



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-04589742v2

HAL Id: hal-04589742

<https://hal.science/hal-04589742v2>

Preprint submitted on 10 Dec 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 ¹ and A. Podgorny¹

¹*Université de Strasbourg & CNRS, IRMA, UMR 7501,
7 rue René Descartes, 67084 Strasbourg Cedex, France*

Abstract

We are interested in the relationship between the large values of a real random variable Y and its associated covariate X , which takes values in a subset \mathcal{X} of \mathbb{R}^p , when the conditional distribution of Y given $X = x$ is heavy-tailed with a conditional tail-index $\gamma(x) > 0$. Estimating this index is a critical step for inferring the conditional distribution, but the task becomes increasingly challenging as the dimension p grows. The objective of this work is to propose a dimension reduction method to obtain a more efficient estimator of $\gamma(x)$. 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. Additionally, we introduce an estimator of the conditional tail-index that leverages this dimension reduction and prove its consistency. We demonstrate the advantages of this dimension reduction approach for estimating the extreme value index through simulations and an application to real-world 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 [13] and [21] and the nearest-neighbors approach (see for instance [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 [2]). A solution to avoid this problem is to assume the existence of a mean dimension reduction subspace (see [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., [15]) or insurance (see for instance [4, 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 [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 [1]. Their approach can be seen as an adaptation of the Sliced Inverse Regression method introduced by [12] and differs significantly from [7]. In particular, condition (1) and the specific condition used in [1] are not equivalent. Moreover, the final objective in [7] is the estimation of extreme quantile while the one in [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 [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 [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} which is then used to estimate the conditional tail-index. The convergences in probability of the proposed estimators towards their target are established under certain mild conditions. 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. The proofs are provided in Section 6, except for the proof of Theorem 3 (see Section 3.3) which can be found in the supplementary file.

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) follows the distribution described by the following model.

- (M) The support \mathcal{X} of X is 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 a position to present a result that will play a central role in defining 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} .

The CTI subspace is a straightforward adaptation of the central mean subspace as defined in [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 and the conditional tail-index

Throughout this section, we assume that the CTI subspace \mathcal{S}_0 exists and that its dimension, q , is known. As previously mentioned, the canonical basis of \mathcal{S}_0 is denoted by B_0 . Naturally, knowing the CTI dimension is unrealistic in practice; therefore, its estimation is addressed in the simulation study (see Section 4). Furthermore, as with most dimension reduction methods in the literature, verifying the existence of \mathcal{S}_0 is not straightforward. In Section 5, we present a graphical tool that can help assess whether this condition is satisfied.

This section is organized as follows. In Section 3.1, we start by introducing an estimator of the CTI subspace \mathcal{S}_0 . Next, given an arbitrary estimator $\widehat{\xi}_{n,B}(B^\top x)$ of $\xi_B(B^\top x)$, we estimate the conditional tail-index $\gamma(x) = \xi_{B_0}(B_0^\top x)$ by

$$\widehat{\gamma}_n(x, \widehat{B}_n) := \widehat{\xi}_{n, \widehat{B}_n}(\widehat{B}_n^\top x),$$

where \widehat{B}_n is the canonical basis of the estimated CTI subspace. In Section 3.2, we establish the consistency of the CTI subspace estimator. More specifically, we prove that

$$\left\| \widehat{B}_n - B_0 \right\| \xrightarrow{\mathbb{P}} 0,$$

for any norm $\|\cdot\|$ in $\mathbb{R}^{p \times q}$. Next, for a given compact subset \mathcal{X}_0 in the interior of \mathcal{X} , we establish the uniform consistent over the set \mathcal{X}_0 of the conditional tail-index estimator:

$$\sup_{x \in \mathcal{X}_0} \left| \widehat{\xi}_{n, \widehat{B}_n}(\widehat{B}_n^\top x) - \gamma(x) \right| \xrightarrow{\mathbb{P}} 0.$$

This last convergence is obtained under the following condition on the subspace \mathcal{X}_0 .

(C.3) If there exists $B \in \mathcal{B}_q$ such that $\xi_B(B^\top x) = \gamma(x)$ for all $x \in \mathcal{X}_0$ then $B = B_0$.

This condition prevents, for instance, the pathological case where $\gamma(\cdot)$ is constant over the set \mathcal{X}_0 , a scenario that can arise even if the CTI subspace exists.

3.1 Construction and definition of the estimators

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 $B^\top X = B^\top x$ with right endpoint $\xi_B(B^\top x)$. If condition (C.3) holds for a compact subset \mathcal{X}_0 in the interior of \mathcal{X} then

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

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

$$\Psi(B, \mathcal{X}_0) := \mathbb{E} \left[\xi_B(B^\top X) \mathbb{I}\{X \in \mathcal{X}_0\} \right].$$

The empirical counterpart of $\Psi(B, \mathcal{X}_0)$ is given by

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

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 $\sup_{x \in \mathcal{X}} M(x) = \bar{M} < \infty$ 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

$$\hat{\Psi}_n(B, \mathcal{X}_0) := \frac{1}{n} \sum_{i=1}^n \hat{\xi}_{n,B}(B^\top X_i) \mathbb{I}\{X_i \in \mathcal{X}_0\}. \quad (3)$$

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 $B \mapsto \hat{\Psi}_n(B, \mathcal{X}_0)$.

Our estimator thus belongs to the class of M-estimators. Note that, in a complete different setting, [11] also consider an M-estimator for the estimation of the direction of a single-index model. As mentioned in the introduction of this section, we estimate the conditional tail-index $\gamma(x)$ by

$$\widehat{\gamma}_n(x, \widehat{B}_n) := \widehat{\xi}_{n, \widehat{B}_n}(\widehat{B}_n^\top x).$$

The theoretical study of \widehat{B}_n and $\widehat{\gamma}_n(x, \widehat{B}_n)$ is done in the next section.

