



HAL
open science

Extreme Partial Least-Squares regression

Meryem Bousebata, Geoffroy Enjolras, Stéphane Girard

► **To cite this version:**

Meryem Bousebata, Geoffroy Enjolras, Stéphane Girard. Extreme Partial Least-Squares regression. 2021. hal-03165399v1

HAL Id: hal-03165399

<https://inria.hal.science/hal-03165399v1>

Preprint submitted on 10 Mar 2021 (v1), last revised 5 Sep 2022 (v3)

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.

Extreme Partial Least-Squares regression

Meryem Bousebata^{1,2}, Geoffroy Enjolras² & Stéphane Girard^{1,*}

¹ Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

² Univ. Grenoble Alpes, CERAG, 38000 Grenoble, France.

* Corresponding author, stephane.girard@inria.fr

Abstract

We propose a new approach, called Extreme-PLS, for dimension reduction in regression and adapted to distribution tails. The objective is to find linear combinations of predictors that best explain the extreme values of the response variable in a non-linear inverse regression model. The asymptotic normality of the Extreme-PLS estimator is established in the single-index framework and under mild assumptions. The performance of the method is assessed on simulated data. A statistical analysis of French farm income data, considering extreme cereal yields, is provided as an illustration.

Keywords. Extreme-value analysis, Dimension reduction, Non-linear inverse regression, Partial Least Squares.

MSC 2020 subject classification: 62G32, 62H25, 62H12, 62E20.

1 Introduction

Regression analysis is widely used to study the relationship between a response variable Y and a p -dimensional vector X of covariates starting from a n -sample. When p grows, a dimension reduction becomes necessary to overcome the well-known "curse of dimensionality" and to reveal the most relevant directions of the high-dimensional covariate space. There exist a number of statistical approaches to dimension reduction in regression problems. One of the most popular is Partial least squares (PLS) regression, introduced by [53], that combines characteristics of Principal component analysis (PCA) for dimension reduction and multiple regression. The development of PLS regression has been initiated within the chemometrics field, see the reference book [39]. Since then, PLS has also received attention in the statistical literature. For example, [29] discusses the statistical properties of the PLS procedure under a factor analysis model, while [20] provides a comparison between PLS and Principal component regression (PCR) from various perspectives. See also [13] for a connection between PLS regression and envelopes and [7]

for a sparse version of PLS. The basic idea of PLS regression is to seek directions, *i.e.* linear combinations of X coordinates having both high variance and high correlation with Y , unlike PCR method which only takes into account high variance components [20, 47]. Sliced inverse regression (SIR) [37] is an alternative method which takes profit of the simplicity of the inverse regression view of X against Y . It aims at replacing X by its projection onto a subspace of smaller dimension without loss of regression information. Many extensions of SIR have been proposed such as Partial inverse regression handling the $p > n$ situation [38], Kernel sliced inverse regression allowing the estimation of a nonlinear subspace [54], Student sliced inverse regression dealing with non-Gaussian and heavy-tailed errors [10], Sliced inverse regression for multivariate response [14], among others. Single-index models provide additional practical tools to overcome the curse of dimensionality, by modelling the non-linear relationship between Y and X through an unknown link function and a single linear combination of the covariates referred to as the index, see [32, Chapter 2]. As such, they provide a reasonable compromise between non-parametric and parametric approaches. Among the numerous works dedicated to the estimation of the index and the link function, the most popular are the average derivative estimation method in the context of kernel smoothing [28, 43], and the M-estimation technique based on spline regression [52, 56]. One can also mention [35, 55] who considered single-index models for the estimation of conditional quantiles. Shrinkage methods aim at reducing the dimension of the covariate by variable selection. As an example, Lasso method [48] penalizes regression coefficients similarly to ridge regression [31] but replacing the L_2 penalization by the L_1 counterpart. Some extensions of this idea have been proposed and include for instance Fused lasso [49] and Elastic net [57] to deal with the case where p is larger than n . Many other shrinkage and variable selection methods are discussed in [21, Chapter 3].

Another way to investigate the relationship between Y and X is to focus on conditional extremes. The goal is then to describe how tail characteristics such as extreme quantiles of Y may depend on the explanatory vector X . Estimation of conditional extreme quantiles has been studied from various points of view. One first approach relies on the fit of a parametric model, see for instance [9] who deals with extreme quantiles in the linear regression model and derives their asymptotic behaviour under several error distributions. Other parametric models are proposed in [8, 18, 45, 50] using extreme-value techniques to model exceedances above a high threshold. The second line of work relies on non-parametric approaches that can be split into three main categories, namely, fixed design, random design and functional covariates. Fixed design methods aim at estimating conditional extreme quantiles depending on a non-random covariate, see [23] for a nearest neighbors technique. In the random design setting, one can cite [15, 25] who studied the estimation of extreme quantiles under a conditional heavy-tail model, later extended in [16] to conditional distributions belonging to any maximum domain of

attraction. Finally, see [24] for the functional covariate situation. Semi-parametric approaches have also been considered for trend modeling in extreme values. Local polynomial fitting of extreme-value models is investigated in [2, 17], a non-parametric estimation of the temporal trend is combined with parametric models for extreme values in [27] and a location-dispersion regression model for heavy-tailed distributions is introduced in [1].

There is a growing interest in many applications for a combination of these two main lines, namely dimension reduction in the regression context and conditional extremes. One motivating example is the study of the influence of meteorological parameters (temperature, humidity,...), agricultural inputs (pesticides, fertilisers,...) and risk management (insurance premiums, subsidies,...) on the low values of farm incomes (see Section 6). Other applications may be found in finance [40], climatology [34], hydrology [23] and environment [45], to name a few. Despite this potential interest, there is a limited literature on the combination of these two lines of work. In [51] a semi-parametric approach is introduced for the estimation of extreme quantiles basing on a tail single-index model but requiring a linearity assumption on the conditional mean. One can also mention [22] where a dimension reduction method dedicated to conditional tail distributions is developed assuming that, when Y is large, X and Y are independent conditionally on a combination of the covariates. Finally, another approach [19] consists in adapting the (unsupervised) dimension reduction method PCA to an extreme setting.

Here, we aim at looking for linear combinations of covariates that best explain the extreme values of Y . More precisely, we first propose a single-index approach to find a direction $\hat{\beta}$ maximizing the covariance between $\beta^t X$ and Y given Y exceeds a high threshold y . An iterative procedure is then exhibited to adapt the method to the multiple-index situation. This adaptation of the PLS estimator to the extreme-value framework, referred to as extreme-PLS (EPLS), is achieved in the context of a non-linear inverse regression model. In practice, $\hat{\beta}$ allows to quantify the effect of the covariates on the extreme values of Y in a simple and interpretable way. Plotting Y against the projection $\hat{\beta}^t X$ provides a visual interpretation of conditional extremes. Moreover, working on the pair $(\hat{\beta}^t X, Y)$ should yield improved results for most estimators dealing with conditional extreme values thanks to the dimension reduction achieved in the projection step. From the theoretical point of view, the asymptotic normality of $\hat{\beta}$ is established without linearity nor independence requirements, when $n \rightarrow \infty$ and p is fixed.

Similarly, [41] showed that (non-extreme) PLS provides a consistent estimator of the direction when Y given X follows a single-index model and $n \rightarrow \infty$ with a fixed dimension p , under the additional assumption of independence between noise and covariates. This result is extended in [7] to the multiple-index situation and when both $n \rightarrow \infty$ and $p \rightarrow \infty$ with $p/n \rightarrow 0$. It is also shown that, at the opposite, the PLS estimator is no more consistent in the case where $p/n \rightarrow k > 0$, and a sparse version of PLS is introduced to circumvent this vexing effect. More

recently, [12] tempered this conclusion by exhibiting some designs under which single-index PLS is consistent in the linear regression situation, when both $n \rightarrow \infty$ and $p \rightarrow \infty$, regardless of the alignment between n and p . These results support the conjecture that PLS is a well-founded dimension reduction method.

The paper is organized as follows. In Section 2, the EPLS approach is introduced in the framework of a single-index model and heavy-tailed distributions. Some preliminary properties are stated in order to justify the above heuristics from a theoretical point of view. The associated estimator is exhibited in Section 3 and its asymptotic distribution is established under mild assumptions. This approach is extended to the multiple-index setting in Section 4. The performances of the method are investigated through a simulation study in Section 5. EPLS approach is then applied in Section 6 to assess the influence of various parameters on cereal yields collected on French farms. A small discussion is provided in Section 7 and proofs are postponed to the Appendix.

2 Single-index EPLS approach

Let us first consider the following single-index non linear inverse regression model:

(\mathbf{M}_1) $X = g(Y)\beta + \varepsilon$, where X and ε are p -dimensional random vectors, Y is a real random variable, $g : \mathbb{R} \rightarrow \mathbb{R}$ is an unknown link function, $\beta \in \mathbb{R}^p$ is an unknown unit vector.

Here, Y is the response variable, X is the multidimensional covariate and ε is a multidimensional error term. Let us highlight that no independence assumption is made on the pair (X, ε) . Model (\mathbf{M}_1) is referred to as an inverse regression model since the covariates are written as functions of the response variable. Similar models were used to establish the theoretical properties of SIR, see for instance [4, 11]. Under model (\mathbf{M}_1), we aim at estimating β by maximizing the covariance between $\beta^t X$ and Y conditionally on large values of Y . Indeed, roughly speaking, when Y is large, provided the distribution tail of ε is negligible, one has $X \simeq g(Y)\beta$ leading to the approximate single-index forward model $Y \simeq g^{-1}(\beta^t X)$. Note that the considered model does not require a linear conditional expectation as in [51] nor a conditional independence assumption as in [22, 44].

Let us denote by $w(y)$ the unit vector maximizing the covariance between $w^t X$ and Y given that Y exceeds a large threshold y :

$$w(y) = \arg \max_{\|w\|=1} \text{cov}(w^t X, Y | Y \geq y). \quad (1)$$

This linear optimization problem under a quadratic constraint benefits from a closed-form solution obtained with Lagrange multipliers method and given in the next Proposition. To this

end, for all $y \in \mathbb{R}$, denote by $\bar{F}(y) = \mathbb{P}(Y \geq y)$ the survival function of Y and consider the three tail-moments, whenever they exist, $m_Y(y) = \mathbb{E}(Y\mathbf{1}_{\{Y \geq y\}}) \in \mathbb{R}$, $m_X(y) = \mathbb{E}(X\mathbf{1}_{\{Y \geq y\}}) \in \mathbb{R}^p$, $m_{XY}(y) = \mathbb{E}(XY\mathbf{1}_{\{Y \geq y\}}) \in \mathbb{R}^p$.

Proposition 1. *Suppose that $\mathbb{E}\|X\| < \infty$, $\mathbb{E}|Y| < \infty$ and $\mathbb{E}\|XY\| < \infty$. Then, the unique solution of the optimization problem (1) is given for all $y \in \mathbb{R}$ by:*

$$w(y) = v(y)/\|v(y)\| \text{ where } v(y) = \bar{F}(y)m_{XY}(y) - m_X(y)m_Y(y). \quad (2)$$

Let us note that solution (2) is invariant with respect to the scaling and location of X . Besides, when ε is centered and independent of Y , we recover the classical PLS framework and it is easily shown that $w(y) = \pm\beta$ for all $y \in \mathbb{R}$. In the following, no assumption is made on the (in)dependence between Y and ε , but additional assumptions on the link function g and the distribution tail of Y are considered:

(A₁) Y is a real random variable with density function f regularly varying at infinity with index $-1/\gamma - 1$, $\gamma \in (0, 1)$ *i.e.* for all $t > 0$,

$$\lim_{y \rightarrow \infty} \frac{f(ty)}{f(y)} = t^{-\frac{1}{\gamma}-1}.$$

This property is denoted for short by $f \in RV_{-1/\gamma-1}$.

(A₂) $g \in RV_c$ with $c > 0$.

(A₃) There exists $q > 1/(\gamma c)$ such that $\mathbb{E}(\|\varepsilon\|^q) < \infty$.

Let us note that **(A₁)** implies that $\bar{F} \in RV_{-1/\gamma}$ in view of Karamata's theorem [6, Theorem 1.5.8]. In other words, **(A₁)** entails that Y has a right heavy-tail. This is equivalent to assuming that the distribution of Y is in the Fréchet maximum domain of attraction, with tail-index $\gamma > 0$, see [26, Theorem 1.2.1]. This domain of attraction includes for instance Pareto, Burr and Student distributions, see [3] for further examples of heavy-tailed distributions. The restriction to $\gamma < 1$ ensures that $\mathbb{E}|Y|$ exists. Assumption **(A₂)** means that the link function ultimately behaves like a power function. Finally, **(A₃)** can be interpreted as an assumption on the tails of $\|\varepsilon\|$. It is satisfied for instance by distributions with exponential-like tails such as Gaussian, Gamma or Weibull distributions. More specifically, $\mathbb{E}(\|\varepsilon\|^q) < \infty$ implies that the tail-index, say $\gamma_{\|\varepsilon\|}$, associated with $\|\varepsilon\|$ is such that $\gamma_{\|\varepsilon\|} < 1/q$. Besides, it can readily be shown that $g(Y)$ is heavy-tailed with tail-index $\gamma_{g(Y)} := c\gamma$. Condition **(A₃)** thus imposes that $\gamma_{g(Y)} > \gamma_{\|\varepsilon\|}$, meaning that $g(Y)$ has an heavier right tail than ε . In model **(M₁)**, the tail behavior of $|\beta^t X|$ and $\|X\|$ is thus driven by $g(Y)$ rather than by $\beta^t \varepsilon$, *i.e.* $\gamma_{\|X\|} = \gamma_{g(Y)}$, which is the desired property.

In order to assess the convergence of $w(y)$ to β as $y \rightarrow \infty$, we let

$$\Delta(w(y), \beta) := 1 - \cos^2(w(y), \beta) = 1 - (w(y)^t \beta)^2.$$

A value close to 1 implies a low colinearity ($w(y)$ is almost orthogonal to β) while a value close to 0 means a high colinearity.

Proposition 2. *Assume (\mathbf{M}_1) , (\mathbf{A}_1) , (\mathbf{A}_2) and (\mathbf{A}_3) hold with $\gamma(c+1) < 1$. Then,*

$$\Delta(w(y), \beta) = O \left\{ \left(\frac{1}{g(y)\bar{F}^{1/q}(y)} \right)^2 \right\} \rightarrow 0 \quad \text{and} \quad \|w(y) - \beta\| = O \left(\frac{1}{g(y)\bar{F}^{1/q}(y)} \right) \rightarrow 0,$$

as $y \rightarrow \infty$.

It should be noted that, since $\|w(y) - \beta\| \rightarrow 0$ as $y \rightarrow \infty$, the EPLS axis has asymptotically the same direction as β (without sign issue). Besides, in view of assumptions (\mathbf{A}_1) and (\mathbf{A}_2) , the function $y \mapsto g(y)\bar{F}^{1/q}(y)$ is regularly varying with index $c - 1/(q\gamma) > 0$ from (\mathbf{A}_3) . Unsurprisingly, the above convergence rates are large when c is large (*i.e.* the link function is rapidly increasing), q is large (*i.e.* the noise ε is small) or/and γ is large (*i.e.* the tail of Y is heavy). The estimation of $w(y)$ from data distributed from model (\mathbf{M}_1) is addressed in the following section.

3 Single-index EPLS: Estimators and main properties

Let (X_i, Y_i) , $1 \leq i \leq n$ be independent and identically distributed random variables from model (\mathbf{M}_1) and let $y_n \rightarrow \infty$ as the sample size n tends to infinity. The solution (2) is estimated by its empirical counterpart introducing

$$\hat{v}(y_n) = \hat{\bar{F}}(y_n)\hat{m}_{XY}(y_n) - \hat{m}_X(y_n)\hat{m}_Y(y_n), \quad (3)$$

with $\hat{\bar{F}}$ the empirical survival function and

$$\hat{m}_{XY}(y_n) = \frac{1}{n} \sum_{i=1}^n X_i Y_i \mathbf{1}_{\{Y_i \geq y_n\}}, \quad \hat{m}_Y(y_n) = \frac{1}{n} \sum_{i=1}^n Y_i \mathbf{1}_{\{Y_i \geq y_n\}}, \quad \hat{m}_X(y_n) = \frac{1}{n} \sum_{i=1}^n X_i \mathbf{1}_{\{Y_i \geq y_n\}}.$$

For all $j = 1, \dots, p$, let us denote by $X_{\cdot, j}$ the j th coordinate of X . It readily follows that the j th coordinates of $m_X(y_n)$ and $m_{XY}(y_n)$ are respectively given by $m_{X_{\cdot, j}}(y_n)$ and $m_{X_{\cdot, j}Y}(y_n)$.

