O}(h^q n^{-\eta})$ and, since from (H.2) $|f_B(B^\top x)|$ is uniformly bounded on $(B, x) \in \Theta$,

$$\mathbb{P}\left(h_- \leq \|\tilde{B}(X - \tilde{x})\| \leq h_+\right) = \mathcal{O}(h^{q-1} n^{-\eta}) + \mathcal{O}(h^q n^{-\eta}) = \mathcal{O}(n^{-\eta}). \quad (21)$$

Since $\binom{n}{K} = \mathcal{O}(n^K)$ and $\text{card}(\Theta_n) = \mathcal{O}(n^c)$, we get

$$\text{card}(\Theta_n) \sup_{(\tilde{B}, \tilde{x}) \in \Theta_n} \mathbb{P}\left(\tilde{D}_n(\tilde{B}, \tilde{x}) \geq K\right) = \mathcal{O}\left(n^{c+K(1-\eta)}\right).$$

We conclude the proof by remarking that the condition $K > \lfloor c/(\eta - 1) \rfloor$ entails that $n^{c+K(1-\eta)} \rightarrow 0$. ■

Proof of Proposition 5

We keep the notations used for the definition of the conditional tail-index estimator in Section 3.3. In particular, $M^* = M(B, x, h)$ and $Z_i^* = Z_i(B, x, h)$. In addition, for all $\theta = (B, x) \in \Theta$, let $\widetilde{M}_n^\pm := M(\widetilde{B}_n(\theta), \widetilde{x}_n(\theta), h_\pm)$ and for $i = 1, \dots, \widetilde{M}_n^\pm$, let $\widetilde{Z}_i^{*,\pm} := Z_{(\widetilde{M}_n^\pm - i + 1)}(\widetilde{B}_n(\theta), \widetilde{x}_n(\theta), h_\pm)$. Moreover, for $K > \lfloor c/(\eta - 1) \rfloor$, let

$$\mathfrak{C}_n(K) := \left\{ \sup_{(\widetilde{B}, \widetilde{x}) \in \Theta_n} \widetilde{D}_n(\widetilde{B}, \widetilde{x}) < K \right\} \cap \mathfrak{A}_n^+ \cap \mathfrak{A}_n^-,$$

where

$$\mathfrak{A}_n^\pm := \bigcap_{(\widetilde{B}, \widetilde{x}) \in \Theta_n} \left\{ \frac{M(\widetilde{B}, \widetilde{x}, h_\pm)}{n(h_\pm)^q} \in \left[\frac{C_1 \vartheta_q}{2}, \frac{3C_2 \vartheta_q}{2} \right] \right\},$$

From Lemma 5 and Lemma 8, we know that $\mathbb{P}(\mathfrak{C}_n(K)) \rightarrow 1$. In what follows, we assume that the event $\mathfrak{C}_n(K)$ holds. We have the following inequalities: for all $(B, x) \in \Theta$,

$$\begin{aligned} M^* - K + 1 &\leq \widetilde{M}_n^- \leq M^*, & M^* &\leq \widetilde{M}_n^+ \leq M^* + K - 1, \\ \text{and } \frac{C_1 \vartheta_q}{2} n(h_-)^q &\leq M^* \leq \frac{3C_2 \vartheta_q}{2} n(h_+)^q. \end{aligned} \quad (22)$$

As a straightforward consequence, for n large enough and uniformly on $(B, x) \in \Theta$, we have $\lfloor \alpha M^* \rfloor < \widetilde{M}_n^- < \widetilde{M}_n^+$.