3.2 Main results

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

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

then the estimator \widehat{B}_n in Definition 4 is consistent. Such an estimator is given in Section 3.3.

Theorem 1. *Assume that model (M) holds and let \mathcal{X}_0 be a compact subset in the interior of \mathcal{X} satisfying (C.3). Under conditions (C.1), (C.2) and (C.4), if the conditional tail-index estimator appearing in $\widehat{\Psi}_n(\cdot, \mathcal{X}_0)$ satisfies (C.5) then $\|\widehat{B}_n - B_0\| \xrightarrow{\mathbb{P}} 0$ for any norm $\|\cdot\|$ in $\mathbb{R}^{p \times q}$.*

The uniform consistency on \mathcal{X}_0 of the conditional tail-index estimator $\widehat{\gamma}_n(\cdot, \widehat{B}_n)$ can be deduced from that of \widehat{B}_n .

Theorem 2. *Assume that model (M) holds and let \mathcal{X}_0 be a compact subset in the interior of \mathcal{X} satisfying (C.3). Under conditions (C.1), (C.2) and (C.4), if the conditional tail-index estimator $\widehat{\xi}_{n, B}(B^\top x)$ satisfies (C.5) then*

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

The next section is dedicated to the definition and the theoretical study of a conditional tail-index estimator satisfying condition (C.5).

3.3 A local-Hill estimator

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 a consistent 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 $\alpha \in (1/m, 1)$. 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

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

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 approximatively 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 conditional 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 [8]. Under additional assumptions given below, we show that this estimator is uniformly consistent over the

set $\Theta := \mathcal{B}_q \times \mathcal{X}_0$ where \mathcal{X}_0 is any compact subset in the interior of \mathcal{X} . 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 similar to 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 $\Theta = \mathcal{B}_q \times \mathcal{X}_0$ 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 $\mathcal{B}_q \times \mathcal{X} \ni (B, x) \mapsto \xi_B(B^\top x)$ 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) $\lim_{y \rightarrow \infty} \sup_{B \in \mathcal{B}_q} \sup_{x \in \mathcal{X}} \Delta_{B, x}(y) = 0$.

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

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

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

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

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

Theorem 3. *Assume that model (M) holds and let \mathcal{X}_0 be a compact subset in the interior of \mathcal{X} satisfying (C.3). Under 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,*

$$\sup_{B \in \mathcal{B}_q} \sup_{x \in \mathcal{X}_0} \left| \hat{\xi}_B^{(H)}(B^\top x) - \xi_B(B^\top x) \right| \xrightarrow{\mathbb{P}} 0.$$

The proof of Theorem 3 is quite long and technical. It is available in the supplementary file. 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 and the estimator of $\gamma(x)$ 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 $\hat{B}_n^{(H)}$ of the canonical basis of CTI subspace minimizes the function*

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

The associated conditional tail-index estimator is given for $x \in \mathcal{X}_0$ by

$$\hat{\gamma}_n^{(H)}(x, \hat{B}_n^{(H)}) := \hat{\xi}_{\hat{B}_n^{(H)}}^{(H)}([\hat{B}_n^{(H)}]^\top x).$$

The consistencies of the estimators $\hat{B}_n^{(H)}$ and $\hat{\gamma}_n^{(H)}(x, \hat{B}_n^{(H)})$ follow directly from Theorems 1, 2 and 3.

Corollary 1. *Assume that model (M) holds and let \mathcal{X}_0 be a compact subset in the interior of \mathcal{X} satisfying (C.3). Under conditions (C.1) 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,*

$$\|\hat{B}_n^{(H)} - B_0\| \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \sup_{x \in \mathcal{X}_0} \left| \hat{\gamma}_n^{(H)}(x, \hat{B}_n^{(H)}) - \gamma(x) \right| \xrightarrow{\mathbb{P}} 0.$$

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 the estimator $\hat{\gamma}_n^{(H)}(x, \hat{B}_n^{(H)})$, where $\hat{B}_n^{(H)}$ denotes the local-Hill CTI estimator of B_0 , see Definition 6. 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. A short discussion about the choice of the tuning parameters required to compute these estimators is provided in Section 4.1. 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. The **R** source codes used in the simulation study as well as for the analysis of real data are available on a GitHub page¹.

4.1 Tuning parameters

The local-Hill estimator of the conditional tail-index $\hat{\gamma}_n^{(H)}(\cdot, \hat{B}_n^{(H)})$ depends on two tuning parameters. The first one is the bandwidth h used to select the observations closest to the point of interest. We propose to take h in such a way that the number of points in the ball of center $[\hat{B}_n^{(H)}]^\top x$ and radius h is proportional to n^{1-b} for some $b \in (0, 1)$. In view of Lemma 3 in the supplementary file, we propose to take $h = (n^{-b}/\vartheta_q)^{1/q}$ where ϑ_q is the volume of the unit q -ball. The selection of h thus reduces to the one of $b \in [0, 1)$. In all what follows, we use the uniform norm $\|\cdot\|_\infty$ for which $\vartheta_q = 2^q$.

The second tuning parameter is the probability $\alpha \in (0, 1)$ controlling the number of largest observations. We take $\alpha = n^{-a}$ for some $a \in (0, 1)$.

Proposing a method to select the parameters a and b is beyond the scope of this paper and left for future work. In Section 4.4, we consider a grid of values for these two tuning parameters and keep the couple (a, b) leading to the best estimation, see Section 4.4 for more details.