We first provide a tool establishing the joint asymptotic normality of $\hat{\bar{F}}(y_n)$, $\hat{m}_{XY}(y_n)$, $\hat{m}_Y(y_n)$ and $\hat{m}_X(y_n)$ when $y_n \rightarrow \infty$ and $n\bar{F}(y_n) \rightarrow \infty$. This latter condition ensures that the rate of convergence, which is driven by the number of effective observations $n\bar{F}(y_n)$ used in the estimators, tends to infinity as the sample size increases.

Proposition 3. Assume (\mathbf{M}_1) , (\mathbf{A}_1) , (\mathbf{A}_2) and (\mathbf{A}_3) hold with $2\gamma(c+1) < 1$. Let us denote by d the number of non-zero β_j coefficients in $\beta \in \mathbb{R}^p$, and assume for the sake of simplicity that $\beta_j \neq 0$ for all $j = 1, \dots, d$ and $\beta_{d+1} = \dots = \beta_p = 0$. Let $y_n \rightarrow \infty$ such that $n\bar{F}(y_n) \rightarrow \infty$ and introduce the $\mathbb{R}^{2(d+1)}$ - random vector

$$\Lambda_n := \left\{ \left(\frac{\hat{F}(y_n)}{\bar{F}(y_n)} - 1 \right), \left(\frac{\hat{m}_Y(y_n)}{m_Y(y_n)} - 1 \right), \left(\frac{\hat{m}_{X.,j}(y_n)}{m_{X.,j}(y_n)} - 1 \right)_{1 \leq j \leq d}, \left(\frac{\hat{m}_{X.,jY}(y_n)}{m_{X.,jY}(y_n)} - 1 \right)_{1 \leq j \leq d} \right\}.$$

Then,

$$\sqrt{n\bar{F}(y_n)}\Lambda_n \xrightarrow{d} \mathcal{N}(0, B),$$

where B is the $2(d+1) \times 2(d+1)$ covariance matrix defined by

$$B = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 1 & b_{22} & b_{23} & \dots & b_{23} & b_{24} & \dots & b_{24} \\ 1 & b_{23} & b_{33} & \dots & b_{33} & b_{34} & \dots & b_{34} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & b_{23} & b_{33} & \dots & b_{33} & b_{34} & \dots & b_{34} \\ 1 & b_{24} & b_{34} & \dots & b_{34} & b_{44} & \dots & b_{44} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & b_{24} & b_{34} & \dots & b_{34} & b_{44} & \dots & b_{44} \end{pmatrix}$$

and where

$$\begin{aligned} b_{22} &= \frac{(1-\gamma)^2}{1-2\gamma}, & b_{23} &= \frac{(1-\gamma)(1-\gamma c)}{1-\gamma(c+1)}, \\ b_{33} &= \frac{(1-\gamma c)^2}{1-2\gamma c}, & b_{24} &= \frac{(1-\gamma(c+1))(1-\gamma)}{1-\gamma(c+2)}, \\ b_{44} &= \frac{(1-\gamma(c+1))^2}{1-2\gamma(c+1)}, & b_{34} &= \frac{(1-\gamma c)(1-\gamma(c+1))}{1-\gamma(2c+1)}. \end{aligned}$$

Let us remark that the above result only provides the (joint) asymptotic distribution of $\hat{m}_{XY}(y_n)$ and $\hat{m}_X(y_n)$ in the directions associated with non-zero β_j . It is however sufficient to establish the asymptotic normality of $\hat{v}(y_n)$ centered on $v(y_n)$, the direction provided by the EPLS criterion, see Proposition 1.

Proposition 4. Assume (\mathbf{M}_1) , (\mathbf{A}_1) , (\mathbf{A}_2) and (\mathbf{A}_3) hold with $2\gamma(c+1) < 1$. Let $y_n \rightarrow \infty$ such that $n\bar{F}(y_n) \rightarrow \infty$. Then,

$$\sqrt{n\bar{F}(y_n)} \left(\frac{\hat{v}(y_n) - v(y_n)}{\|v(y_n)\|} \right) \xrightarrow{d} \xi\beta,$$

where $\xi \sim \mathcal{N}(0, \lambda(c, \gamma))$ and

$$\lambda(c, \gamma) = a_1^2(3 + b_{44}) + a_2^2(2b_{23} + b_{22} + b_{33}) - 2a_1a_2(2 + b_{24} + b_{34}), \quad (4)$$

with $a_1 = (1 - \gamma)(1 - \gamma c)/(\gamma^2 c)$ and $a_2 = (1 - \gamma(c + 1))/(\gamma^2 c)$.

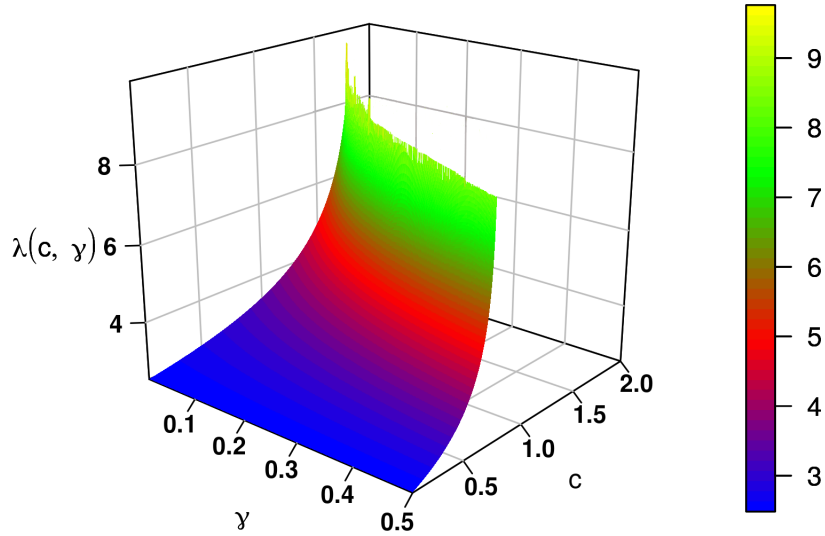


Figure 1: Graph of the asymptotic variance $(c, \gamma) \in [1/2, 2] \times [0, 1/2] \mapsto \lambda(c, \gamma)$ given in Proposition 4, Equation (4), on a logarithmic scale.

The asymptotic variance $\lambda(c, \gamma)$ is plotted on Figure 1 as a function of $(c, \gamma) \in [1/2, 2] \times [0, 1/2]$ and under the constraint $2\gamma(c + 1) < 1$. This condition imposes an upper bound on the tail-index of $\|X\|$: $\gamma_{\|X\|} < c/(2(c + 1))$, see Section 2. Similarly, asymptotic properties of usual dimension-reduction methods are established under the assumption that $\mathbb{E}(\|X\|^4) < \infty$ which implies $\gamma_{\|X\|} < 1/4$, see for instance [33, 44] in the SIR case. The latter bound is the strongest one when $c > 1$.

It appears from Proposition 4 that the asymptotic distribution of $\hat{v}(y_n)$ is Gaussian and degenerated in every direction orthogonal to β . Combining the above result with Proposition 2 provides an asymptotic normality result for $\hat{v}(y_n)$ centered on the true direction β .

Theorem 1. *Assume (\mathbf{M}_1) , (\mathbf{A}_1) , (\mathbf{A}_2) and (\mathbf{A}_3) hold with $2\gamma(c + 1) < 1$. Let $y_n \rightarrow \infty$ such that $n\bar{F}(y_n) \rightarrow \infty$ and $n\bar{F}(y_n)^{1-2/q}/g^2(y_n) \rightarrow 0$ as $n \rightarrow \infty$. Then,*

$$\sqrt{n\bar{F}(y_n)} \left(\frac{\hat{v}(y_n)}{\|\hat{v}(y_n)\|} - \beta \right) \xrightarrow{d} \xi\beta,$$

with $\xi \sim \mathcal{N}(0, \lambda(c, \gamma))$ and where $\lambda(c, \gamma)$ is defined in (4).

Assumption $n\bar{F}(y_n) \rightarrow \infty$ ensures that the variance of the estimator tends to zero while condition $n\bar{F}(y_n)^{1-2/q}/g^2(y_n) \rightarrow 0$ entails that the bias (bounded above by $1/(g(y_n)\bar{F}^{1/q}(y_n))$, see Proposition 2) is asymptotically small compared to the standard deviation $1/\sqrt{n\bar{F}(y_n)}$. Choosing $y_n = \bar{F}^{-1}(\tau_n)$ with $\tau_n = n^{-\nu}$, these conditions are fulfilled provided that $\nu \in \left(\frac{q}{(2\gamma c+1)q-2}, 1\right)$ since both functions g and \bar{F} are assumed to be regularly-varying. Let us also highlight that the above interval is not empty under (\mathbf{A}_3) . Finally, Theorem 1 shows that the estimated direction $\hat{v}(y_n)$ is asymptotically aligned with the true direction β .

4 Extension to several directions

The single-index model (\mathbf{M}_1) can be extended to a multi-index setting by considering, for some $K \geq 1$:

(\mathbf{M}_K) $X = \sum_{\ell=1}^K g_\ell(Y)\beta^{(\ell)} + \varepsilon$, where X and ε are p -dimensional random vectors, Y is a real random variable, $g_\ell : \mathbb{R} \rightarrow \mathbb{R}$ are unknown link functions, $\beta^{(\ell)} \in \mathbb{R}^p$ are unknown orthogonal unit vectors.

For a known value of K , the following iterative procedure is considered to estimate the directions $\beta^{(1)}, \dots, \beta^{(K)}$:

1. Initialization: Set $R_i^{(0)} := X_i$ for all $i = 1, \dots, n$.

2. For $\ell \in \{1, \dots, K\}$,

- Estimation of the ℓ th direction:

$$\hat{\beta}^{(\ell)} = \hat{v}^{(\ell)}(y_n) / \|\hat{v}^{(\ell)}(y_n)\| \text{ where } \hat{v}^{(\ell)}(y_n) = \hat{F}(y_n)\hat{m}_{R^{(\ell-1)}Y}(y_n) - \hat{m}_{R^{(\ell-1)}}(y_n)\hat{m}_Y(y_n).$$

- Computation of the residuals: $R_i^{(\ell)} := R_i^{(\ell-1)} - (\hat{\beta}^{(\ell)})^t R_i^{(\ell-1)} \hat{\beta}^{(\ell)}$ for all $i = 1, \dots, n$.

The idea is the following. At the first iteration, $\hat{\beta}^{(1)}$ corresponds to the estimator computed in the single-index model by (2) and (3). From model (\mathbf{M}_K) , in view of the orthogonality of the directions, one has $(\hat{\beta}^{(1)})^t X \simeq (\beta^{(1)})^t X \simeq g_1(Y)$ and thus the residual

$$R^{(1)} := X - (\hat{\beta}^{(1)})^t X \hat{\beta}^{(1)} \simeq X - g_1(Y)\beta^{(1)} = \sum_{\ell=2}^K g_\ell(Y)\beta^{(\ell)} + \varepsilon$$

approximately satisfies the same inverse regression model with $K - 1$ directions. It is then natural to iterate the process and estimate $\beta^{(2)}$ basing on (2) and (3) computed on the residuals. Moreover, since the residuals are by construction orthogonal to $\hat{\beta}^{(1)}$, one necessarily has the orthogonality property $\hat{\beta}^{(2)} \perp \hat{\beta}^{(1)}$. The estimation of the number of directions K is not

addressed here. We limit ourselves to proposing a visual criterion in the real data experiment (Section 6). Let us however highlight that, thanks to the above orthogonality property, the estimated number of directions is necessarily smaller than p . We refer to [29] for a similar result on the original PLS method.

5 Validation on simulations

Let us consider a sample of size $n = 1000$ and dimension p from model (\mathbf{M}_1) with a power link function $g(t) = t^c$, $t > 0$, $c \in \{1/4, 1/2, 1, 3/2\}$. The behavior of the EPLS estimator $\hat{v}(y_n)$ is illustrated on this inverse regression model and compared to the estimator introduced in [51] referred to as SIMEXQ (single-index model extreme quantile) in the sequel. SIMEXQ method estimates the direction of interest by fitting a linear quantile regression model to the exceedances. Two heavy-tailed distributions (with common tail-index $\gamma = 1/5$) are selected for the response variable Y :

- a Pareto distribution with survival function $\bar{F}(y) = (y/2)^{-5}$, $y \geq 2$,
- a Student t_5 distribution with 5 degrees of freedom.

Two dimensions of the covariate are considered:

- $p = 3$ with $\beta = (1, 1, 0)^t/\sqrt{2}$,
- $p = 30$ with $\beta = (1, \dots, 1, 0, \dots, 0)^t/\sqrt{15}$.

Each component $\varepsilon^{(j)}$, $j = 1, \dots, p$ of the error ε is simulated from the $\mathcal{N}(0, \sigma^2)$ distribution and depending on Y using Frank copula defined for all $(u, v) \in [0, 1]^2$ by

$$C_\theta(u, v) = -\frac{1}{\theta} \log \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right),$$

where $\theta \in \mathbb{R}$ is a parameter tuning the dependence between the margins. Frank copula is an Archimedean copula, see for instance [42, Section 4.2], able to model the full range of dependence: $\theta \rightarrow -\infty$ yields the counter-monotonicity copula, $\theta \rightarrow +\infty$ yields the co-monotonicity copula while $\theta = 0$ corresponds to independence. Here, we choose $\theta \in \{0, 10, 20\}$ corresponding to Kendall's $\tau \in \{0, 0.67, 0.82\}$. The standard deviation σ is selected such that the Signal to Noise Ratio (SNR) defined as $\text{SNR} := g(\bar{F}^{-1}(1/n))/\sigma$ is equal to 10. Note that $g(\bar{F}^{-1}(1/n))$ represents the approximate maximum value of g on a n -sample from the distribution with associated survival function \bar{F} . Finally, the mean proximity criterion between $\hat{v}(y_n)$ and β is computed on $N = 100$ replications as follows:

$$\text{PC}(y_n) = \frac{1}{N} \sum_{r=1}^N \cos^2(\hat{v}(y_n)^{(r)}, \beta) = 1 - \frac{1}{N} \sum_{r=1}^N \Delta(\hat{v}(y_n)^{(r)}, \beta), \quad (5)$$

where $\hat{v}(y_n)^{(r)}$ denotes the estimator (3) computed on the r^{th} replication. The closer $\text{PC}(y_n)$ is to 1, the better the estimator is. The performance of EPLS and SIMEXQ methods are compared by computing (5) in the $4 \times 2 \times 2 \times 3 = 48$ considered situations. To this end, denoting by $Y_{1,n} \leq Y_{2,n} \leq \dots \leq Y_{n,n}$ the order statistics of the sample (Y_1, \dots, Y_n) , the quality measure $\text{PC}(Y_{n-k+1,n})$ is plotted in Figures 2–5 as a function of the number of exceedances $k \in \{1, \dots, 200\}$.

It first appears that the performance of the EPLS estimator does not depend on the distribution of the response variable. Besides, in small dimension ($p = 3$), the EPLS method yields very accurate results (with $\text{PC} \geq 0.8$) for a wide range of choices of k and whatever the exponent c and the dependence coefficient are. In contrast, the SIMEXQ method appears to be very sensitive to the distribution of Y and to the dependence strength. In this small dimension situation, EPLS outperforms SIMEXQ as soon as independence does not hold. In a high dimension setting ($p = 30$), EPLS still provides very good results (with $\text{PC} \geq 0.8$) for a wide range of choices of k when $c \geq 1$ for all dependence situations. Good results ($\text{PC} \geq 0.6$) can also be obtained when $c = 1/2$ for well-chosen values of k . Here again, the SIMEXQ method is not robust to dependence and is outperformed by EPLS.

Finally, it appears that the choice of the number of exceedances k may be a crucial point in difficult situations (high dimension p , high dependence and c small) and is therefore discussed in the next section on a real dataset.

6 Application to farm income modelling

Our approach is applied to data extracted from the Farm Accountancy Data Network¹ (FADN), an annual database of commercial-sized farm holdings. This dataset of $n = 949$ observations contains significant accounting and financial information about French professional farm incomes in 2014. Our goal is to investigate the impact of various factors on farm yields.

The response variable Y is the inverse of the wheat yield (in quintals/hectare), as we focus in the analysis of low yields, and the covariate X includes 12 continuous variables: selling prices (euro/quintal), pesticides, fertilizers, crop insurance purchase, insurance claims, farm subsidies, seeds and seedlings costs, works and services purchase for crops, other insurance premiums, farm income taxes, farmer’s personal social security costs (euro/hectare) and temperature average (degree Celsius). We first carry out, in Figure 6, a number of visual checks of whether the heavy-tailed assumption makes sense for these data. First, the histogram of the Y_i (top left panel) gives descriptive evidence that Y has a heavy right tail. The second step consists in

¹A detailed presentation of the database can be found at: <http://agreste.agriculture.gouv.fr> (in French).