Now, Lemma 7 ensures that for all $i \in \{1, \dots, \lfloor \alpha M^* \rfloor\}$,

$$\widetilde{Z}_{(\widetilde{M}_n^- - i + 1)}^{*,-} \leq Z_{(M^* - i + 1)}^* \leq \widetilde{Z}_{(\widetilde{M}_n^+ - i + 1)}^{*,+}. \quad (23)$$

Moreover, for all $k_+ \geq \lfloor \alpha M^* \rfloor + K$, we have $Z_{(M^* - \lfloor \alpha M^* \rfloor)}^* \geq \widetilde{Z}_{(\widetilde{M}_n^+ - k_+)}^{*,+}$. Indeed, the largest value of $\widetilde{Z}_{(\widetilde{M}_n^+ - k_+)}^{*,+}$ is obtained when $\widetilde{M}_n^+ = M^* + K - 1$ with the additional $(K - 1)$ observations larger than the maximal observation $Z_{(M^*)}^*$. In the same way, for all $k_- \leq \lfloor \alpha M^* \rfloor - K + 1$, we have $Z_{(M^* - \lfloor \alpha M^* \rfloor)}^* \leq \widetilde{Z}_{(\widetilde{M}_n^- - k_-)}^{*,-}$. Note that for n large enough and uniformly on $(B, x) \in \Theta$, $\lfloor \alpha M^* \rfloor > K - 1$. To sum up, for $k_+ \geq \lfloor \alpha M^* \rfloor + K$, and $k_- \leq \lfloor \alpha M^* \rfloor - K + 1$, we have for n large enough, that

$$\widetilde{Z}_{(\widetilde{M}_n^- - k_-)}^{*,-} \geq Z_{(M^* - \lfloor \alpha M^* \rfloor)}^* \geq \widetilde{Z}_{(\widetilde{M}_n^+ - k_+)}^{*,+}. \quad (24)$$

Taking $\alpha_+ := \alpha + 2(K + 2)/[C_1 \vartheta_q n(h_+)^q]$, we have

$$\lfloor \alpha_+ \widetilde{M}_n^+ \rfloor \geq \alpha \widetilde{M}_n^+ + \frac{2(K + 2)}{C_1 \vartheta_q n(h_+)^q} \widetilde{M}_n^+ + 1 \geq \alpha \widetilde{M}_n^+ + K + 1 \geq \lfloor \alpha M^* \rfloor + K.$$

Similarly, we show that for $\alpha_- := \alpha - 2(K + 1)/[C_1 \vartheta_q n(h_-)^q]$,

$$\lfloor \alpha_- \widetilde{M}_n^- \rfloor \leq \lfloor \alpha M^* \rfloor - K + 1.$$

We can thus use inequality (23) and inequality (24) with $k_- = \lfloor \alpha_- \widetilde{M}_n^- \rfloor$ and $k_+ = \lfloor \alpha_+ \widetilde{M}_n^+ \rfloor$ to obtain for n large enough and uniformly on $\theta = (B, x) \in \Theta$,

$$\begin{aligned} \frac{\lfloor \alpha_- \widetilde{M}_n^- \rfloor}{\lfloor \alpha M^* \rfloor} \widehat{\xi}_{\widetilde{B}_n(\theta)}^{(H)} \left([\widetilde{B}_n(\theta)]^\top \widetilde{x}_n(\theta); \alpha_-, h_- \right) &\leq \widehat{\xi}_B^{(H)} \left(B^\top x; \alpha, h \right) \\ &\leq \frac{\lfloor \alpha_+ \widetilde{M}_n^+ \rfloor}{\lfloor \alpha M^* \rfloor} \widehat{\xi}_{\widetilde{B}_n(\theta)}^{(H)} \left([\widetilde{B}_n(\theta)]^\top \widetilde{x}_n(\theta); \alpha_+, h_+ \right). \end{aligned}$$

Note from (22), it easy to check that there exist constants $C_- > 0$ and $C_+ > 0$ such that for all $(B, x) \in \Theta$

$$1 - \frac{C_-}{nh^q} \leq \frac{\lfloor \alpha_- \widetilde{M}_n^- \rfloor}{\lfloor \alpha M^* \rfloor} \leq 1 \text{ and } 1 \leq \frac{\lfloor \alpha_+ \widetilde{M}_n^+ \rfloor}{\lfloor \alpha M^* \rfloor} \leq 1 + \frac{C_+}{nh^q}$$