A third tuning parameter $n_0 \ll n$ is considered in order to reduce the computational

¹<https://github.com/Alex-Podgorny/Dimension-Reduction-Tail-Index/tree/master>

burden associated with the evaluation of $\widehat{\Psi}_n^{(H)}(B, \mathcal{X}_0)$. More precisely, we randomly select n_0 observations $\tilde{X}_1, \dots, \tilde{X}_{n_0}$ in the original sample $\{X_1, \dots, X_n\}$ and replace the evaluation of $\widehat{\Psi}_n^{(H)}(B, \mathcal{X}_0)$ by the one of

$$\tilde{\Psi}_{n_0}^{(H)}(B, \mathcal{X}_0) := \frac{1}{n_0} \sum_{i=1}^{n_0} \tilde{\xi}_B^{(H)}(B^\top \tilde{X}_i) \mathbb{I}\{\tilde{X}_i \in \mathcal{X}_0\}.$$

This approach significantly decreases the computation time without noticeably affecting the accuracy of the results, this is also used in [7]. If $n_0 \sim cn$ for some constant $c \in (0, 1)$, the theoretical results presented above remain unchanged. We set $n_0 = \lfloor n/20 \rfloor$, which provides a good trade-off between computational efficiency and precision.

4.2 Finding the solution of the optimization problem

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

$$(P) \quad \min_{B \in \mathbb{R}^{p \times q}} \tilde{\Psi}_{n_0}^{(H)}(B, \mathcal{X}_0) \quad \text{subject to} \quad B \in \mathcal{B}_q.$$

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

$$(P') \quad \min_{B \in \mathbb{R}^{p \times q}} \tilde{\Psi}_{n_0}^{(H)}(B, \mathcal{X}_0) \quad \text{subject to} \quad B^\top B = I_q.$$

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

$$\tilde{\Psi}_{n_0}^{(H)}(B', \mathcal{X}_0) \leq \tilde{\Psi}_{n_0}^{(H)}(B, \mathcal{X}_0).$$

To solve the constrained optimization problem (P') , we apply row by row the Simulated Annealing (SANN) algorithm (which is available in the `optim` function in `R`). The algorithm is described below:

1. Initialization: We generate 100 points from a Halton sequence (cf. [10]) in the hypercube $[-1, 1]^{pq}$. The function $\tilde{\Psi}_{n_0}^{(H)}(B, \mathcal{X}_0)$ is evaluated for each $p \times q$ matrix obtained with these points, and the matrix corresponding to the lowest value of $\tilde{\Psi}_{n_0}^{(H)}$ is selected as the initial matrix $B^{(0)}$.
2. For $i \geq 0$, until convergence:

- Let $B^{(i,0)} = B^{(i)}$.
- For $j = 0, \dots, p-1$:
 - ↔ optimize the function $B \mapsto \tilde{\Psi}_{n_0}^{(H)}(B, \mathcal{X}_0)$ over the $(j+1)$ th row of $B^{(i,j)}$ (all the other rows are fixed). We use the SANN algorithm with 20 iterations;
 - ↔ apply a QR decomposition on the optimized matrix to obtain an orthonormalized version denoted by $B^{(i,j+1)}$.
- Let $B^{(i+1)} = B^{(i,p)}$

The convergence is achieved when

$$\|P_{B^{(i+1)}} - P_{B^{(i)}}\|_2 < 10^{-2},$$

where $P_B = BB^\top$ is the projector on the subspace spanned by the orthogonal basis B and $\|\cdot\|_2$ denotes the spectral norm.

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 $p \in \{4, 30\}$. 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. Two different slowly varying functions are used in our simulation study:

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

with $B_1 = (0, 0, 5, 5, 0, \dots, 0)^\top \in \mathbb{R}^p$. 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 [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

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

4.4 Simulation results

Our aim is to compare the finite sample performance of a local-Hill type estimator $\hat{\gamma}_n^{(H)}(x, \check{B}_n)$ of the conditional tail-index $\gamma(x)$ with different estimated dimension reduction subspaces $\text{span}(\check{B}_n)$. More specifically, we consider 6 different estimated basis:

- $\check{B}_n = \hat{B}_n^{(H)}$, the proposed local-Hill CTI estimator, see Definition 6;
- $\check{B}_n = \hat{B}_n^{(G)}$, the TDR estimator proposed by Gardes [7];
- $\check{B}_n = \hat{B}_n^{(T_1)}$ and $\check{B}_n = \hat{B}_n^{(T_2)}$ which are the two estimators proposed by Aghbalou et al. [1] (respectively, TIREX1 and TIREX2 as named by the authors);
- $\check{B}_n = B_0$ corresponding to the ideal situation where the true basis B_0 is known;
- $\check{B}_n = I_p$ corresponding to the case where no dimension reduction is used.

Note that $\hat{B}_n^{(H)}$ is consistent only if $\gamma(x) > 0$ for all $x \in \mathcal{X}$ while $\hat{B}_n^{(G)}$, $\hat{B}_n^{(T_1)}$ and $\hat{B}_n^{(T_2)}$ are consistent regardless the sign of the conditional tail-index. The TDR estimator in [7] is also the solution of a constrained optimization problem which is solved with the procedure described in Section 4.2. Moreover, because the execution time increases exponentially with the dimension p , the estimator $\hat{B}_n^{(G)}$ is only considered in the case $p = 4$.