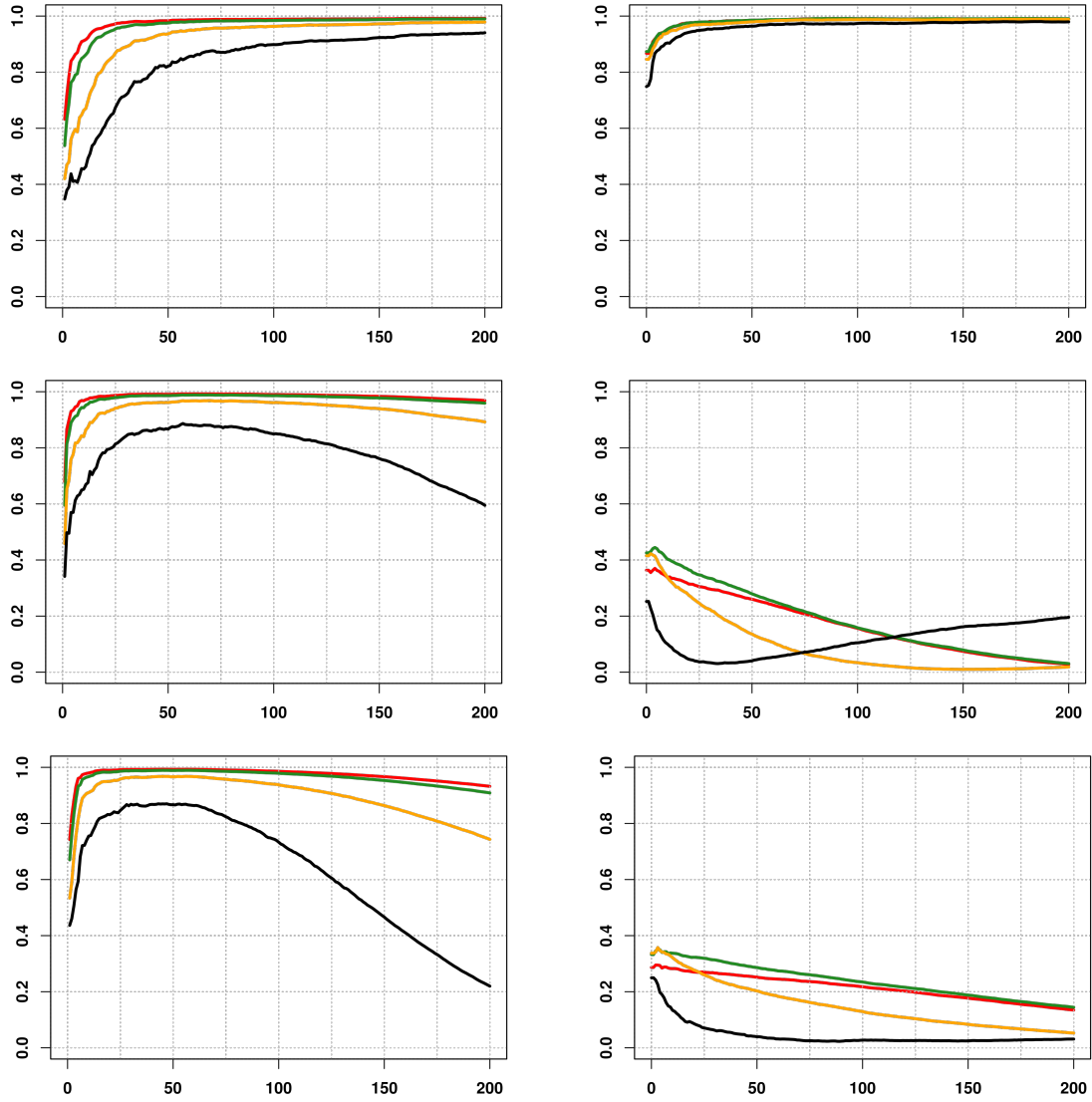


Figure 2: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ [51] (right) estimators, on simulated data from a Pareto distribution, dimension $p = 3$. $PC(Y_{n-k+1,n})$ quality measure as a function of the number $k \in \{1, \dots, 200\}$ of exceedances. From top to bottom, Frank copula parameter $\theta \in \{0, 10, 20\}$. The powers $c \in \{1/4, 1/2, 1, 3/2\}$ of the link function $g(t) = t^c$ are displayed in {black, yellow, green, red}.

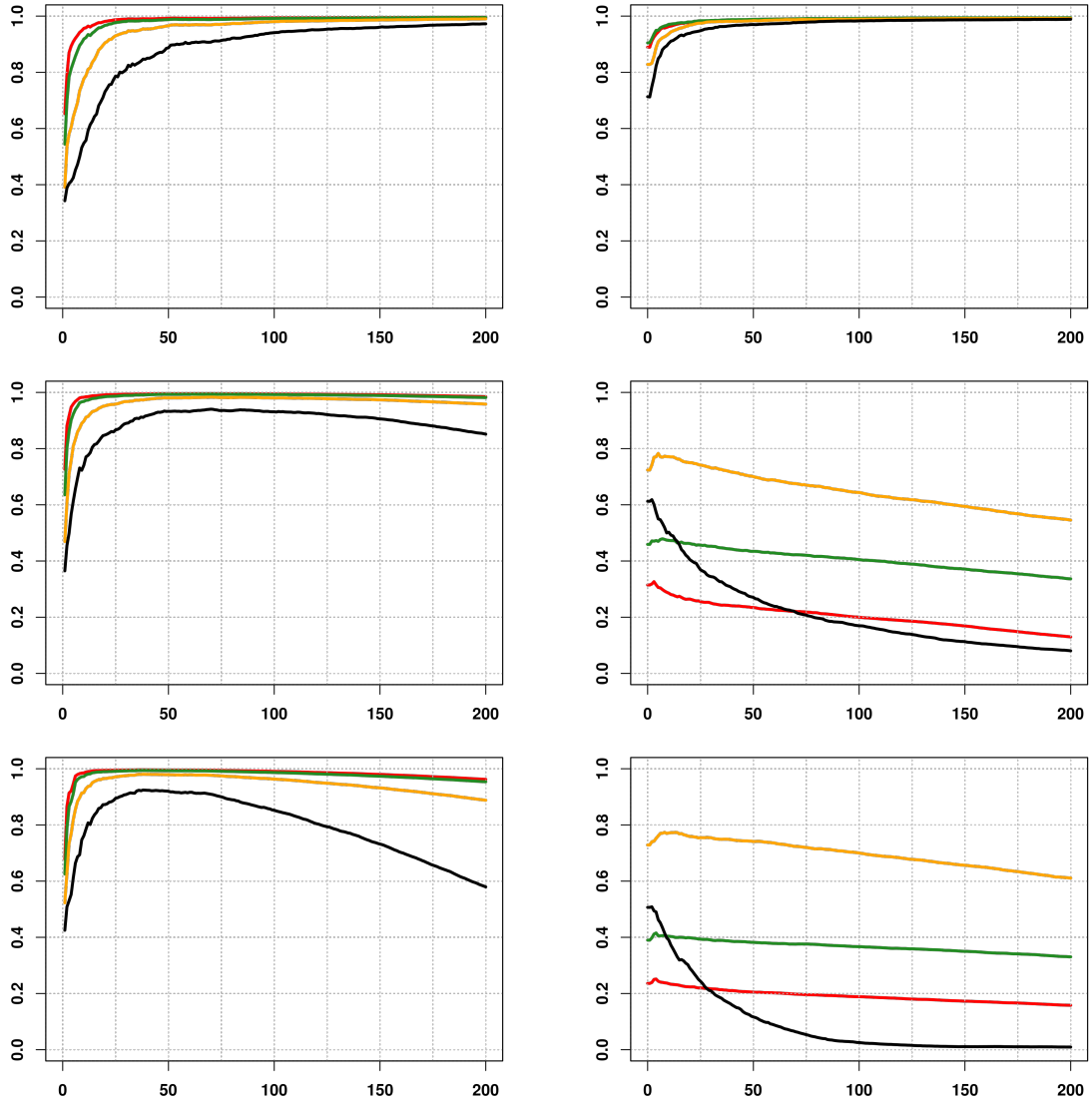


Figure 3: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ [51] (right) estimators, on simulated data from a Student distribution, dimension $p = 3$. $PC(Y_{n-k+1,n})$ quality measure as a function of the number $k \in \{1, \dots, 200\}$ of exceedances. From top to bottom, Frank copula parameter $\theta \in \{0, 10, 20\}$. The powers $c \in \{1/4, 1/2, 1, 3/2\}$ of the link function $g(t) = t^c$ are displayed in {black, yellow, green, red}.

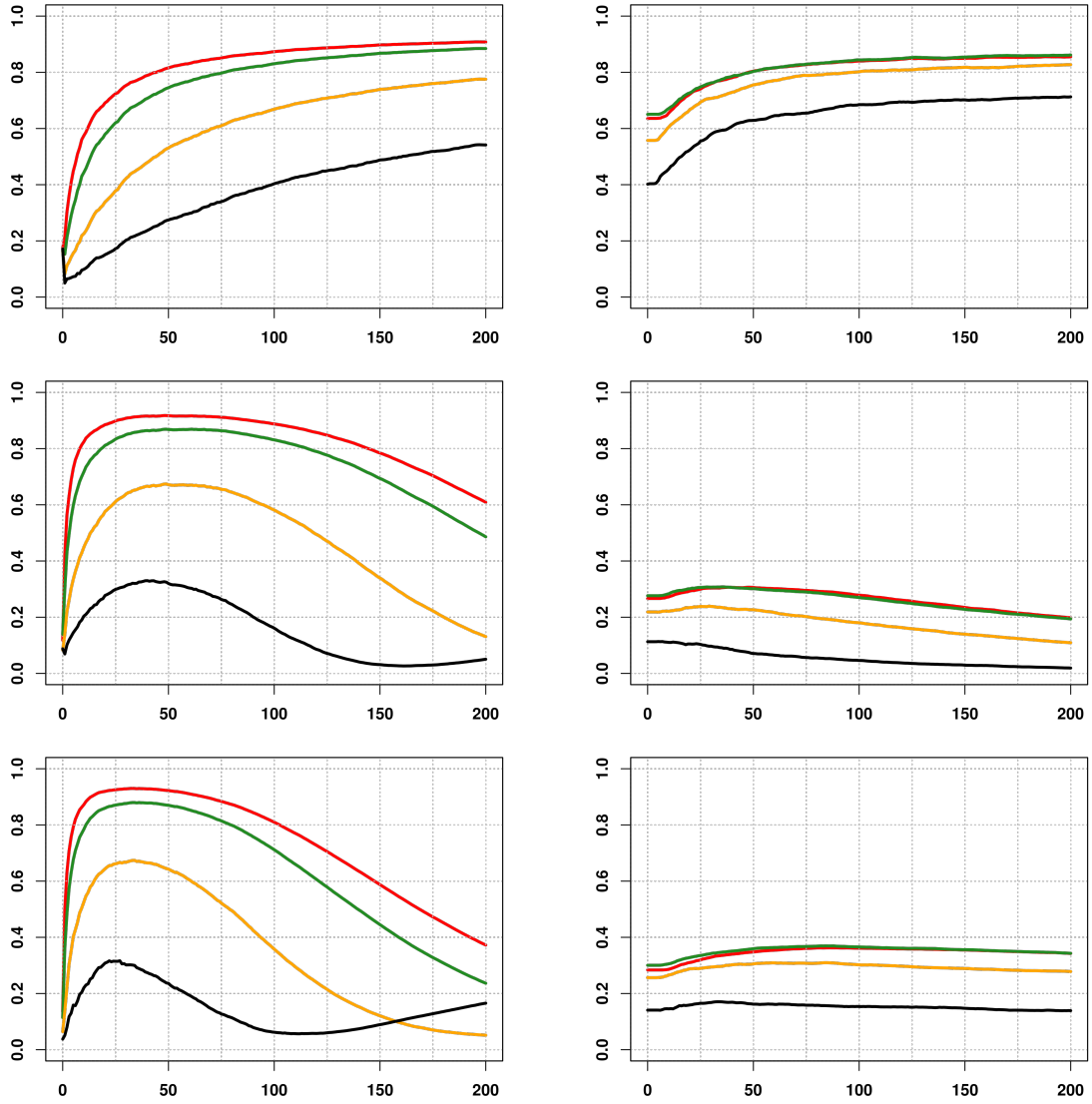


Figure 4: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ [51] (right) estimators, on simulated data from a Pareto distribution, dimension $p = 30$. $PC(Y_{n-k+1,n})$ quality measure as a function of the number $k \in \{1, \dots, 200\}$ of exceedances. From top to bottom, Frank copula parameter $\theta \in \{0, 10, 20\}$. The powers $c \in \{1/4, 1/2, 1, 3/2\}$ of the link function $g(t) = t^c$ are displayed in {black, yellow, green, red}.

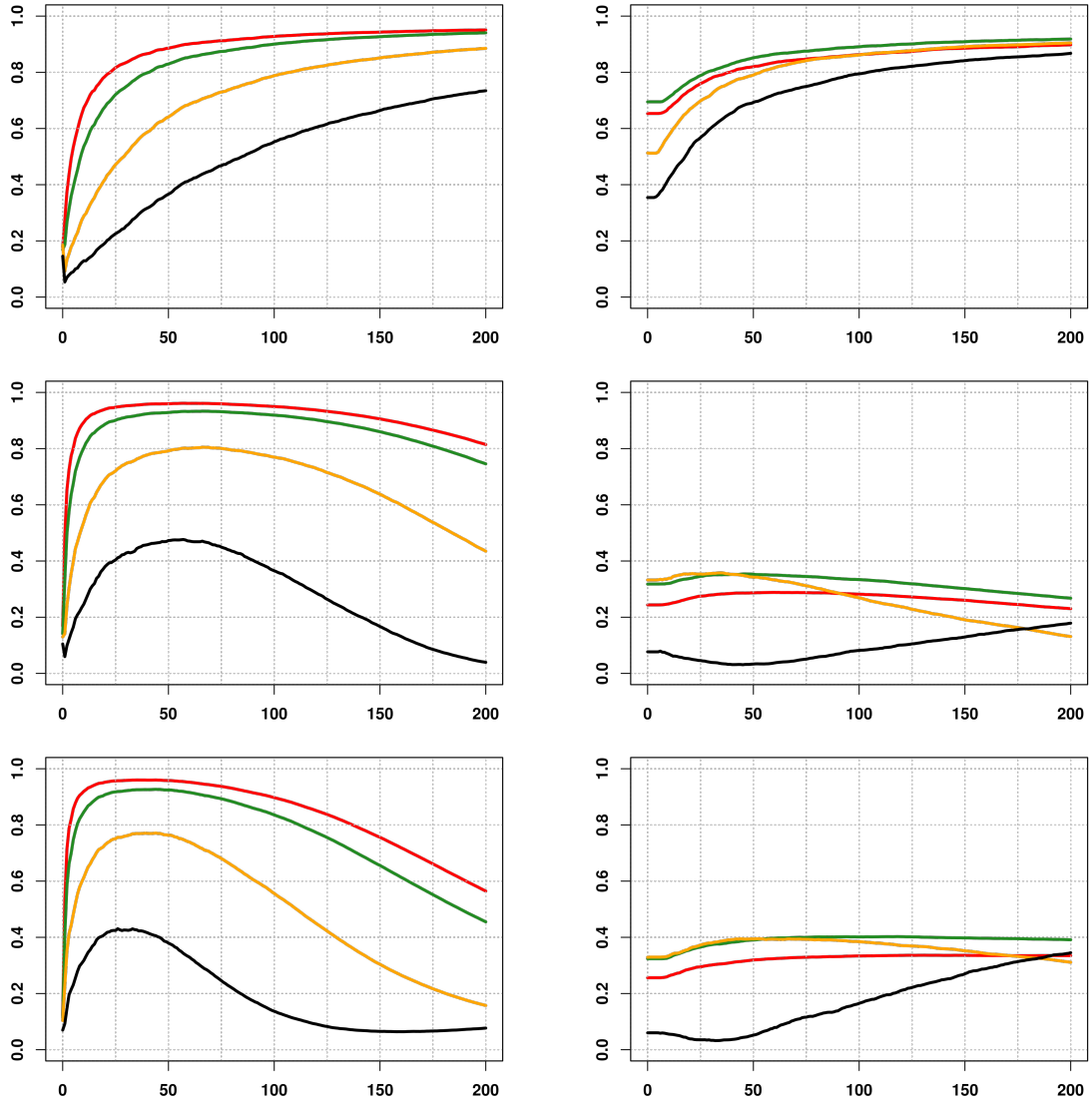


Figure 5: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ [51] (right) estimators, on simulated data from a Student distribution, dimension $p = 30$. $PC(Y_{n-k+1,n})$ quality measure as a function of the number $k \in \{1, \dots, 200\}$ of exceedances. From top to bottom, Frank copula parameter $\theta \in \{0, 10, 20\}$. The powers $c \in \{1/4, 1/2, 1, 3/2\}$ of the link function $g(t) = t^c$ are displayed in {black, yellow, green, red}.