Hence, uniformly on $(B, x) \in \Theta$,

$$\begin{aligned} \left(1 - \frac{C_-}{nh^q} \right) \widehat{\xi}_{\widetilde{B}_n(\theta)}^{(H)} \left([\widetilde{B}_n(\theta)]^\top \widetilde{x}_n(\theta); \alpha_-, h_- \right) &\leq \widehat{\xi}_B^{(H)} \left(B^\top x; \alpha, h \right) \\ &\leq \left(1 + \frac{C_+}{nh^q} \right) \widehat{\xi}_{\widetilde{B}_n(\theta)}^{(H)} \left([\widetilde{B}_n(\theta)]^\top \widetilde{x}_n(\theta); \alpha_+, h_+ \right). \end{aligned} \quad (25)$$

Recall that all the previous results hold under the event $\mathfrak{C}_n(K)$. We thus write for all $\varepsilon > 0$,

$$\mathbb{P} \left(\sup_{\theta=(B,x) \in \Theta} \left| \widehat{\xi}_{\widetilde{B}_n(\theta)}^{(H)} \left([\widetilde{B}_n(\theta)]^\top \widetilde{x}_n(\theta) \right) - \widehat{\xi}_B^{(H)} \left(B^\top x \right) \right| > \varepsilon \right) \leq T_n + (1 - \mathbb{P}(\mathfrak{C}_n(K))),$$

where

$$T_n := \mathbb{P} \left(\left\{ \sup_{\theta=(B,x) \in \Theta} \left| \widehat{\xi}_{\widetilde{B}_n(\theta)}^{(H)} \left([\widetilde{B}_n(\theta)]^\top \widetilde{x}_n(\theta) \right) - \widehat{\xi}_B^{(H)} \left(B^\top x \right) \right| > \varepsilon \right\} \cap \mathfrak{C}_n(K) \right).$$

Since $\mathbb{P}(\mathfrak{C}_n(K)) \rightarrow 1$, we only need to focus on T_n . We have $T_n \leq T_{n,1} + T_{n,2}$ where

$$T_{n,1} = \mathbb{P} \left(\left\{ \sup_{\theta=(B,x) \in \Theta} \left| \widehat{\xi}_B^{(H)} \left(B^\top x \right) - \xi_{\widetilde{B}_n(\theta)} \left([\widetilde{B}_n(\theta)]^\top \widetilde{x}_n(\theta) \right) \right| > \varepsilon/2 \right\} \cap \mathfrak{C}_n(K) \right)$$

$$\text{and } T_{n,2} = \mathbb{P} \left(\sup_{(\widetilde{B}, \widetilde{x}) \in \Theta_n} \left| \widehat{\xi}_{\widetilde{B}}^{(H)} \left(\widetilde{B}^\top \widetilde{x} \right) - \xi_{\widetilde{B}} \left(\widetilde{B}^\top \widetilde{x} \right) \right| > \varepsilon/2 \right).$$

From Proposition 4, we have $T_{n,2} \rightarrow 0$. It thus remain to show that $T_{n,1} \rightarrow 0$.

Using (25), $T_{n,1} \leq T_{n,3}^+ + T_{n,3}^-$ where

$$T_{n,3}^\pm = \mathbb{P} \left(\sup_{(\widetilde{B}, \widetilde{x}) \in \Theta_n} \left| \left(1 \pm \frac{C_\pm}{nh^q} \right) \widehat{\xi}_{\widetilde{B}}^{(H)} \left(\widetilde{B}^\top \widetilde{x}; \alpha_\pm, h_\pm \right) - \xi_{\widetilde{B}} \left(\widetilde{B}^\top \widetilde{x} \right) \right| > \varepsilon/2 \right).$$

Since $h_\pm \sim h$, $\alpha_\pm \sim \alpha$ and $nh^q \rightarrow \infty$, we can use Proposition 4 to conclude the proof. ■