For each model, we generate $N = 100$ samples of size $n = 2000$ for $p = 4$ and $n = 4000$ for $p = 30$. The accuracy of the estimator $\hat{\gamma}_n^{(H)}(x, \check{B}_n)$ for the conditional tail-index $\gamma(x)$ is evaluated on the compact $\mathcal{X}_0 := [\varepsilon_p, 1 - \varepsilon_p]^p$. We take $\varepsilon_p = (1 - 0.9^p)/2$ so that approximatively 90% of observations belong to \mathcal{X}_0 .

As mentioned in Section 4.1, the tuning parameters are given by $h = n^{-b/q}/2$ and $\alpha = n^{-a}$ for $(a, b) \in (0, 1)^2$. The couple (a, b) is taken in the set

$$\mathcal{G} := \{0.25, 0.3, 0.35, 0.4, 0.45, 0.5\} \times \{0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4\}.$$

For $r \in \{1, \dots, N\}$, let $\hat{\gamma}_n^{(H,r)}(x, \check{B}_n)$ be the local-Hill estimator of $\gamma(x)$ evaluated on the r th replication. For each $r \in \{1, \dots, N\}$ and each $(a, b) \in \mathcal{G}$, we compute the error

$$\mathbb{E}^{(r)}(\check{B}_n, a, b) := \frac{1}{M} \sum_{i=1}^M \left(\hat{\gamma}_n^{(H,r)}(x_i, \check{B}_n) - \gamma(x_i) \right)^2,$$

where $M = 10.000$ and $\{x_i; i = 1, \dots, M\}$ are the points of a Halton sequence on the compact \mathcal{X}_0 . Note that $\mathbb{E}^{(r)}(\check{B}_n, a, b)$ is an approximation of the integral

$$\int_{\mathcal{X}_0} \left(\hat{\gamma}_n^{(H,r)}(x, \check{B}_n) - \gamma(x) \right)^2 dx.$$

For $r \in \{1, \dots, N\}$, let $\mathbb{E}^{(r,opt)}(\check{B}_n) := \mathbb{E}^{(r)}(\check{B}_n, \hat{a}_n^{(opt)}, \hat{b}_n^{(opt)})$ where

$$(\hat{a}_n^{(opt)}, \hat{b}_n^{(opt)}) := \arg \min_{(a,b) \in \mathcal{G}} \widehat{\text{MISE}}(\check{B}_n, a, b).$$

where

$$\widehat{\text{MISE}}(\check{B}_n, a, b) := \frac{1}{N} \sum_{r=1}^N \mathbb{E}^{(r)}(\check{B}_n, a, b),$$

is an estimation of the Mean Integrated Squared Error

$$\mathbb{E} \left[\int_{\mathcal{X}_0} \left(\hat{\gamma}_n^{(H,r)}(x, \check{B}_n) - \gamma(x) \right)^2 dx \right].$$

For each estimator \check{B}_n , the boxplot of the values $\{\mathbb{E}^{(r,opt)}(\check{B}_n); r = 1, \dots, N\}$ are given in Figure 1 and 2. The CTI approach consistently achieves lower errors compared to the alternatives, indicating its effectiveness in estimating the tail-index $\gamma(x)$ when dimension reduction is properly applied. As expected, the ideal scenario where $\check{B}_n = B_0$ achieves slightly better results, but this serves only as a benchmark, as the true subspace is generally unknown in practice. In contrast, using $\check{B}_n = I_p$ leads to significantly higher errors, underscoring the importance of reducing dimensionality for accurate tail-index estimation. For Model 2, where the CTI subspace is included in the dimension reduction subspace introduced in [7], the estimation of B_0 by $\hat{B}_n^{(G)}$ 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 $y = \pi/4$, $\hat{B}_n^{(G)}$, $\hat{B}_n^{(T_1)}$ and $\hat{B}_n^{(T_2)}$ are 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'

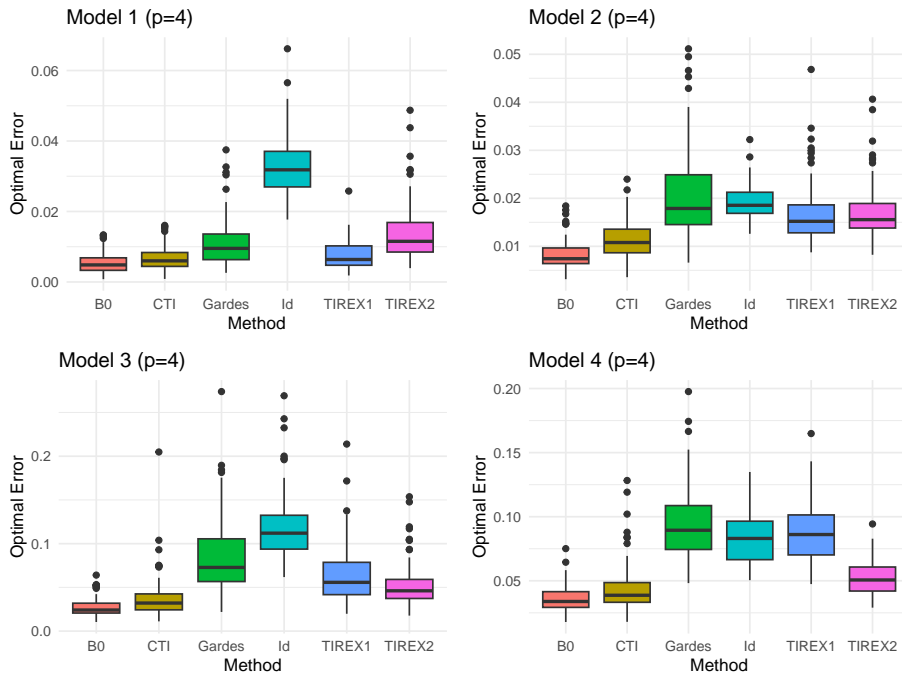


Figure 1: Boxplots of the values $\{E^{(r,opt)}(\check{B}_n); r = 1, \dots, N\}$ for the 6 choices for \check{B}_n on $N = 100$ replications of a sample of size $n = 2000$ from Models 1 to 4 with $p = 4$.

estimator and TIREX1 failed to estimate the dimension reduction subspace in Model 4 for which $q = 2$.