drawing a Hill plot [30] (top right panel):

$$\left(k, \hat{\gamma}_k = \frac{1}{k} \sum_{i=1}^{k-1} \log(Y_{n-i,n}/Y_{n-k,n}) \right), k = 1, \dots, 150.$$

The resulting graph shows a nice stability of the Hill estimator $\hat{\gamma}_k$ of the tail index γ as a function of the sample fraction for $k \in \{50, \dots, 150\}$ pointing towards $\gamma \simeq 0.25$. To further confirm that the heavy-tailed framework is appropriate, we draw a quantile-quantile plot of the log-excesses against the quantiles of the unit exponential distribution. Formally, let $Z_{i,n} = \log(Y_{n-i+1,n}/Y_{n-k,n})$, $1 \leq i \leq k$ denote the log-excesses computed from the consecutive top order statistics. If Y is heavy-tailed with tail-index γ then, for small i , the $Z_{i,n}$ are approximately independent copies of an exponential random variable with mean γ , see for instance [3, pp.109–110]. The bottom panel of Figure 6 displays such a quantile-quantile plot for $k = 150$. The relationship appearing in this plot is approximately linear, which constitutes an empirical evidence that the heavy-tail assumption on Y makes sense. We secondly compute the first two estimators $\hat{v}^{(\ell)}(y_n)$ for $\ell \in \{1, 2\}$ and $y_n = Y_{n-k+1,n}$, $k \in \{1, \dots, 150\}$. For the sake of interpretation, we define the following two conditional correlations as:

$$\rho(X^t \hat{v}^{(\ell)}(y_n), Y | Y \geq y_n) = \frac{\text{cov}(X^t \hat{v}^{(\ell)}(y_n), Y | Y \geq y_n)}{\sigma(X^t \hat{v}^{(\ell)}(y_n) | Y \geq y_n) \sigma(Y | Y \geq y_n)}, \quad (6)$$

$$\rho(X^t \hat{v}^{(\ell)}(y_n), X^{(j)} | Y \geq y_n) = \frac{\text{cov}(X^t \hat{v}^{(\ell)}(y_n), X^{(j)} | Y \geq y_n)}{\sigma(X^t \hat{v}^{(\ell)}(y_n) | Y \geq y_n) \sigma(X^{(j)} | Y \geq y_n)}. \quad (7)$$

The results are depicted on Figure 7 for the 12 considered covariates. The top panel displays the conditional correlation (6) between the projected covariate $X^t \hat{v}^{(1)}(y_n)$ on the first direction and the response variable Y . It benefits from a stable behaviour with respect to the choice of $k \in \{50, \dots, 150\}$, confirming together with the previous Hill-plot and quantile-quantile plot that this range of number of exceedances is well-suited to the dataset. It also appears that the first index captures about 60% of the correlation ($\ell = 1$ in blue), while the second index fails to represent a significant correlation ($\ell = 2$ in red). The second direction is thus discarded in the sequel. The bottom panel of Figure 7 represents the conditional correlation (7) between the projected covariate $X^t \hat{v}^{(1)}(y_n)$ on the first direction and each coordinate $X^{(j)}$ of the covariate. One can note that small yields are mainly linked to operating costs that can be divided into two categories: agricultural inputs (fertilisers, pesticides, seeds and seedlings, works and services purchase, personal social security costs) and risk management (claims, crop insurance purchase, farm subsidies). This result could be expected since, in 2014, yields were strongly impacted by agricultural inputs, despite mild winter temperatures (which are favourable for wheat crops)².

²<http://agreste.agriculture.sg-ppd.maaf.ate.info/IMG/pdf/conjbilan2014.pdf> (in French).

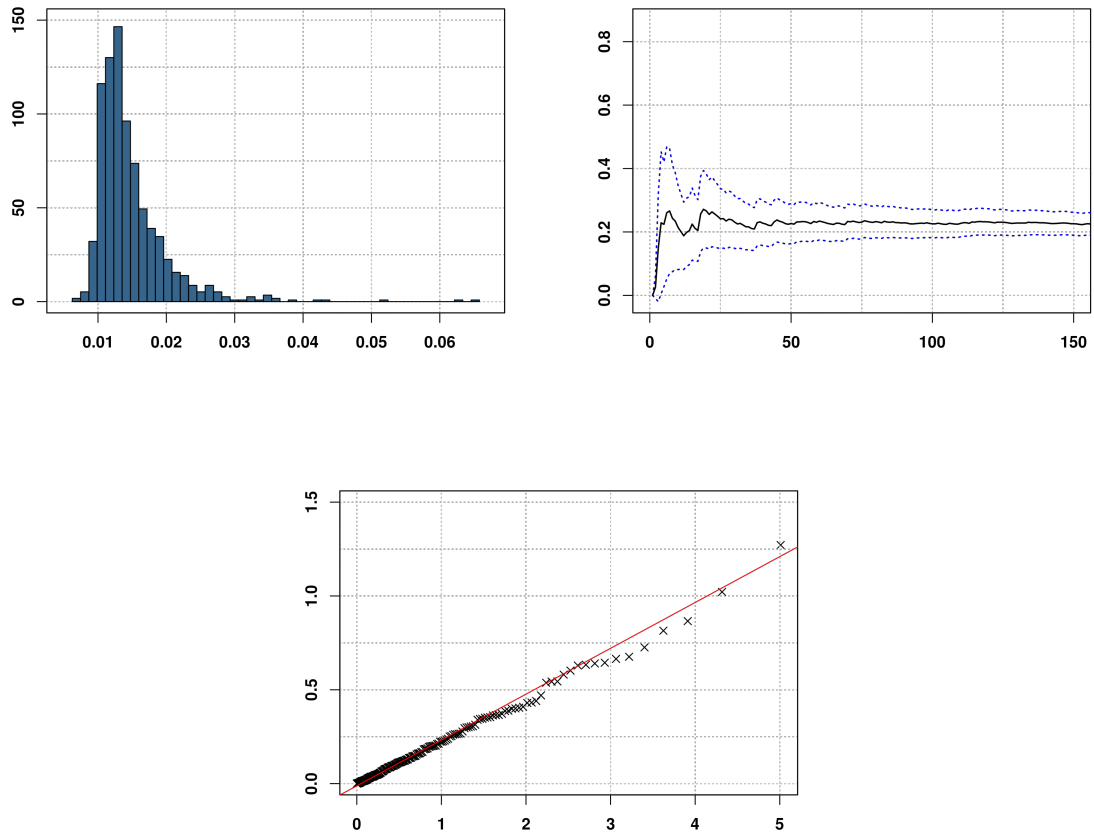


Figure 6: Farm income data. Top left panel: Histogram of the inverse yields Y_i , $i = 1, \dots, n$. Top right panel: Hill plot (Horizontally: $k \in \{1, \dots, 150\}$, vertically: Hill estimator $\hat{\gamma}_k$, dashed blue line: empirical 95% confidence interval). Bottom panel: quantile-quantile plot (horizontally: $\log(k/i)$, vertically: $\log(Y_{n-i+1,n}/Y_{n-k,n})$ for $i \in \{1, \dots, k = 150\}$, red: regression line).

Besides, the effect of crop insurance purchase could be explained by moral hazard that leads insured farmers to use fewer agricultural inputs [46].

In the following, we select $k = 97$ which corresponds to the highest correlation between $X^t \hat{v}^{(1)}(y_n)$ and Y in the range $\{50, \dots, 150\}$, see the top panel of Figure 7. The projected scatter plot $(Y_i, \hat{v}^{(1)}(y_n)^t X_i)$, $i = 1, \dots, n$ is displayed in a logarithmic scale for the visualization sake on the top panel of Figure 8 together with two estimations (linear and non-linear) of the conditional mean $\mathbb{E}(\hat{v}^{(1)}(y_n)^t X|Y)$. A positive trend appears for large values of Y in accordance to the inverse regression model (\mathbf{M}_1). Let us now focus on the conditional quantiles $\hat{q}(\alpha|\hat{v}^{(1)}(y_n)^t X)$ computed through a kernel estimator of the conditional survival function [15]. The results are reported in the bottom panel of Figure 8 together with the scatter plot $(\hat{v}^{(1)}(y_n)^t X_i, Y_i)$, $i = 1, \dots, n$. The vertical and horizontal axes are represented in a logarithmic scale. Both curves of the conditional quantiles associated with levels $\alpha = 0.15$ (blue line) and $\alpha = 0.05$ (red line) behave in a similar way. The estimated conditional quantiles of inverse yields feature an increasing shape for $\log(\hat{v}^{(1)}(y_n)^t X) \leq 9.5$: Lowest yields are (mainly) linked to high operating costs. The interpretation of the results for $\log(\hat{v}^{(1)}(y_n)^t X) > 9.5$ is difficult, the estimation being unreliable for large values of the covariate because of data sparsity in this area and boundary effects of kernel estimators, see for instance [36].

7 Discussion

We introduced a new approach EPLS for dimension reduction adapted to distribution tails. It allows to quantify the effect of covariates X on the extreme values of Y in a simple and interpretable way. The asymptotic properties of the estimated direction are established under an inverse single-index model and a heavy tail assumption but without recourse to linearity nor independence assumptions. The proposed method can then be used to facilitate the estimation of extreme conditional quantiles thanks to the dimension reduction which circumvents the curse of dimensionality. Quantifying the gain in terms of convergence rates would be of great interest and is the subject of our current work.

Acknowledgments. This work is supported by the French National Research Agency (ANR) in the framework of the Investissements d’Avenir Program (ANR-15-IDEX-02). S. Girard also acknowledges the support of the Chair Stress Test, Risk Management and Financial Steering, led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas.

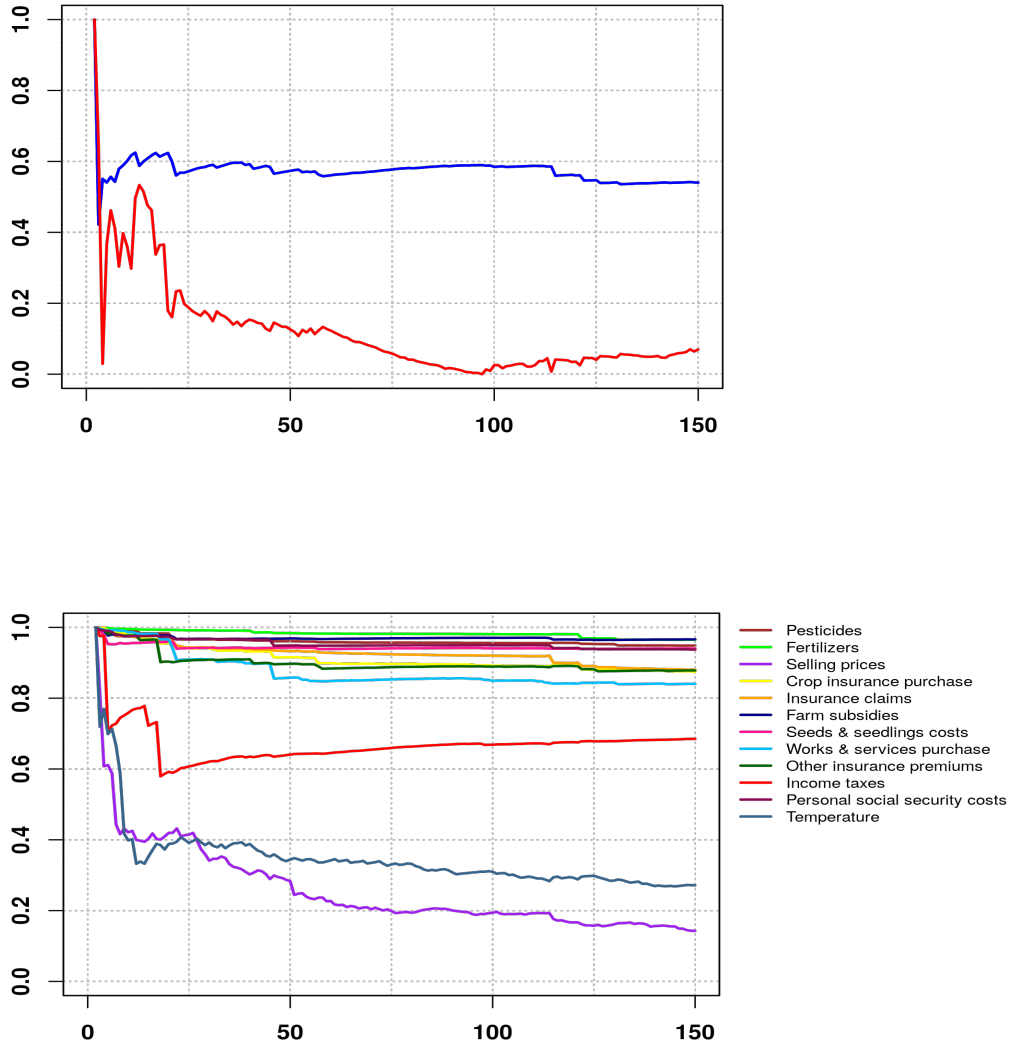


Figure 7: Farm income data. Top: Graph of the estimated conditional correlation function $y \mapsto \hat{\rho}(X^{t\hat{v}^{(\ell)}}(y), Y|Y \geq y)$, for $\ell = 1$ in blue and $\ell = 2$ in red. Bottom: Graph of the estimated conditional correlation function $y \mapsto \rho(X^{t\hat{v}^{(1)}}(y), X^{(j)}|Y \geq y)$ for $j = 1, \dots, 12$ (horizontally: number of exceedances k , vertically: conditional correlation estimated by its empirical counterpart using the threshold $y = Y_{n-k+1,n}$).

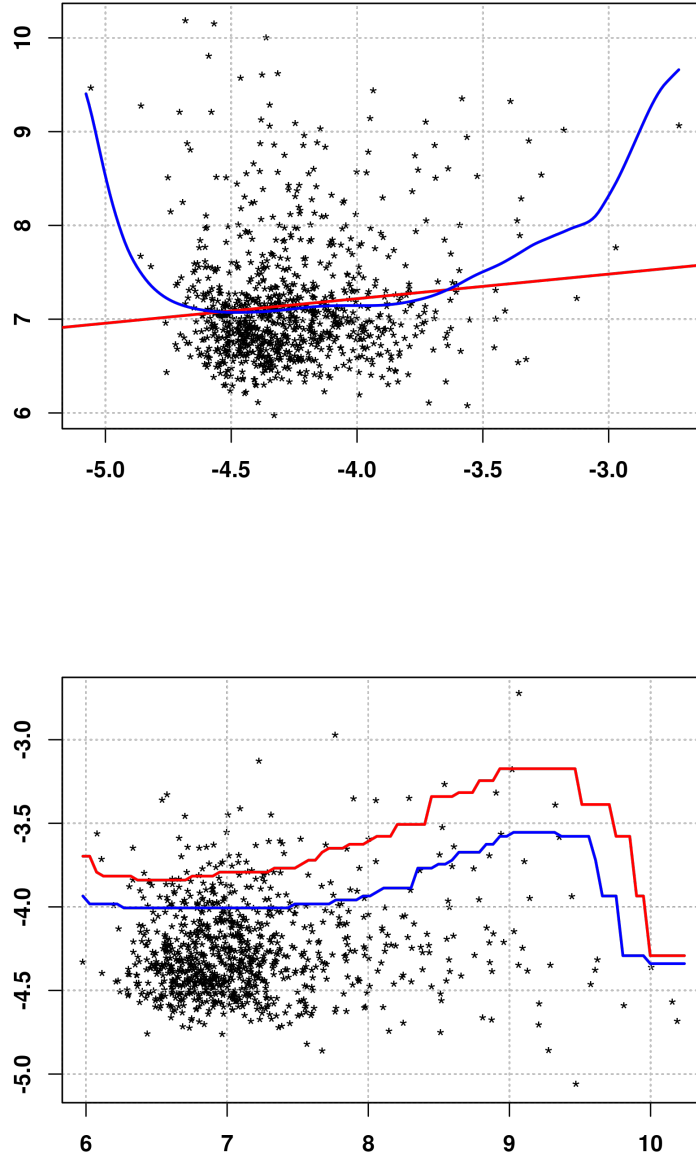


Figure 8: Farm income data. Top: scatter-plot $(Y_i, \hat{v}^{(1)}(y)^t X_i)$ in log scale obtained for $y = Y_{n-k+1,n}$ and $k = 97$ exceedances. The regression line (red) and a kernel estimate of the link function (blue) are superimposed. Bottom: scatter-plot $(\hat{v}^{(1)}(y)^t X_i, Y_i)$ in log scale obtained for $y = Y_{n-k+1,n}$ and $k = 97$ exceedances. The estimated conditional quantiles $\hat{q}(\alpha | \hat{v}(y)^t X)$ are superimposed ($\alpha = 0.15$: blue line, $\alpha = 0.05$: red line).

A Proofs

Some preliminary lemmas are first provided in Subsection A.1. They will reveal useful in the proofs of the main results collected in Subsection A.2.

A.1 Preliminary results

We begin with a technical tool to compute limits of integrals involving a regularly varying function.

Lemma 1. *Let $q_2 > 0$ and suppose Y is a random variable satisfying (\mathbf{A}_1) with $\gamma q_2 < 1$. Let $\phi(\cdot)$ be a continuous function on $[1, \infty)$ such that $\phi(t) \rightarrow \kappa > 1$ as $t \rightarrow \infty$. Then,*

$$\lim_{t \rightarrow \infty} \int_1^{+\infty} |x - \phi(t)|^{q_2} \frac{f(tx)}{f(t)} dx = \int_1^{+\infty} |x - \kappa|^{q_2} x^{-\frac{1}{\gamma}-1} dx < \infty.$$

Proof. Potter bounds entail that for all $\epsilon > 0$, $x \geq 1$ and t large enough:

$$0 \leq \frac{f(tx)}{f(t)} \leq (1 + \epsilon)x^{-\frac{1}{\gamma}-1+\epsilon}, \quad (8)$$

see for instance [26, Proposition B.1.9]. Besides, for t large enough, $\kappa/2 \leq \phi(t) \leq 2\kappa$, and therefore:

$$(1 - 2\kappa)x \leq x - 2\kappa \leq x - \phi(t) \leq x - \kappa/2 \leq x \leq (2\kappa - 1)x.$$

It follows that $|x - \phi(t)| \leq (2\kappa - 1)x$ and, since $q_2 > 0$, we have:

$$|x - \phi(t)|^{q_2} \leq (2\kappa - 1)^{q_2} x^{q_2}. \quad (9)$$

Collecting (8) and (9) yields:

$$0 \leq |x - \phi(t)|^{q_2} \frac{f(tx)}{f(t)} \leq (2\kappa - 1)^{q_2} x^{q_2 - \frac{1}{\gamma} - 1 + \epsilon}.$$

Recalling that $1/\gamma - q_2 > 0$, one can choose $0 < \epsilon < 1/\gamma - q_2$ such that the function $x \mapsto x^{q_2 - \frac{1}{\gamma} - 1 + \epsilon}$ is integrable on $[1, \infty)$. Then, Lebesgue's dominated convergence theorem together with the regular variation property of f conclude the proof. \square

The next lemma is an adaptation of [6, Proposition 1.5.10] to our setting. It provides an asymptotic equivalent of conditional expectations above a large threshold.

Lemma 2. *Suppose $\rho \in RV_\mu$ with $\mu \geq 0$ and Y is a random variable satisfying (\mathbf{A}_1) with $\gamma\mu < 1$. Then, as $y \rightarrow \infty$,*

$$\mathbb{E}[\rho(Y)|Y \geq y] \sim \frac{1}{1 - \gamma\mu} \rho(y).$$

Proof. Let us consider

$$\mathbb{E}[\rho(Y)|Y \geq y] = \frac{1}{\bar{F}(y)} \int_y^{+\infty} \rho(t)f(t)dt.$$

Since $\rho(\cdot)f(\cdot) \in RV_{\mu-1/\gamma-1}$, there exists a slowly-varying function L such that $\rho(t)f(t) = t^{\mu-\frac{1}{\gamma}-1}L(t)$. Then, [6, Proposition 1.5.10] shows that

$$\mathbb{E}[\rho(Y)|Y \geq y] \sim \frac{1}{\bar{F}(y)} \frac{y^{\mu-\frac{1}{\gamma}}L(y)}{1/\gamma-\mu} = \frac{\gamma}{1-\gamma\mu} \frac{y\rho(y)f(y)}{\bar{F}(y)},$$

as $y \rightarrow \infty$. Recalling that $f \in RV_{-1/\gamma-1}$ and using again [6, Proposition 1.5.10] prove that $\gamma y f(y) \sim \bar{F}(y)$ as $y \rightarrow \infty$. Finally,

$$\mathbb{E}[\rho(Y)|Y \geq y] \sim \frac{\rho(y)}{1-\gamma\mu},$$

as $y \rightarrow \infty$ and the conclusion follows. \square

The following lemma establishes sufficient conditions such that the moment conditions of Proposition 1 hold in the context of the inverse regression model (\mathbf{M}_1) .

Lemma 3. *Assume (\mathbf{M}_1) , (\mathbf{A}_1) , (\mathbf{A}_2) and (\mathbf{A}_3) hold with $\gamma(c+1) < 1$. Then, $\mathbb{E}|Y| < \infty$, $\mathbb{E}\|XY\| < \infty$ and $\mathbb{E}\|X\| < \infty$.*

Proof. First, let us recall that the existence of $\mathbb{E}|Y|$ in the Fréchet maximum domain of attraction is a consequence of $\gamma < 1$. Second, the triangle inequality yields:

$$\mathbb{E}\|X\| < \mathbb{E}|g(Y)| + \mathbb{E}\|\varepsilon\|.$$

Let us note that $\mathbb{E}|g(Y)| < \infty$ since $c\gamma < 1$. Besides, for all $q \geq 1/(\gamma c) \geq 1$, Jensen's inequality entails $(\mathbb{E}\|\varepsilon\|)^q \leq \mathbb{E}(\|\varepsilon\|^q) < \infty$ under (\mathbf{A}_3) . Hence, $\mathbb{E}\|\varepsilon\| < \infty$ and $\mathbb{E}\|X\| < \infty$. Third,

$$\mathbb{E}\|XY\| < \mathbb{E}|Yg(Y)| + \mathbb{E}\|Y\varepsilon\|,$$

and $\mathbb{E}|Yg(Y)| < \infty$ in view of $\gamma(c+1) < 1$. Furthermore, Hölder inequality shows that

$$\mathbb{E}\|Y\varepsilon\| < [\mathbb{E}(\|\varepsilon\|^q)]^{1/q} [\mathbb{E}(|Y|^{q_2})]^{1/q_2},$$

for all $q_2 \geq 1$ such that $1/q + 1/q_2 = 1$. As already remarked, $\mathbb{E}(\|\varepsilon\|^q) < \infty$ from (\mathbf{A}_3) with $q > 1/(\gamma c)$. Moreover, taking account of condition $\gamma(c+1) < 1$ yields $q > 1/(1-\gamma)$ which is equivalent to $q_2 < 1/\gamma$, and therefore $\mathbb{E}(|Y|^{q_2}) < \infty$ as well. As a conclusion, $\mathbb{E}\|XY\| < \infty$ and the result is proved. \square

Lemma 4 provides, in the framework of model (\mathbf{M}_1) , an alternative expression of the direction $v(y)$ defined in Proposition 1.

Lemma 4. *Assume model (\mathbf{M}_1) and suppose the assumptions of Proposition 1 hold. Then, for all $y \in \mathbb{R}$, $v(y)$ can be rewritten as*

$$v(y) = \bar{F}(y)\mathbb{E}[g(Y)\Psi_y(Y)](\beta + \eta(y)), \quad (10)$$

where

$$\eta(y) := \frac{\mathbb{E}[\varepsilon\Psi_y(Y)]}{\mathbb{E}[g(Y)\Psi_y(Y)]} \text{ and } \Psi_y(Y) := \left(Y - \frac{m_Y(y)}{\bar{F}(y)}\right) \mathbf{1}_{\{Y \geq y\}}. \quad (11)$$

Proof. Proposition 1 states that

$$v(y) = \bar{F}(y)\mathbb{E}[XY\mathbf{1}_{\{Y \geq y\}}] - \mathbb{E}[X\mathbf{1}_{\{Y \geq y\}}]m_Y(y).$$

Recalling that $X = g(Y)\beta + \varepsilon$ from model (\mathbf{M}_1) yields

$$\begin{aligned} v(y) &= \bar{F}(y)\mathbb{E}[(g(Y)\beta + \varepsilon)Y\mathbf{1}_{\{Y \geq y\}}] - \mathbb{E}[(g(Y)\beta + \varepsilon)\mathbf{1}_{\{Y \geq y\}}]m_Y(y) \\ &= \beta\mathbb{E}[g(Y)\mathbf{1}_{\{Y \geq y\}}(\bar{F}(y)Y - m_Y(y))] + \mathbb{E}[\varepsilon\mathbf{1}_{\{Y \geq y\}}(\bar{F}(y)Y - m_Y(y))] \\ &= \beta\bar{F}(y)\mathbb{E}[g(Y)\Psi_y(Y)] + \bar{F}(y)\mathbb{E}[\varepsilon\Psi_y(Y)] \\ &= \bar{F}(y)\mathbb{E}[g(Y)\Psi_y(Y)] \left(\beta + \frac{\mathbb{E}[\varepsilon\Psi_y(Y)]}{\mathbb{E}[g(Y)\Psi_y(Y)]} \right). \end{aligned}$$

Hence the result. \square

We first establish a precise control of the moments of the random variable $\Psi_y(Y)$ appearing in the numerator of the remainder term $\eta(y)$, see (11) in Lemma 4.

Lemma 5. *Let $q_2 > 0$ and suppose Y is a random variable satisfying (\mathbf{A}_1) with $\gamma q_2 < 1$. Then,*

$$\mathbb{E}(|\Psi_y(Y)|^{q_2}) \sim \lambda_1(\gamma, q_2)y^{q_2+1}f(y),$$

as $y \rightarrow \infty$ and where $\lambda_1(\gamma, q_2)$ is a positive constant.

Proof. From (11), one has

$$\mathbb{E}|\Psi_y(Y)|^{q_2} = \int_y^{+\infty} \left| t - \frac{m_Y(y)}{\bar{F}(y)} \right|^{q_2} f(t) dt = y^{q_2+1} f(y) \int_1^{+\infty} \left| x - \frac{1}{y} \mathbb{E}[Y|Y \geq y] \right|^{q_2} \frac{f(yx)}{f(y)} dx,$$

thanks to the change of variable $x = t/y$ and recalling that $m_Y(y)/\bar{F}(y) = \mathbb{E}[Y|Y \geq y]$. Since $f \in RV_{-1/\gamma-1}$, $\gamma \in (0, 1)$, Lemma 2 applied with $\rho(t) = t$ and $\mu = 1$ entails that

$$\phi(y) := \frac{1}{y} \mathbb{E}[Y|Y \geq y] \rightarrow \frac{1}{1-\gamma} =: \kappa \geq 1,$$

as $y \rightarrow \infty$. Lemma 1 then yields, as $y \rightarrow \infty$,

$$\int_1^{+\infty} \left| x - \frac{1}{y} \mathbb{E}[Y|Y \geq y] \right|^{q_2} \frac{f(yx)}{f(y)} dx \rightarrow \int_1^{+\infty} \left| x - \frac{1}{1-\gamma} \right|^{q_2} x^{-\frac{1}{\gamma}-1} dx =: \lambda_1(\gamma, q_2),$$

As a conclusion, $\mathbb{E}(|\Psi_y(Y)|^{q_2}) \sim \lambda_1(\gamma, q_2) y^{q_2+1} f(y)$, as $y \rightarrow \infty$ and the result is proved. \square

Similarly, we provide an asymptotic equivalent of the moments of the random variable $g(Y)\Psi_y(Y)$ appearing in the denominator of the remainder term $\eta(y)$, see (11) in Lemma 4.

Lemma 6. *Let Y be a random variable satisfying (\mathbf{A}_1) and (\mathbf{A}_2) with $\gamma(c+1) < 1$. Then,*

$$\mathbb{E}[g(Y)\Psi_y(Y)] \sim \lambda_2(\gamma, c) yg(y)\bar{F}(y),$$

as $y \rightarrow \infty$ and where $\lambda_2(\gamma, c) := \frac{\gamma^2 c}{(1-\gamma(c+1))(1-\gamma c)(1-\gamma)}$.

Proof. From (11), we have:

$$\begin{aligned} \frac{\mathbb{E}[g(Y)\Psi_y(Y)]}{\bar{F}(y)} &= \frac{\mathbb{E}[Yg(Y)\mathbf{1}_{\{Y \geq y\}}]}{\bar{F}(y)} - \frac{\mathbb{E}[Y\mathbf{1}_{\{Y \geq y\}}]}{\bar{F}(y)} \frac{\mathbb{E}[g(Y)\mathbf{1}_{\{Y \geq y\}}]}{\bar{F}(y)} \\ &= \mathbb{E}[Yg(Y)|Y \geq y] - \mathbb{E}[Y|Y \geq y]\mathbb{E}[g(Y)|Y \geq y]. \end{aligned}$$

Let us consider $\rho_1(y) = yg(y) \in RV_{c+1}$, $\rho_2(y) = y \in RV_1$ and $\rho_3(y) = g(y) \in RV_c$. Lemma 2 entails as $y \rightarrow \infty$:

$$\begin{aligned} \mathbb{E}(\rho_1(Y)|Y \geq y) &\sim \frac{1}{1-\gamma(c+1)} yg(y), \\ \mathbb{E}(\rho_2(Y)|Y \geq y) &\sim \frac{1}{1-\gamma} y, \\ \mathbb{E}(\rho_3(Y)|Y \geq y) &\sim \frac{1}{1-\gamma c} g(y), \end{aligned}$$

which concludes the proof. \square

The next lemma applied successively with $\zeta = 0$ and $\zeta = 1$ yields asymptotic equivalents for these two quantities in the two situations where $\beta_j = 0$ and $\beta_j \neq 0$.

Lemma 7. *Assume (\mathbf{M}_1) , (\mathbf{A}_1) , (\mathbf{A}_2) and (\mathbf{A}_3) hold with $\gamma(c+1) < 1$. Let $\zeta \in \{0, 1\}$. Then,*

(i) *For all $j \in \{1, \dots, p\}$ such that $\beta_j \neq 0$, we have*

$$m_{X_{\cdot, j} Y^\zeta}(y) := \mathbb{E}(X_{\cdot, j} Y^\zeta \mathbf{1}_{\{Y \geq y\}}) \sim \frac{\beta_j}{1-\gamma(c+\zeta)} y^\zeta g(y) \bar{F}(y),$$

as $y \rightarrow \infty$.

(ii) For all $j \in \{1, \dots, p\}$ such that $\beta_j = 0$, we have

$$m_{X_{\cdot,j}Y^\zeta}(y) := \mathbb{E}(\varepsilon_{\cdot,j}Y^\zeta \mathbf{1}_{\{Y \geq y\}}) = O\left(y^\zeta \bar{F}(y)^{1-1/q}\right),$$

as $y \rightarrow \infty$.

Remark that, in view of the above lemma, condition (\mathbf{A}_3) implies that, for all $j \in \{1, \dots, p\}$ such that $\beta_j = 0$, $m_{X_{\cdot,j}Y^\zeta}(y)$ is negligible compared to each $m_{X_{\cdot,\ell}Y^\zeta}(y)$ associated with $\beta_\ell \neq 0$.