In addition to the boxplots, heatmaps are used to display the values of $E^{(r)}(\check{B}_n, a, b)$ over the entire grid of tuning parameters (a, b) . These heatmaps provide a detailed visualization of the sensitivity of each method to the choice of parameters. All the methods demonstrate robust performance across a wide range of tuning parameters. Moreover it seems that the choice of the optimal parameters does not strongly depend of the model. The choice $(a, b) = (0.3, 0.2)$ give good results in all cases for our estimation procedure, this is why we recommend it. Figure 3 shows the heatmaps for our method on the different models. The heatmaps for the other methods are available in the Github page mentioned in the introduction of Section 4.

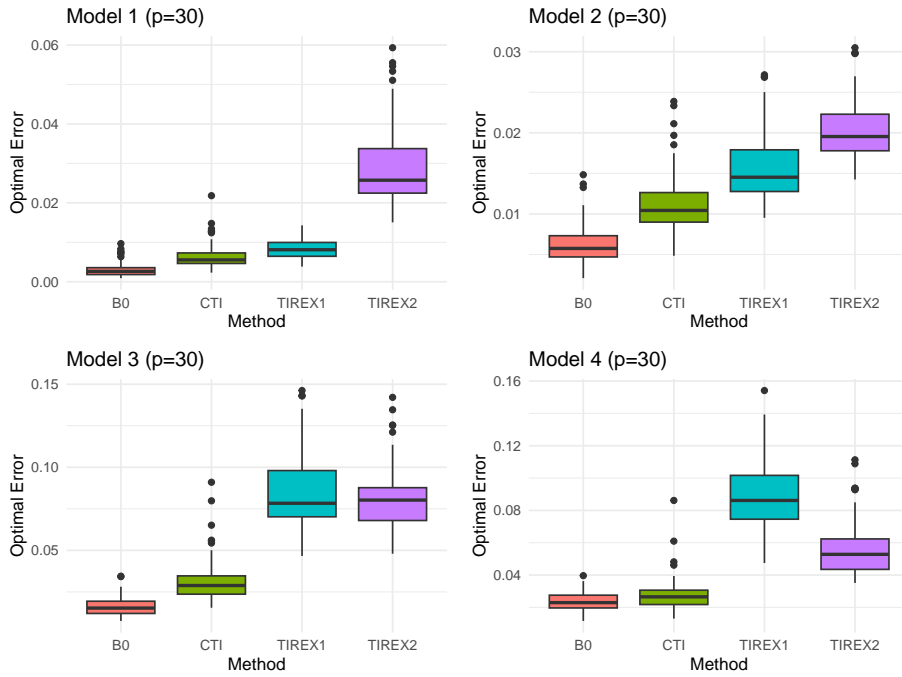


Figure 2: Boxplots of the values $\{E^{(r,opt)}(\check{B}_n); r = 1, \dots, N\}$ for different choices for \check{B}_n on $N = 100$ replications of a sample of size $n = 4000$ from Models 1 to 4 with $p = 30$. For the case $\check{B}_n = I_p$ the errors are very high and are not represented (For instance, medians for the four models are respectively 5.64, 0.33, 6.25, 5.91).

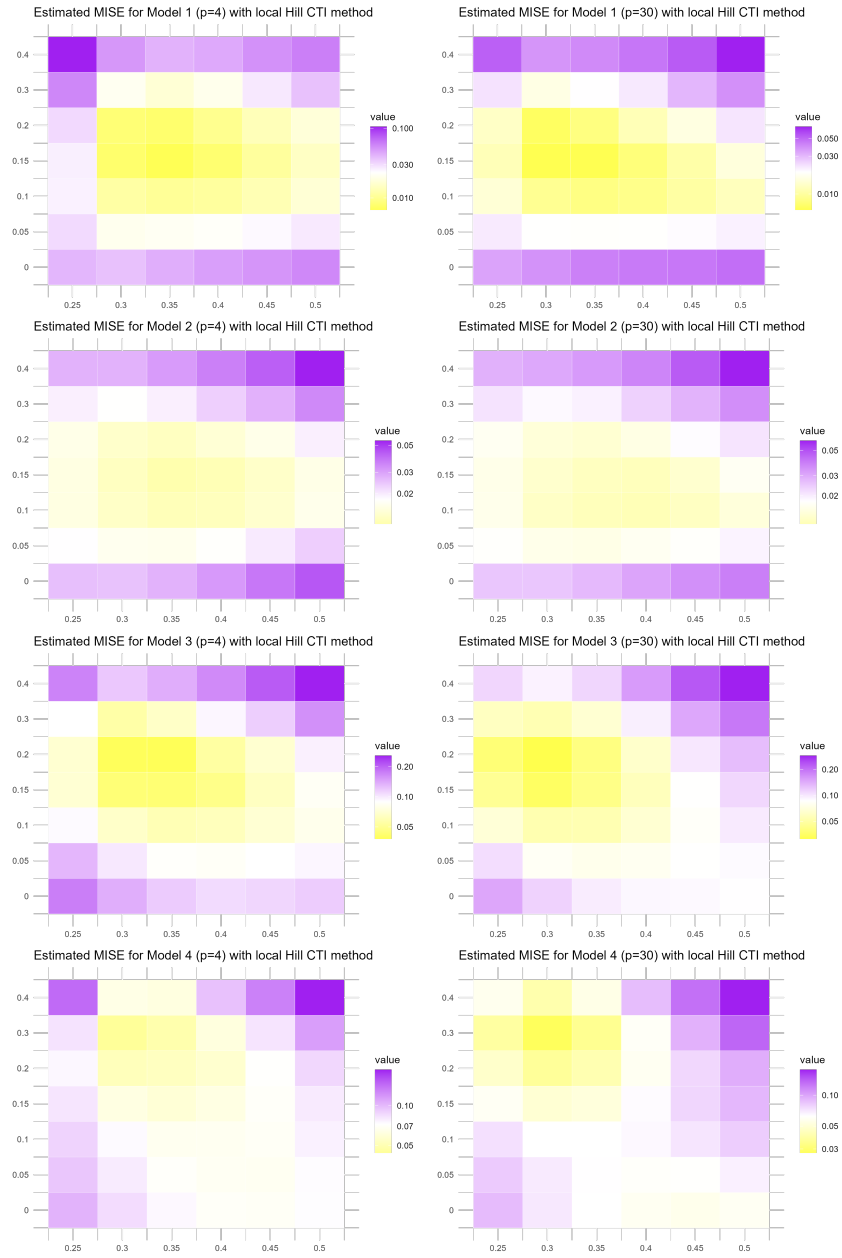


Figure 3: Estimated MISE values over the entire grid of tuning parameters (a, b) (a on x-axis, b on y-axis).

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. Our procedure of estimation lies on the following result. 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 $B \in \mathcal{B}_d$. Let also us introduce the notation,

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

Proposition 2. *Assume that (C.3) holds. If for all $d \in \{1, \dots, p-1\}$ there exists a matrix $C_{d+1} \in \mathcal{B}_{d+1}$ with $\text{span}(B_{0,d}) \subset \text{span}(C_{d+1})$ and such that*

$$\mathbb{P} \left[\{\xi_{C_{d+1}}(C_{d+1}^\top X) \neq \xi_{B_{0,d}}(B_{0,d}^\top X)\} \cap \{X \in \mathcal{X}_0\} \right] > 0,$$

then, for all $d \in \{1, \dots, q-1\}$, one has $c(d+1, \mathcal{X}_0) < c(d, \mathcal{X}_0)$ and for all $d \in \{q, \dots, p\}$, $c(d, \mathcal{X}_0) = \mathbb{E}[\gamma(X)\mathbb{I}\{X \in \mathcal{X}_0\}]$.

As a consequence, 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 (with tuning parameters $a = 0.3$ and $b = 0.2$) 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

$$\hat{c}_n(d, \mathcal{X}_0) = \min_{B \in \mathcal{B}_d} \tilde{\Psi}_{n_0}^{(H)}(B, \mathcal{X}_0).$$

As shown in Table 2, the worst result is obtained for Model 4 with $p = 4$ where our estimator failed to recover the true dimension in 18% of the replications.

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

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

We compute the estimated Mean Integrated Squared Error $\widehat{\text{MISE}}(\hat{B}_n^{(d,H)}, a, b)$ with the tuning parameters $a = 0.3$ and $b = 0.2$ as indicated in the previous section. As expected, the smallest value is observed when d is the true dimension; for all the consider models, using the estimated dimension \hat{q}_n instead of the true dimension provides very similar results.

Estimation of q				Estimated MISE			
Model	$\hat{q}_n = 1$	$\hat{q}_n = 2$	$\hat{q}_n = 3$	$d = 1$	$d = 2$	$d = 3$	$d = \hat{q}_n$
1 ($q = 1$)	100 %	0 %	0 %	0.012	0.015	0.026	0.012
2 ($q = 1$)	96 %	4 %	0 %	0.014	0.014	0.016	0.014
3 ($q = 1$)	99 %	1 %	0 %	0.066	0.075	0.107	0.065
4 ($q = 2$)	18 %	81 %	1 %	0.085	0.063	0.075	0.065

Table 1: 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: Estimated MISE of the local-Hill tail-index estimator $\hat{\gamma}_n^{(H)}(x, \hat{B}_n^{(d,H)})$ for different values of d .

Estimation of q				Estimated MISE			
Model	$\hat{q}_n = 1$	$\hat{q}_n = 2$	$\hat{q}_n = 3$	$d = 1$	$d = 2$	$d = 3$	$d = \hat{q}_n$
1 ($q = 1$)	99 %	1 %	0 %	0.010	0.013	0.020	0.010
2 ($q = 1$)	93 %	6 %	1 %	0.014	0.016	0.018	0.013
3 ($q = 1$)	98 %	2 %	0 %	0.053	0.066	0.090	0.052
4 ($q = 2$)	12 %	88 %	0 %	0.073	0.050	0.062	0.052

Table 2: Left part of the table: estimation of the CTI subspace dimension over $N = 100$ replications of a sample of size $n = 4000$ generated from Models 1 to 4 with $p = 30$. Right part of the table: Estimated MISE 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, 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 $X = (X_{NO_2}, X_{SO_2}, X_{PM_{10}}, X_{CO})^T \in [0, 1]^p$ with $p = 4$ is normalized version of the random vector $(NO_2, SO_2, PM_{10}, CO)$, where $X_\bullet = [\bullet - \max(\bullet)]/[\min(\bullet) - \max(\bullet)]$. 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 4, 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 estimation procedure proposed in this paper to this data set with $\mathcal{X}_0 := [0.02, 0.4]^p$. This compact contains around 88% of the observations. Following the recommendation made in the simulation study, the tuning parameters are set to $a = 0.3$ and $b = 0.2$. Applying the procedure described in Section 4.5, we find that $\hat{q}_n = 1$ (with $\hat{c}_n(1, \mathcal{X}_0) = 0.531$ and $\hat{c}_n(2, \mathcal{X}_0) = 0.614$). By applying the minimization