Proof. From (\mathbf{M}_1) , we have:

$$\begin{aligned} \mathbb{E}(X_{\cdot,j}Y^\zeta \mathbf{1}_{\{Y \geq y\}}) &= \mathbb{E}(Y^\zeta g(Y) \mathbf{1}_{\{Y \geq y\}}) \beta_j + \mathbb{E}(Y^\zeta \varepsilon_{\cdot,j} \mathbf{1}_{\{Y \geq y\}}) \\ &= \mathbb{E}(Y^\zeta g(Y) | Y \geq y) \bar{F}(y) \beta_j + \mathbb{E}(Y^\zeta \varepsilon_{\cdot,j} \mathbf{1}_{\{Y \geq y\}}) \\ &= \frac{\beta_j}{1 - \gamma(c + \zeta)} y^\zeta g(y) \bar{F}(y) (1 + o(1)) + \mathbb{E}(Y^\zeta \varepsilon_{\cdot,j} \mathbf{1}_{\{Y \geq y\}}), \end{aligned} \quad (12)$$

by applying Lemma 2 to the first term. Let $q_2 \geq 1$ such that $1/q + 1/q_2 = 1$. Combining Hölder inequality with (\mathbf{A}_3) yields:

$$\begin{aligned} |\mathbb{E}(\varepsilon_{\cdot,j}Y^\zeta \mathbf{1}_{\{Y \geq y\}})| &\leq [\mathbb{E}|\varepsilon_{\cdot,j}|^q]^{1/q} [\mathbb{E}(|Y|^{\zeta q_2} \mathbf{1}_{\{Y \geq y\}})]^{1/q_2} \\ &= [\mathbb{E}|\varepsilon_{\cdot,j}|^q]^{1/q} [\mathbb{E}(|Y|^{\zeta q_2} | Y \geq y)]^{1/q_2} \bar{F}(y)^{1/q_2} \\ &= [\mathbb{E}|\varepsilon_{\cdot,j}|^q]^{1/q} \left(\frac{|y|^{\zeta q_2}}{1 - q_2 \zeta \gamma} \right)^{1/q_2} \bar{F}(y)^{1/q_2} (1 + o(1)) \\ &= O\left(y^\zeta \bar{F}(y)^{1-1/q}\right), \end{aligned} \quad (13)$$

as $y \rightarrow \infty$, according to Lemma 2 and since $q_2 \zeta \gamma < 1$. This proves (ii) when $\beta_j = 0$. Focusing on the situation where $\beta_j \neq 0$, we have from (12) and (13),

$$\begin{aligned} \mathbb{E}(X_{\cdot,j}Y^\zeta \mathbf{1}_{\{Y \geq y\}}) &= \frac{\beta_j}{1 - \gamma(c + \zeta)} y^\zeta g(y) \bar{F}(y) (1 + o(1)) + O(y^\zeta \bar{F}(y)^{1-1/q}) \\ &= \frac{\beta_j}{1 - \gamma(c + \zeta)} y^\zeta g(y) \bar{F}(y) \left(1 + o(1) + O\left(\frac{1}{\bar{F}(y)^{1/q} g(y)}\right) \right). \end{aligned}$$

As a consequence of (\mathbf{A}_1) and (\mathbf{A}_2) , $\bar{F}(\cdot)^{1/q} g(\cdot)$ is a regularly varying function with index $c - 1/(q\gamma) > 0$. Therefore, $\bar{F}(y)^{1/q} g(y) \rightarrow \infty$ when $y \rightarrow \infty$ and (i) is proved. \square

The last lemma proves that the noise term ε does not contribute to the asymptotic distribution of the estimators.

Lemma 8. Assume (\mathbf{M}_1) , (\mathbf{A}_1) , (\mathbf{A}_2) and (\mathbf{A}_3) hold with $2\gamma(c + 1) < 1$. For all $\zeta \in \{0, 1\}$ let

$$T_{\cdot,n}^{(\zeta)} = \sqrt{\bar{F}(y_n)} \left(\sum_{\beta_j \neq 0} \frac{\alpha_j^{(\zeta)} \varepsilon_{\cdot,j}}{m_{X_{\cdot,j}Y^\zeta}(y_n)} \right) Y^\zeta \mathbf{1}_{\{Y \geq y_n\}},$$

where $\alpha_j^{(\zeta)} \in \mathbb{R}$ for all $j = 1, \dots, p$. Then,

$$\chi_n^{(\zeta)} := \frac{1}{\sqrt{n}} \sum_{i=1}^n (T_{i,n}^{(\zeta)} - \mathbb{E}(T_{i,n}^{(\zeta)})) \xrightarrow{P} 0.$$

Proof. Clearly, $\chi_n^{(\zeta)}$ is centered by definition. Let us consider its variance:

$$\begin{aligned} \text{var}(\chi_n^{(\zeta)}) &= \text{var}(T_{\cdot,n}^{(\zeta)}) \\ &= \sum_{\beta_j \neq 0} \sum_{\beta_\ell \neq 0} \frac{\alpha_j^{(\zeta)} \alpha_\ell^{(\zeta)} \bar{F}(y_n)}{m_{X_{\cdot,j} Y^\zeta}(y_n) m_{X_{\cdot,\ell} Y^\zeta}(y_n)} \text{cov}(\varepsilon_{\cdot,j} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}, \varepsilon_{\cdot,\ell} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}) \\ &\sim \frac{(1 - \gamma(c + \zeta))^2}{y_n^{2\zeta} g(y_n)^2 \bar{F}(y_n)} \sum_{\beta_j \neq 0} \sum_{\beta_\ell \neq 0} \frac{\alpha_j^{(\zeta)} \alpha_\ell^{(\zeta)}}{\beta_j \beta_\ell} \text{cov}(\varepsilon_{\cdot,j} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}, \varepsilon_{\cdot,\ell} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}), \end{aligned}$$

as $n \rightarrow \infty$, from Lemma 7(i). The covariance can be expanded as

$$\begin{aligned} \text{cov}(\varepsilon_{\cdot,j} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}, \varepsilon_{\cdot,\ell} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}) &= \mathbb{E}(\varepsilon_{\cdot,j} \varepsilon_{\cdot,\ell} Y^{2\zeta} \mathbb{1}_{\{Y \geq y_n\}}) \\ &\quad - \mathbb{E}(\varepsilon_{\cdot,j} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}) \mathbb{E}(\varepsilon_{\cdot,\ell} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}). \end{aligned}$$

The first term is bounded using Hölder inequality, applied for all $q_3 \geq 1$ such that $2/q + 1/q_3 = 1$:

$$\begin{aligned} \mathbb{E}(\varepsilon_{\cdot,j} \varepsilon_{\cdot,\ell} Y^{2\zeta} \mathbb{1}_{\{Y \geq y_n\}}) &\leq [\mathbb{E}|\varepsilon_{\cdot,j}|^q]^{1/q} [\mathbb{E}|\varepsilon_{\cdot,\ell}|^q]^{1/q} [\mathbb{E}|Y^{2\zeta} \mathbb{1}_{\{Y \geq y_n\}}|^{q_3}]^{1/q_3} \\ &\leq [\mathbb{E}|\varepsilon_{\cdot,j}|^q]^{1/q} [\mathbb{E}|\varepsilon_{\cdot,\ell}|^q]^{1/q} [\mathbb{E}|Y^{2\zeta q_3} | Y \geq y_n|]^{1/q_3} \bar{F}(y_n)^{1/q_3} \\ &= [\mathbb{E}|\varepsilon_{\cdot,j}|^q]^{1/q} [\mathbb{E}|\varepsilon_{\cdot,\ell}|^q]^{1/q} \left(\frac{y_n^{2\zeta q_3}}{1 - 2\zeta \gamma p'} \right)^{1/q_3} \bar{F}(y_n)^{1/q_3} (1 + o(1)) \\ &= O\left(y_n^{2\zeta} \bar{F}(y_n)^{1-2/q}\right), \end{aligned} \tag{14}$$

in view of Lemma 2 and (\mathbf{A}_3) . Indeed, condition $2\gamma(c + 1) < 1$ is equivalent to $\gamma c < 1/2 - \gamma$. Besides, from (\mathbf{A}_3) , $q > 1/(\gamma c)$ and thus $q > 2/(1 - 2\gamma)$ leading to $2\gamma q_3 < 1$. The second term is controlled with Lemma 7(ii) and is negligible compared to the first one:

$$|\mathbb{E}(\varepsilon_{\cdot,j} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}) \mathbb{E}(\varepsilon_{\cdot,\ell} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}})| = O(y_n^{2\zeta} \bar{F}(y_n)^{2-2/q}) = o\left(y_n^{2\zeta} \bar{F}(y_n)^{1-2/q}\right). \tag{15}$$

Taking account of (14) and (15) yields

$$\text{cov}(\varepsilon_{\cdot,j} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}, \varepsilon_{\cdot,\ell} Y^\zeta \mathbb{1}_{\{Y \geq y_n\}}) = O\left(y_n^{2\zeta} \bar{F}(y_n)^{1-2/q}\right), \tag{16}$$

and therefore

$$\text{var}(\chi_n^{(\zeta)}) = O\left(\frac{1}{\bar{F}(y_n)^{2/q} g(y_n)^2}\right) \rightarrow 0,$$

as $n \rightarrow \infty$ since $\bar{F}(\cdot)^{2/q} g(\cdot)^2$ is regularly-varying with index $2(c - 1/(q\gamma)) > 0$. The conclusion follows. \square

A.2 Proofs of main results

Proof of Proposition 1. Let us rewrite the optimization problem as

$$\begin{aligned}
w(y) &= \arg \max_{\|\beta\|=1} \text{cov}(\beta^t X, Y | Y \geq y) \\
&= \arg \max_{\|\beta\|=1} \frac{\mathbb{E}(\beta^t X Y \mathbf{1}_{\{Y \geq y\}})}{\bar{F}(y)} - \frac{\mathbb{E}(\beta^t X \mathbf{1}_{\{Y \geq y\}}) \mathbb{E}(Y \mathbf{1}_{\{Y \geq y\}})}{\bar{F}(y)^2} \\
&= \arg \max_{\|\beta\|=1} \bar{F}(y) \beta^t m_{XY}(y) - \beta^t m_X(y) m_Y(y) \\
&= \arg \max_{\|\beta\|=1} \beta^t v(y).
\end{aligned}$$

This constrained optimization problem is solved using Lagrange multipliers method. Introducing

$$\mathcal{L}(w, \lambda) = w^t v(y) - \frac{\lambda}{2} (\|w\|^2 - 1), \quad \lambda \in \mathbb{R},$$

and setting the partial derivatives to zero yield:

$$\begin{cases} \nabla_{\lambda} \mathcal{L}(w, \lambda) &= -\frac{1}{2} (\|w\|^2 - 1) = 0, \\ \nabla_w \mathcal{L}(w, \lambda) &= v(y) - \lambda w = 0, \end{cases}$$

or equivalently,

$$\begin{cases} \|w\|^2 &= 1, \\ w &= v(y)/\lambda. \end{cases}$$

It straightforwardly follows that $\lambda = \|v(y)\|$ and $w = v(y)/\|v(y)\|$. \square

Proof of Proposition 2. From Lemma 3, $\mathbb{E}|Y|$, $\mathbb{E}\|XY\|$ and $\mathbb{E}\|X\|$ exist. We may then apply Lemma 4 to get:

$$\cos(w(y), \beta) = w(y)^t \beta = \text{sign}(\mathbb{E}[g(Y)\Psi_y(Y)]) \frac{1 + \beta^t \eta(y)}{\|\beta + \eta(y)\|}, \quad (17)$$

with $\eta(y) = \mathbb{E}[\varepsilon \Psi_y(Y)] / \mathbb{E}[g(Y)\Psi_y(Y)]$, see (11), and where $\text{sign}(u) = 1$ if $u \geq 0$ and $\text{sign}(u) = -1$ otherwise. Straightforward calculations yield

$$\cos^2(w(y), \beta) - 1 = \frac{(\beta^t \eta(y))^2 - \|\eta(y)\|^2}{\|\beta + \eta(y)\|^2},$$

and therefore it is sufficient to prove that $\|\eta(y)\| \rightarrow 0$ as $y \rightarrow \infty$ to get

$$1 - \cos^2(w(y), \beta) = O(\|\eta(y)\|^2), \quad (18)$$

as $y \rightarrow \infty$. Under assumption (\mathbf{A}_3) , there exists $q > 1/(\gamma c)$ such that $\mathbb{E}\|\varepsilon\|^q < \infty$. Hölder inequality thus yields

$$\|\mathbb{E}[\varepsilon\Psi_y(Y)]\| \leq [\mathbb{E}\|\varepsilon\|^q]^{1/q}[\mathbb{E}|\Psi_y(Y)|^{q_2}]^{1/q_2},$$

for all $q_2 \geq 1$ such that $1/q + 1/q_2 = 1$. As a consequence, $\gamma(c+1) < 1$ and $\gamma c > 1/q$ imply $\gamma q_2 < 1$ and then Lemma 5 shows that $\mathbb{E}|\Psi_y(Y)|^{q_2} \sim \lambda_1(\gamma, q_2)y^{q_2+1}f(y)$ as $y \rightarrow \infty$, with $\lambda_1(\gamma, q_2) > 0$. Therefore,

$$\|\mathbb{E}[\varepsilon\Psi_y(Y)]\| \leq [\mathbb{E}\|\varepsilon\|^q]^{1/q}(\lambda_1(\gamma, q_2))^{1/q_2}y^{1+1/q_2}f(y)^{1/q_2}. \quad (19)$$

and Lemma 6 shows that, as $y \rightarrow \infty$,

$$\mathbb{E}[g(Y)\Psi_y(Y)] \sim \lambda_2(\gamma, c)yg(y)\bar{F}(y), \quad (20)$$

with $\lambda_2(\gamma, c) > 0$. Collecting (19) and (20) thus yields:

$$\|\eta(y)\| \leq \frac{[\mathbb{E}\|\varepsilon\|^q]^{1/q}(\lambda_1(\gamma, q_2))^{1/q_2}}{\lambda_2(\gamma, c)} \frac{(yf(y))^{1/q_2}}{g(y)\bar{F}(y)}(1+o(1)) = O\left(\frac{1}{g(y)\bar{F}^{1/q}(y)}\right). \quad (21)$$

Under assumptions (\mathbf{A}_1) and (\mathbf{A}_2) , $\bar{F} \in RV_{-1/\gamma}$ and $g \in RV_c$ so that $y \mapsto g(y)\bar{F}^{1/q}(y)$ is also regularly varying with index $c - 1/(q\gamma) > 0$. As a consequence, $\|\eta(y)\| \rightarrow 0$ as $y \rightarrow \infty$ and the first part of the result is proved. Second, (20) shows that, for y large enough, $\mathbb{E}[g(Y)\Psi_y(Y)] > 0$. Consequently, $\cos(w(y), \beta) \geq 0$ eventually in view of (17), and then (18) entails that $\cos(w(y), \beta) \rightarrow 1$ as $y \rightarrow \infty$ leading to

$$\|w(y) - \beta\| = \sqrt{2(1 - \cos(w(y), \beta))} \sim \sqrt{1 - \cos^2(w(y), \beta)} = O(\|\eta(y)\|) = O\left(\frac{1}{g(y)\bar{F}^{1/q}(y)}\right),$$

as $y \rightarrow \infty$, from (21). The result is proved. \square

Proof of Proposition 3. To establish the joint asymptotic normality of the $2(d+1)$ - random vector Λ_n , we shall prove that any non-zero linear combination of its components is asymptotically Gaussian. Set $\alpha = (\alpha_1, \alpha_2, \alpha_{3,1}, \dots, \alpha_{3,d}, \alpha_{4,1}, \dots, \alpha_{4,d})^t \in \mathbb{R}^{2(d+1)}$ and let us investigate the asymptotic distribution of χ_n defined as follows:

$$\begin{aligned} \chi_n &= \sqrt{n\bar{F}(y_n)} \left\{ \alpha_1 \left(\frac{\hat{F}(y_n)}{\bar{F}(y_n)} - 1 \right) + \alpha_2 \left(\frac{\hat{m}_Y(y_n)}{m_Y(y_n)} - 1 \right) \right\} \\ &+ \sqrt{n\bar{F}(y_n)} \left\{ \sum_{j=1}^d \alpha_{3,j} \left(\frac{\hat{m}_{X_{\cdot,j}}(y_n)}{m_{X_{\cdot,j}}(y_n)} - 1 \right) + \alpha_{4,j} \left(\frac{\hat{m}_{X_{\cdot,j}Y}(y_n)}{m_{X_{\cdot,j}Y}(y_n)} - 1 \right) \right\}, \end{aligned}$$

and which can be rewritten as $\chi_n = \sum_{i=1}^n \chi_{i,n} := \sum_{i=1}^n (Z_{i,n} - \mathbb{E}(Z_{i,n}))$, where

$$Z_{i,n} = \sqrt{\frac{\bar{F}(y_n)}{n}} \left(\frac{\alpha_1}{\bar{F}(y_n)} + \frac{\alpha_2 Y_i}{m_Y(y_n)} + \sum_{j=1}^d \frac{\alpha_{3,j} X_{i,j}}{m_{X_{\cdot,j}}(y_n)} + \sum_{j=1}^d \frac{\alpha_{4,j} X_{i,j} Y_i}{m_{X_{\cdot,j}Y}(y_n)} \right) \mathbb{1}_{\{Y_i \geq y_n\}}.$$