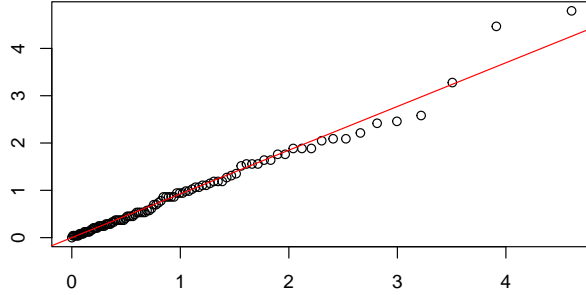


Figure 4: Pareto quantile plot (x-axis: $\ln(k/i)$, y-axis: $\ln(Y_{n-i,n}/Y_{n-k,n})$ for $i \in \{0, \dots, k = 200\}$. The fitted line is depicted in red.)

procedure described in Section 4.2, the estimated basis of the CTI subspace is given by

$$\widehat{B}_n^{(H)} = (0.831, 0.095, 0.500, 0.223)^\top.$$

Note that the two largest components of $\widehat{B}_n^{(H)}$ correspond to NO_2 and PM_{10} . In Figure 5, 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 conditional tail-index function. The estimated values $\widehat{\gamma}_n^{(H)}(x_i, \widehat{B}_n^{(H)})$ for $x_i \in \mathcal{X}_0$ as a function of $[\widehat{B}_n^{(H)}]^\top x_i$ are depicted in Figure 6. 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 [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.

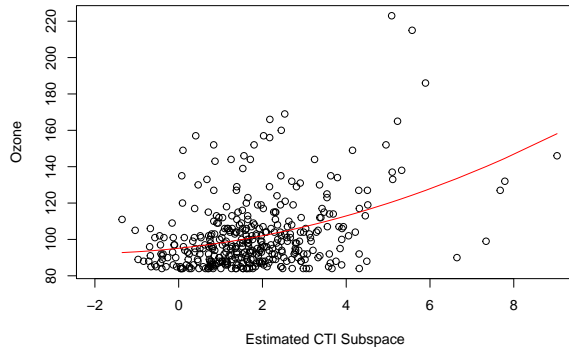


Figure 5: Largest observations of O_3 versus the projected covariates on the linear subspace spanned by $\hat{B}_n^{(H)}$. The smooth curve is obtained by fitting a polynomial of order 2.

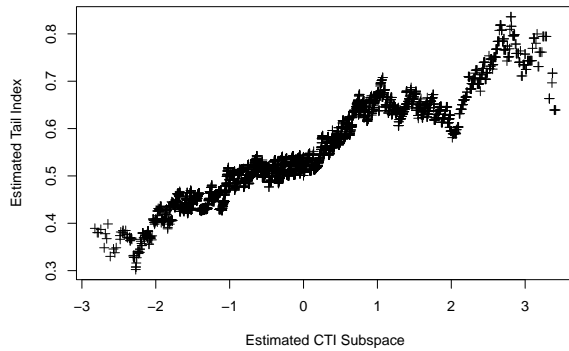


Figure 6: Estimated tail-index versus the projected covariates on the linear subspace spanned by $\hat{B}_n^{(H)}$.

Conclusion

When the tail-index function is positive and defined on a lower-dimensional linear subspace (referred to as the CTI subspace), this paper proposes an estimation procedure for a basis of this subspace, as well as an estimation of the conditional tail-index.

The consistency of these estimators is established, and a simulation study demonstrates that our method outperforms those of our main competitors. Additionally, we propose a method for estimating the unknown dimension of the dimension reduction subspace, which yields satisfactory results on synthetic data. The implementation of the estimation procedure described in this paper in an **R** package is currently in progress; however, all the codes are available on a GitHub page.

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

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

As a consequence,

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

leading to

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

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

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

Remarking that

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

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

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

which is the desired result. \blacksquare

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 $T_{1,n}$, we must prove a uniform law of large numbers for the set of parametric functions $\mathcal{F}_q := \{x \in \mathcal{X} \mapsto \xi_B(B^\top x) \mathbb{I}\{X \in \mathcal{X}_0\}; B \in \mathcal{B}_q\}$. This result is established by compiling different results that can be found for instance in the monograph of [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 compactness 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

$$y \in (0, \infty) \mapsto \mathbb{E}\left[c(X)y^{1/\gamma(X)} \mid B^\top X\right], \quad (4)$$

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

$$\lim_{y \rightarrow \infty} \frac{\mathbb{E}[\ln(Z/y) \mathbb{I}\{Z > y\} \mid B^\top X]}{\mathbb{P}(Z > y \mid B^\top X)} = \xi_B(B^\top X), \quad (5)$$

holds almost surely if and only if the function

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

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 $y > [c(x)]^{\gamma(x)}$,

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

proving (4) is thus equivalent to prove (5). 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

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

Moreover

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

Hence,

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

As a consequence, prove (5) 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 $t \in (0, \xi_B(B^\top X))$,

$$\lim_{y \rightarrow \infty} y^{1/t} \mathbb{P}(\{\gamma(X) < t\} \cap \{Z > y\} | B^\top X) = 0, \quad (6)$$

and

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

Let us start by proving (6). We have

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

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

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

proving (6). Let us now turn to the proof of (7). First, note that

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

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

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

Hence, for all $y > 0$,

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

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

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

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\inf_{B \notin U(B_0, \rho)} \widehat{\Psi}_n(B, \mathcal{X}_0) \leq \widehat{\Psi}_n(B_0, \mathcal{X}_0)\right) = 0, \quad (8)$$