Under model (\mathbf{M}_1) , $X_{i,j} = g(Y_i)\beta_j + \varepsilon_{i,j}$, for $j \in \{1, \dots, d\}$ and $i \in \{1, \dots, n\}$, we get the following decomposition:

$$Z_{i,n} \stackrel{d}{=} \frac{1}{\sqrt{n}} (T_{1,i,n} + T_{2,i,n} + T_{3,i,n} + T'_{3,i,n} + T_{4,i,n} + T'_{4,i,n}),$$

where

$$\begin{aligned} T_{1,i,n} &= \frac{\alpha_1}{\sqrt{\bar{F}(y_n)}} \mathbb{1}_{\{Y_i \geq y_n\}}, & T_{2,i,n} &= \sqrt{\bar{F}(y_n)} \frac{\alpha_2}{m_Y(y_n)} Y_i \mathbb{1}_{\{Y_i \geq y_n\}}, \\ T_{3,i,n} &= \sqrt{\bar{F}(y_n)} \left(\sum_{j=1}^d \frac{\alpha_{3,j} \beta_j}{m_{X_{\cdot,j}}(y_n)} \right) g(Y_i) \mathbb{1}_{\{Y_i \geq y_n\}}, & T'_{3,i,n} &= \sqrt{\bar{F}(y_n)} \left(\sum_{j=1}^d \frac{\alpha_{3,j} \varepsilon_{i,j}}{m_{X_{\cdot,j}}(y_n)} \right) \mathbb{1}_{\{Y_i \geq y_n\}}, \\ T_{4,i,n} &= \sqrt{\bar{F}(y_n)} \left(\sum_{j=1}^d \frac{\alpha_{4,j} \beta_j}{m_{X_{\cdot,j}Y}(y_n)} \right) Y_i g(Y_i) \mathbb{1}_{\{Y_i \geq y_n\}}, & T'_{4,i,n} &= \sqrt{\bar{F}(y_n)} \left(\sum_{j=1}^d \frac{\alpha_{4,j} \varepsilon_{i,j}}{m_{X_{\cdot,j}Y}(y_n)} \right) Y_i \mathbb{1}_{\{Y_i \geq y_n\}}. \end{aligned}$$

Substituting yields $\chi_n = \chi'_{0,n} + \chi'_{3,n} + \chi'_{4,n}$ where

$$\begin{aligned} \chi'_{0,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{\ell=1}^4 (T_{\ell,i,n} - \mathbb{E}(T_{\ell,i,n})), \\ \chi'_{3,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (T'_{3,i,n} - \mathbb{E}(T'_{3,i,n})), \\ \chi'_{4,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (T'_{4,i,n} - \mathbb{E}(T'_{4,i,n})). \end{aligned}$$

Lemma 8 shows that both $\chi'_{3,n}$ and $\chi'_{4,n}$ converge in probability to zero and we therefore focus on the limiting distribution of $\chi'_{0,n}$. Clearly, $\mathbb{E}(\chi'_{0,n}) = 0$. Turning to the variance of $\chi'_{0,n}$ and, since we deal with independent and identically distributed random variables, one has

$$\text{var}(\chi'_{0,n}) = \text{var} \left(\sum_{\ell=1}^4 T_{\ell,i,n} \right) = \sum_{\ell=1}^4 \text{var}(T_{\ell,i,n}) + 2 \sum_{1 \leq \ell < m \leq 4} \text{cov}(T_{\ell,i,n}, T_{m,i,n}),$$

for all $i \in \{1, \dots, n\}$. As a preliminary result, the following asymptotic equivalent holds for all $(\zeta, \omega) \in \{0, 1, 2\}^2$:

$$\mathbb{E}(Y^\zeta g(Y)^\omega \mathbb{1}_{\{Y \geq y_n\}}) \sim \frac{1}{1 - \gamma(\omega c + \zeta)} y_n^\zeta g(y_n)^\omega \bar{F}(y_n), \quad (22)$$

as $n \rightarrow \infty$, in view of Lemma 2 and assumption $2\gamma(c+1) < 1$. Moreover, Lemma 2 and Lemma 7(i) yield the following asymptotic equivalents:

$$m_Y(y_n) \sim \frac{1}{1-\gamma} y_n \bar{F}(y_n), \quad (23)$$

$$m_{X_{\cdot,j}}(y_n) \sim \frac{\beta_j}{1-\gamma c} g(y_n) \bar{F}(y_n), \quad (24)$$

$$m_{X_{\cdot,j}Y}(y_n) \sim \frac{\beta_j}{1-\gamma(c+1)} y_n g(y_n) \bar{F}(y_n), \quad (25)$$

which will reveal useful in evaluation of variances and covariances below. We shall also use the notation $\langle \alpha_\ell \rangle = \sum_{j=1}^d \alpha_{\ell,j}$ for $\ell \in \{3, 4\}$. A straightforward calculation yields:

$$\text{var}(T_{1,i,n}) = \alpha_1^2 (1 - \bar{F}(y_n)) \rightarrow \alpha_1^2 \text{ as } n \rightarrow \infty.$$

Combining (22) and (23) leads to:

$$\text{var}(T_{2,i,n}) = \alpha_2^2 \bar{F}(y_n) \left(\frac{\mathbb{E}(Y^2 \mathbf{1}_{\{Y \geq y_n\}})}{m_Y^2(y_n)} - 1 \right) \rightarrow b_{22} \alpha_2^2 \text{ as } n \rightarrow \infty.$$

Similarly, and taking account of Lemma 2 and Lemma 7(i), one has:

$$\begin{aligned} \text{var}(T_{3,i,n}) &= \left(\sqrt{\bar{F}(y_n)} \sum_{j=1}^d \frac{\alpha_{3,j} \beta_j}{m_{X_{\cdot,j}}(y_n)} \right)^2 \left[\mathbb{E}(g(Y)^2 \mathbf{1}_{\{Y \geq y_n\}}) - \mathbb{E}(g(Y) \mathbf{1}_{\{Y \geq y_n\}})^2 \right] \\ &\sim \frac{(1-\gamma c)^2 \langle \alpha_3 \rangle^2}{g(y_n)^2 \bar{F}(y_n)} \left(\frac{g(y_n)^2 \bar{F}(y_n)}{1-2\gamma c} - \frac{g(y_n)^2 \bar{F}(y_n)^2}{(1-\gamma c)^2} (1+o(1)) \right) \\ &\rightarrow b_{33} \langle \alpha_3 \rangle^2 \text{ as } n \rightarrow \infty, \end{aligned}$$

$$\begin{aligned} \text{var}(T_{4,i,n}) &= \left(\sqrt{\bar{F}(y_n)} \sum_{j=1}^d \frac{\alpha_{4,j} \beta_j}{m_{X_{\cdot,j}Y}(y_n)} \right)^2 \left[\mathbb{E}(Y^2 g(Y)^2 \mathbf{1}_{\{Y \geq y_n\}}) - \mathbb{E}(Y g(Y) \mathbf{1}_{\{Y \geq y_n\}})^2 \right] \\ &\sim \frac{(1-\gamma(c+1))^2 \langle \alpha_4 \rangle^2}{y_n^2 g(y_n)^2 \bar{F}(y_n)} \left(\frac{(y_n g(y_n))^2 \bar{F}(y_n)}{1-2\gamma(c+1)} - \frac{(y_n g(y_n) \bar{F}(y_n))^2}{(1-\gamma(c+1))^2} (1+o(1)) \right) \\ &\rightarrow b_{44} \langle \alpha_4 \rangle^2 \text{ as } n \rightarrow \infty. \end{aligned}$$

The covariances are evaluated in a similar way. First, terms involving $T_{1,i,n}$ can be readily calculated:

$$\begin{aligned} \text{cov}(T_{1,i,n}, T_{2,i,n}) &= \text{cov} \left(\frac{\alpha_1}{\sqrt{\bar{F}(y_n)}} \mathbf{1}_{\{Y \geq y_n\}}, \sqrt{\bar{F}(y_n)} \frac{\alpha_2}{m_Y(y_n)} Y \mathbf{1}_{\{Y \geq y_n\}} \right) \\ &= \alpha_1 \alpha_2 (1 - \bar{F}(y_n)) \\ &\rightarrow \alpha_1 \alpha_2 \text{ as } n \rightarrow \infty, \end{aligned}$$

$$\begin{aligned}
\text{cov}(T_{1,i,n}, T_{3,i,n}) &= \text{cov} \left(\frac{\alpha_1}{\sqrt{\bar{F}(y_n)}} \mathbf{1}_{\{Y \geq y_n\}}, \sqrt{\bar{F}(y_n)} \left(\sum_{j=1}^d \frac{\alpha_{3,j} \beta_j}{m_{X_{\cdot,j}}(y_n)} \right) g(Y) \mathbf{1}_{\{Y \geq y_n\}} \right) \\
&\sim \frac{(1-\gamma c) \alpha_1 \langle \alpha_3 \rangle}{g(y_n) \bar{F}(y_n)} \mathbb{E}(g(Y) \mathbf{1}_{\{Y \geq y_n\}}) (1 - \bar{F}(y_n)) \\
&\rightarrow \alpha_1 \langle \alpha_3 \rangle \text{ as } n \rightarrow \infty,
\end{aligned}$$

$$\begin{aligned}
\text{cov}(T_{1,i,n}, T_{4,i,n}) &= \text{cov} \left(\frac{\alpha_1}{\sqrt{\bar{F}(y_n)}} \mathbf{1}_{\{Y \geq y_n\}}, \left(\sqrt{\bar{F}(y_n)} \sum_{j=1}^d \frac{\alpha_{4,j} \beta_j}{m_{X_{\cdot,j}Y}(y_n)} \right) Y g(Y) \mathbf{1}_{\{Y \geq y_n\}} \right) \\
&\sim \frac{(1-\gamma(c+1)) \alpha_1 \langle \alpha_4 \rangle}{y_n g(y_n) \bar{F}(y_n)} \mathbb{E}(Y g(Y) \mathbf{1}_{\{Y \geq y_n\}}) (1 - \bar{F}(y_n)) \\
&\rightarrow \alpha_1 \langle \alpha_4 \rangle \text{ as } n \rightarrow \infty.
\end{aligned}$$

Second, the remaining terms require repeated uses of Lemma 7(i):

$$\begin{aligned}
&\text{cov}(T_{2,i,n}, T_{3,i,n}) \\
&= \text{cov} \left(\sqrt{\bar{F}(y_n)} \frac{\alpha_2}{m_Y(y_n)} Y \mathbf{1}_{\{Y \geq y_n\}}, \left(\sqrt{\bar{F}(y_n)} \sum_{j=1}^d \frac{\alpha_{3,j} \beta_j}{m_{X_{\cdot,j}}(y_n)} \right) g(Y) \mathbf{1}_{\{Y \geq y_n\}} \right) \\
&\sim \frac{(1-\gamma)(1-\gamma c) \alpha_2 \langle \alpha_3 \rangle}{y_n g(y_n) \bar{F}(y_n)} [\mathbb{E}(Y g(Y) \mathbf{1}_{\{Y \geq y_n\}}) - \mathbb{E}(Y \mathbf{1}_{\{Y \geq y_n\}}) \mathbb{E}(g(Y) \mathbf{1}_{\{Y \geq y_n\}})] \\
&\sim \frac{(1-\gamma)(1-\gamma c) \alpha_2 \langle \alpha_3 \rangle}{y_n g(y_n) \bar{F}(y_n)} \left(\frac{y_n g(y_n) \bar{F}(y_n)}{1-\gamma(c+1)} - \frac{y_n g(y_n) \bar{F}(y_n)^2}{(1-\gamma)(1-\gamma c)} (1+o(1)) \right) \\
&\rightarrow b_{23} \alpha_2 \langle \alpha_3 \rangle \text{ as } n \rightarrow \infty,
\end{aligned}$$

$$\begin{aligned}
&\text{cov}(T_{2,i,n}, T_{4,i,n}) \\
&= \text{cov} \left(\sqrt{\bar{F}(y_n)} \frac{\alpha_2}{m_Y(y_n)} Y \mathbf{1}_{\{Y \geq y_n\}}, \left(\sqrt{\bar{F}(y_n)} \sum_{j=1}^d \frac{\alpha_{4,j} \beta_j}{m_{X_{\cdot,j}Y}(y_n)} \right) Y g(Y) \mathbf{1}_{\{Y \geq y_n\}} \right) \\
&\sim \frac{(1-\gamma(c+1))(1-\gamma) \alpha_2 \langle \alpha_4 \rangle}{y_n^2 g(y_n) \bar{F}(y_n)} [\mathbb{E}(Y^2 g(Y) \mathbf{1}_{\{Y \geq y_n\}}) - \mathbb{E}(Y \mathbf{1}_{\{Y \geq y_n\}}) \mathbb{E}(Y g(Y) \mathbf{1}_{\{Y \geq y_n\}})] \\
&\sim \frac{(1-\gamma(c+1))(1-\gamma) \alpha_2 \langle \alpha_4 \rangle}{y_n^2 g(y_n) \bar{F}(y_n)} \left(\frac{y_n^2 g(y_n) \bar{F}(y_n)}{1-\gamma(c+2)} - \frac{y_n^2 g(y_n) \bar{F}(y_n)^2}{(1-\gamma)(1-\gamma(c+1))} (1+o(1)) \right) \\
&\rightarrow b_{24} \alpha_2 \langle \alpha_4 \rangle \text{ as } n \rightarrow \infty,
\end{aligned}$$

$$\begin{aligned}
& \text{cov}(T_{3,i,n}, T_{4,i,n}) \\
&= \text{cov} \left(\sqrt{\bar{F}(y_n)} \sum_{j=1}^d \frac{\alpha_{3,j} \beta_j}{m_{X.,j}(y_n)} g(Y) \mathbf{1}_{\{Y \geq y_n\}}, \sqrt{\bar{F}(y_n)} \sum_{j=1}^d \frac{\alpha_{4,j} \beta_j}{m_{X.,jY}(y_n)} Y g(Y) \mathbf{1}_{\{Y \geq y_n\}} \right) \\
&\sim \frac{(1-\gamma c)(1-\gamma(c+1)) \langle \alpha_3 \rangle \langle \alpha_4 \rangle}{y_n g(y_n)^2 \bar{F}(y_n)} [\mathbb{E}(Y g(Y)^2 \mathbf{1}_{\{Y \geq y_n\}}) - \mathbb{E}(g(Y) \mathbf{1}_{\{Y \geq y_n\}}) \mathbb{E}(Y g(Y) \mathbf{1}_{\{Y \geq y_n\}})] \\
&\sim \frac{(1-\gamma c)(1-\gamma(c+1)) \langle \alpha_3 \rangle \langle \alpha_4 \rangle}{y_n g(y_n)^2 \bar{F}(y_n)} \left(\frac{y_n g(y_n)^2 \bar{F}(y_n)}{1-\gamma(2c+1)} - \frac{y_n g(y_n)^2 \bar{F}(y_n)^2}{(1-\gamma c)(1-\gamma(c+1))} (1+o(1)) \right) \\
&\rightarrow b_{34} \langle \alpha_3 \rangle \langle \alpha_4 \rangle \text{ as } n \rightarrow \infty.
\end{aligned}$$

Finally, it follows that, as $n \rightarrow \infty$,

$$\begin{aligned}
\text{var}(\chi'_{0,n}) &\rightarrow \alpha_1^2 + b_{22} \alpha_2^2 + b_{33} \langle \alpha_3 \rangle^2 + b_{44} \langle \alpha_4 \rangle^2 + 2\alpha_1 (\alpha_2 + \langle \alpha_3 \rangle + \langle \alpha_4 \rangle) \\
&\quad + 2\alpha_2 (b_{23} \langle \alpha_3 \rangle + b_{24} \langle \alpha_4 \rangle) + 2b_{34} \langle \alpha_3 \rangle \langle \alpha_4 \rangle \\
&= \alpha^t B \alpha,
\end{aligned}$$

where B is given in the statement of the Proposition. Remarking that $\chi'_{0,n}$ is the sum of a triangular array of independent, identically distributed and centered random variables, one may use Lyapunov criterion [5, Theorem 27.3], to prove its asymptotic normality. To this end, consider $\delta \in (0, \frac{1}{2(c+1)} - \gamma)$ and let us show that

$$n \mathbb{E} \left| \sum_{\ell=1}^4 \frac{T_{\ell,1,n} - \mathbb{E}(T_{\ell,1,n})}{\sqrt{n}} \right|^{2+\delta} \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (26)$$

Using both triangle and Jensen inequalities, we get

$$\begin{aligned}
\left\{ \mathbb{E} \left| \sum_{\ell=1}^4 [T_{\ell,1,n} - \mathbb{E}(T_{\ell,1,n})] \right|^{2+\delta} \right\}^{1/(2+\delta)} &\leq \sum_{\ell=1}^4 \left(\{\mathbb{E}|T_{\ell,1,n}|^{2+\delta}\}^{1/(2+\delta)} + \mathbb{E}|T_{\ell,1,n}| \right) \\
&\leq 8 \max_{1 \leq \ell \leq 4} \{\mathbb{E}|T_{\ell,1,n}|^{2+\delta}\}^{1/(2+\delta)}.
\end{aligned}$$