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 (8), 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

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

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. ■

Proof of Theorem 2 – Let $\varepsilon > 0$. We need to prove that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sup_{x \in \mathcal{X}_0} \left| \hat{\gamma}_n(x, \hat{B}_n) - \gamma(x) \right| > \varepsilon \right] = \lim_{n \rightarrow \infty} \mathbb{P} \left[\sup_{x \in \mathcal{X}_0} \left| \hat{\xi}_{n, \hat{B}_n}([\hat{B}_n]^\top x) - \xi_{B_0}(B_0^\top x) \right| > \varepsilon \right] = 0.$$

We start with the following decomposition: let $\eta \in]0, \varepsilon/(2\bar{M})[$,

$$\mathbb{P} \left[\sup_{x \in \mathcal{X}_0} \left| \hat{\gamma}_n(x, \hat{B}_n) - \gamma(x) \right| > \varepsilon \right] =: T_{1,n} + T_{2,n},$$

where

$$T_{1,n} = \mathbb{P} \left[\left\{ \sup_{x \in \mathcal{X}_0} \left| \hat{\gamma}_n(x, \hat{B}_n) - \gamma(x) \right| > \varepsilon \right\} \cap \left\{ \|\hat{B}_n - B_0\| < \eta \right\} \right]$$

and $T_{2,n} = \mathbb{P} \left[\left\{ \sup_{x \in \mathcal{X}_0} \left| \hat{\gamma}_n(x, \hat{B}_n) - \gamma(x) \right| > \varepsilon \right\} \cap \left\{ \|\hat{B}_n - B_0\| \geq \eta \right\} \right]$

Of course, from Theorem 1,

$$T_{2,n} \leq \mathbb{P} \left(\|\hat{B}_n - B_0\| \geq \eta \right) \rightarrow 0.$$

Next,

$$\begin{aligned} T_{1,n} &\leq \mathbb{P} \left[\sup_{B: \|B - B_0\| < \eta} \sup_{x \in \mathcal{X}_0} \left| \hat{\xi}_{n,B}(B^\top x) - \xi_{B_0}(B_0^\top x) \right| > \varepsilon \right] \\ &\leq \mathbb{P} \left[\sup_{B: \|B - B_0\| < \eta} \sup_{x \in \mathcal{X}_0} \left| \hat{\xi}_{n,B}(B^\top x) - \xi_B(B^\top x) \right| > \varepsilon/2 \right] \\ &\quad + \mathbb{I} \left\{ \sup_{B: \|B - B_0\| < \eta} \sup_{x \in \mathcal{X}_0} \left| \xi_B(B^\top x) - \xi_{B_0}(B_0^\top x) \right| > \varepsilon/2 \right\}. \end{aligned}$$

The first term converges to 0 according to (C.5) and the second term is equal to 0 since from (C.4),

$$\sup_{B: \|B - B_0\| < \eta} \sup_{x \in \mathcal{X}_0} \left| \xi_B(B^\top x) - \xi_{B_0}(B_0^\top x) \right| \leq \bar{M}\eta < \varepsilon/2.$$

The proof is then complete. \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 $C_{d+1} \in \mathcal{B}_{d+1}$ be a matrix satisfying the condition of Proposition 2. Since $\mathcal{A}_x(C_{d+1}) \subset \mathcal{A}_x(B_{0,d})$, for all $x \in \mathcal{X}$, we have $\xi_{C_{d+1}}(C_{d+1}^\top x) \leq \xi_{B_{0,d}}(B_{0,d}^\top x)$ and thus

$$\begin{aligned} \min_{B \in \mathcal{B}_{d+1}} \Psi(B, \mathcal{X}_0) &= \min_{B \in \mathcal{B}_{d+1}} \mathbb{E} \left[\xi_B(B^\top X) \mathbb{I}\{X \in \mathcal{X}_0\} \right] \leq \mathbb{E} \left[\xi_{C_{d+1}}(C_{d+1}^\top X) \mathbb{I}\{X \in \mathcal{X}_0\} \right] \\ &\leq \mathbb{E} \left[\xi_{B_{0,d}}(B_{0,d}^\top X) \mathbb{I}\{X \in \mathcal{X}_0\} \right] = \min_{B \in \mathcal{B}_d} \Psi(B, \mathcal{X}_0), \end{aligned} \quad (9)$$

for all $d \in \{1, \dots, p\}$. In particular, the assumption on the matrix C_{d+1} ensures that, for all $d \in \{1, \dots, q-1\}$,

$$\min_{B \in \mathcal{B}_{d+1}} \Psi(B, \mathcal{X}_0) < \min_{B \in \mathcal{B}_d} \Psi(B, \mathcal{X}_0).$$

Moreover, 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 $d \in \{1, \dots, p\}$,

$$\mathbb{E}[\gamma(X)\mathbb{I}\{X \in \mathcal{X}_0\}] \leq \min_{B \in \mathcal{B}_d} \Psi(B, \mathcal{X}_0). \quad (10)$$

From (9), we also have that for all $d \geq q$,

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

Gathering (10) and (11) entails that for all $d \geq q$,

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

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] Halton, J. H. (1964). Algorithm 247: Radical-inverse quasi-random point sequence, *Communications of the ACM*, **7(12)**, 701–702.
- [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.