Lemma 2 and Lemma 7(i) yield, as $n \rightarrow \infty$,

$$\begin{aligned}
\mathbb{E}|T_{1,1,n}|^{2+\delta} &= \bar{F}(y_n)^{-\delta/2} |\alpha_1|^{2+\delta}, \\
\mathbb{E}|T_{2,1,n}|^{2+\delta} &\sim \bar{F}(y_n)^{-\delta/2} |\alpha_2|^{2+\delta} \frac{(1-\gamma)^{2+\delta}}{1-\gamma(2+\delta)}, \\
\mathbb{E}|T_{3,1,n}|^{2+\delta} &\sim \bar{F}(y_n)^{-\delta/2} |\langle \alpha_3 \rangle|^{2+\delta} \frac{(1-\gamma c)^{2+\delta}}{1-\gamma c(2+\delta)}, \\
\mathbb{E}|T_{4,1,n}|^{2+\delta} &\sim \bar{F}(y_n)^{-\delta/2} |\langle \alpha_4 \rangle|^{2+\delta} \frac{(1-\gamma(c+1))^{2+\delta}}{1-\gamma(2+\delta)(c+1)},
\end{aligned}$$

leading to

$$n\mathbb{E}\left|\sum_{\ell=1}^4 \frac{T_{\ell,1,n} - \mathbb{E}(T_{\ell,1,n})}{\sqrt{n}}\right|^{2+\delta} = O\left([n\bar{F}(y_n)]^{-\delta/2}\right) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

which proves (26). As a conclusion, $\chi'_{0,n} \xrightarrow{d} \mathcal{N}(0, \alpha^t B \alpha)$ and $\sqrt{n\bar{F}(y_n)}\Lambda_n \xrightarrow{d} \mathcal{N}(0, B)$. \square

Proof of Proposition 4. Let us denote by $\sigma_n^{-1} := \sqrt{n\bar{F}(y_n)}$ and prove in a first step that

$$\sigma_n^{-1} \left(\frac{\hat{v}_j(y_n) - v_j(y_n)}{\|v(y_n)\|} \right)_{1 \leq j \leq d} \xrightarrow{d} \xi(\beta_j)_{1 \leq j \leq d} \text{ with } \xi \sim \mathcal{N}(0, \lambda(c, \gamma)), \quad (27)$$

using the notations introduced in Proposition 3. Denoting by $\vartheta_n := \sigma_n^{-1}\Lambda_n$, Proposition 3 shows that, for all $j \in \{1, \dots, d\}$:

$$\begin{aligned} \hat{F}(y_n) &= \bar{F}(y_n)(1 + \sigma_n\vartheta_{1,n}), & \hat{m}_Y(y_n) &= m_Y(y_n)(1 + \sigma_n\vartheta_{2,n}), \\ \hat{m}_{X_{\cdot,j}}(y_n) &= m_{X_{\cdot,j}}(y_n)(1 + \sigma_n\vartheta_{j+2,n}), & \hat{m}_{X_{\cdot,j}Y}(y_n) &= m_{X_{\cdot,j}Y}(y_n)(1 + \sigma_n\vartheta_{j+2+d,n}). \end{aligned}$$

with $\vartheta_n \xrightarrow{d} \mathcal{N}(0, B)$. Substituting in (3), we get

$$\begin{aligned} \hat{v}_j(y_n) &= \hat{F}(y_n)\hat{m}_{X_{\cdot,j}Y}(y_n) - \hat{m}_{X_{\cdot,j}}(y_n)\hat{m}_Y(y_n) \\ &= \bar{F}(y_n)m_{X_{\cdot,j}Y}(y_n) - m_{X_{\cdot,j}}(y_n)m_Y(y_n) \\ &+ \bar{F}(y_n)m_{X_{\cdot,j}Y}(y_n)\sigma_n[\vartheta_{1,n} + \vartheta_{j+2+d,n} + \sigma_n\vartheta_{1,n}\vartheta_{j+2+d,n}] \\ &- m_{X_{\cdot,j}}(y_n)m_Y(y_n)\sigma_n[\vartheta_{2,n} + \vartheta_{j+2,n} + \sigma_n\vartheta_{2,n}\vartheta_{j+2,n}], \end{aligned}$$

and taking account of $v_j(y_n) = \bar{F}(y_n)m_{X_{\cdot,j}Y}(y_n) - m_{X_{\cdot,j}}(y_n)m_Y(y_n)$ and $\sigma_n \rightarrow 0$ yields

$$\begin{aligned} \sigma_n^{-1}(\hat{v}_j(y_n) - v_j(y_n)) &= \bar{F}(y_n)m_{X_{\cdot,j}Y}(y_n)[\vartheta_{1,n} + \vartheta_{j+2+d,n} + o_P(1)] \\ &- m_{X_{\cdot,j}}(y_n)m_Y(y_n)[\vartheta_{2,n} + \vartheta_{j+2,n} + o_P(1)]. \end{aligned}$$

From (23)–(25) in the proof of Proposition 3, it follows that, for all $j \in \{1, \dots, D\}$:

$$\begin{aligned} \frac{\sigma_n^{-1}}{y_n g(y_n) \bar{F}(y_n)^2} (\hat{v}_j(y_n) - v_j(y_n)) &= \frac{\beta_j}{1 - \gamma(c+1)} [\vartheta_{1,n} + \vartheta_{j+2+d,n} + o_P(1)] \\ &- \frac{\beta_j}{(1 - \gamma)(1 - \gamma c)} [\vartheta_{2,n} + \vartheta_{j+2,n} + o_P(1)]. \end{aligned} \quad (28)$$

Besides, Lemma 4 yields

$$\|v(y_n)\| = \bar{F}(y_n)\mathbb{E}[g(Y)\Psi_{y_n}(Y)]\|\beta + \eta(y_n)\|,$$

with $\mathbb{E}[g(Y)\Psi_{y_n}(Y)] \sim \lambda_2(\gamma, c) y_n g(y_n) \bar{F}(y_n)$ and $\|\beta + \eta(y_n)\| \rightarrow \|\beta\| = 1$ as $y_n \rightarrow \infty$, from (20) and (21) in the proof of Proposition 2. It follows that

$$\|v(y_n)\| = \lambda_2(\gamma, c) y_n g(y_n) \bar{F}^2(y_n) (1 + o(1)). \quad (29)$$

Collecting (28) and (29) entails, for all $j = 1, \dots, d$:

$$\frac{\sigma_n^{-1}}{\|v(y_n)\|} (\hat{v}_j(y_n) - v_j(y_n)) = \beta_j [a_1(\vartheta_{1,n} + \vartheta_{j+2+d,n}) - a_2(\vartheta_{2,n} + \vartheta_{j+2,n})] + o_P(1),$$

where we have defined $a_1 = (1-\gamma)(1-\gamma c)/(\gamma^2 c)$ and $a_2 = (1-\gamma(c+1))/(\gamma^2 c)$. As a consequence, we have proved that

$$\frac{\sigma_n^{-1}}{\|v(y_n)\|} (\hat{v}_j(y_n) - v_j(y_n))_{1 \leq j \leq d} \xrightarrow{d} \mathcal{N}(0, \text{diag}(\beta_1, \dots, \beta_d) A B A^t \text{diag}(\beta_1, \dots, \beta_d)), \quad (30)$$

where A is the $d \times 2(d+1)$ matrix defined as follows

$$A = \left(\begin{array}{c|c|cccccc|cccc} a_1 & -a_2 & -a_2 & 0 & \dots & \dots & 0 & a_1 & 0 & \dots & \dots & 0 \\ a_1 & -a_2 & 0 & -a_2 & \ddots & & 0 & 0 & a_1 & \ddots & & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & & \ddots & -a_2 & 0 & \vdots & & \ddots & a_1 & 0 \\ a_1 & -a_2 & 0 & \dots & \dots & 0 & -a_2 & 0 & \dots & \dots & 0 & a_1 \end{array} \right).$$

Straightforward algebra shows that

$$\text{diag}(\beta_1, \dots, \beta_d) A B A^t \text{diag}(\beta_1, \dots, \beta_d) = \lambda(c, \gamma) (\beta_1, \dots, \beta_d)^t (\beta_1, \dots, \beta_d),$$

so that the limiting Gaussian distribution is non-degenerated in the only direction $(\beta_1, \dots, \beta_d)^t$. Therefore, the convergence in distribution (30) can be rewritten as in (27). The second step consists in proving that

$$\sigma_n^{-1} \left(\frac{\hat{v}_j(y_n) - v_j(y_n)}{\|v(y_n)\|} \right)_{d+1 \leq j \leq p} \xrightarrow{\mathbb{P}} 0. \quad (31)$$

For all $j = d+1, \dots, p$ and $\zeta \in \{0, 1\}$, the inverse model (\mathbf{M}_1) shows that

$$\hat{m}_{X_{\cdot, j} Y^\zeta}(y_n) = \frac{1}{n} \sum_{i=1}^n X_{ij} Y_i^\zeta \mathbf{1}_{\{Y_i^\zeta \geq y_n\}} = \frac{1}{n} \sum_{i=1}^n \varepsilon_{ij} Y_i^\zeta \mathbf{1}_{\{Y_i^\zeta \geq y_n\}}.$$

Moreover, from (16) in the proof of Lemma 8, we have

$$\text{var}(\hat{m}_{X_{\cdot, j} Y^\zeta}(y_n)) = O\left(\frac{y_n^{2\zeta} \bar{F}(y_n)^{1-2/q}}{n}\right),$$

and recalling that $\mathbb{E}\left(\hat{m}_{X_{\cdot,j}Y^c}(y_n)\right) = m_{X_{\cdot,j}Y^c}(y_n)$, a straightforward application of Markov inequality yields

$$\hat{m}_{X_{\cdot,j}Y^c}(y_n) = m_{X_{\cdot,j}Y^c}(y_n) + O_P\left(\frac{y_n^\zeta \bar{F}(y_n)^{\frac{1}{2}-\frac{1}{q}}}{\sqrt{n}}\right).$$

Substituting in (3) and taking into account Lemma 7(ii) yield

$$\begin{aligned} \hat{v}_j(y_n) &= \hat{F}(y_n)\hat{m}_{X_{\cdot,j}Y}(y_n) - \hat{m}_{X_{\cdot,j}}(y_n)\hat{m}_Y(y_n) \\ &= \bar{F}(y_n)(1 + \sigma_n\vartheta_{1,n})\left(m_{X_{\cdot,j}Y}(y_n) + O_P\left(\frac{y_n\bar{F}(y_n)^{\frac{1}{2}-\frac{1}{q}}}{\sqrt{n}}\right)\right) \\ &\quad - m_Y(y_n)(1 + \sigma_n\vartheta_{2,n})\left(m_{X_{\cdot,j}}(y_n) + O_P\left(\frac{\bar{F}(y_n)^{\frac{1}{2}-\frac{1}{q}}}{\sqrt{n}}\right)\right) \\ &= v_j(y_n) + O_P\left(\frac{y_n\bar{F}(y_n)^{\frac{3}{2}-\frac{1}{q}}}{\sqrt{n}}\right). \end{aligned}$$

Therefore, in view of (29), we have

$$\frac{\sigma_n^{-1}}{\|v(y_n)\|}(\hat{v}_j(y_n) - v_j(y_n)) = O_P\left(\frac{1}{g(y_n)\bar{F}(y_n)^{1/q}}\right),$$

for all $j \in \{d+1, \dots, p\}$, as $n \rightarrow \infty$ and since $\bar{F}(\cdot)^{1/q}g(\cdot)$ is regularly-varying with index $c - 1/(q\gamma) > 0$. Finally, we can then infer that (31) holds, hence the result. \square

Proof of Theorem 1. Let us recall that $\sigma_n^{-1} = \sqrt{n\bar{F}(y_n)}$ and consider the expansion

$$\sigma_n^{-1}\left(\frac{\hat{v}(y_n)}{\|v(y_n)\|} - \beta\right) = \sigma_n^{-1}\left(\frac{\hat{v}(y_n)}{\|v(y_n)\|} - w(y_n)\right) + \sigma_n^{-1}(w(y_n) - \beta).$$

First, Proposition 4 shows that

$$\sigma_n^{-1}\left(\frac{\hat{v}(y_n)}{\|v(y_n)\|} - w(y_n)\right) \xrightarrow{d} \xi\beta,$$

where $\xi \sim \mathcal{N}(0, \lambda(c, \gamma))$. Second, Proposition 2 entails that

$$\sigma_n^{-2}\|w(y_n) - \beta\|^2 = O\left(\frac{n\bar{F}^{1-2/q}(y_n)}{g^2(y_n)}\right) \rightarrow 0,$$

as $y_n \rightarrow \infty$, and the result is proved. \square

References

- [1] Ahmad, A., Deme, E., Diop, A., Girard, S., & Usseglio Carleve, A. (2020). Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model. *Electronic Journal of Statistics*, 14, 4421–4456.
- [2] Beirlant, J., & Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis*, 89(1), 97–118.
- [3] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- [4] Bernard-Michel, C., Gardes, L., & Girard, S. (2009). Gaussian Regularized Sliced Inverse Regression, *Statistics and Computing*, 19, 85–98.
- [5] Billingsley, P. (1995). *Measure and Probability (third edition)*. John Wiley & Sons.
- [6] Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1989). *Regular variation* (Vol. 27). Cambridge University Press.
- [7] Chun, H., & Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B*, 72(1), 3–25.
- [8] Chavez-Demoulin, V., & Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C*, 54(1), 207–222.
- [9] Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics*, 33(2), 806–839.
- [10] Chiancone, A., Forbes, F., & Girard, S. (2017). Student sliced inverse regression. *Computational Statistics & Data Analysis*, 113, 441–456.
- [11] Cook, R.D. (2007). Fisher lecture: Dimension reduction in regression. *Statistical Science*, 22(1), 1–26.
- [12] Cook, R. D., & Forzani, L. (2018). Big data and partial least-squares prediction. *Canadian Journal of Statistics*, 46(1), 62–78.
- [13] Cook, R. D., Helland, I. S., & Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B*, 75(5), 851–877.

- [14] Coudret, R., Girard, S., & Saracco, J. (2014). A new sliced inverse regression method for multivariate response. *Computational Statistics & Data Analysis*, 77, 285–299.
- [15] Daouia, A., Gardes, L., Girard, S., & Lekina, A. Kernel estimators of extreme level curves, *Test*, 20(2), 311–333.
- [16] Daouia, A., Gardes, L., & Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B), 2557–2589.
- [17] Davison, A. C., & Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society: Series B*, 62(1), 191–208.
- [18] Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B*, 52(3), 393–425.
- [19] Drees, H., & Sabourin, A. (2019). Principal component analysis for multivariate extremes. *arXiv preprint arXiv:1906.11043*.
- [20] Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2), 109–135.
- [21] Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*, Springer Series in Statistics, New-York.
- [22] Gardes, L. (2018). Tail Dimension Reduction for extreme quantile estimation, *Extremes*, 21(1), 57–95.
- [23] Gardes, L., & Girard, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2), 177–204.
- [24] Gardes, L., & Girard, S. (2012). Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics*, 6, 1715–1744.
- [25] Goegebeur, Y., Guillou, A., & Schorgen, A. (2014). Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics*, 48(4), 732–755.
- [26] de Haan, L., & Ferreira, A. (2007). *Extreme value theory: An introduction*. Springer Science & Business Media.
- [27] Hall, P., & Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 153–167.
- [28] Hardle, W., & Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408), 986–995.

- [29] Helland, I. S. (1990). Partial least squares regression and statistical models. *Scandinavian Journal of Statistics*, 17(2), 97–114.
- [30] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3, 1163–1174.
- [31] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1), 69–82.
- [32] Horowitz, J. L. (2009). *Single-Index Models*. Semiparametric and Nonparametric Methods in Econometrics, Springer, New-York.
- [33] Hsing, T., & Carroll, R. J. (1992). An asymptotic theory for sliced inverse regression. *The Annals of Statistics*, 20(2), 1040–1061.
- [34] Jagger, T. H., & Elsner, J. B. (2009). Modeling tropical cyclone intensity with quantile regression. *International Journal of Climatology*, 29(10), 1351–1361.
- [35] Kong, E., & Xia, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory*, 28, 730–768.
- [36] Kyung-Joon, C. & Schucany, W. R. (1998). Nonparametric kernel regression estimation near endpoints. *Journal of Statistical Planning and Inference*, 66, 289–304.
- [37] Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414), 316–327.
- [38] Li, L., Cook, R. D., & Tsai, C. L. (2007). Partial inverse regression. *Biometrika*, 94(3), 615–625.
- [39] Martens, H., & Næs, T. (1992). *Multivariate calibration*. John Wiley & Sons.
- [40] Meligkotsidou, L., Vrontos, I. D., & Vrontos, S. D. (2009). Quantile regression analysis of hedge fund strategies. *Journal of Empirical Finance*, 16(2), 264–279.
- [41] Naik, P., & Tsai, C. L. (2000). Partial least squares estimator for single-index models. *Journal of the Royal Statistical Society: Series B*, 62(4), 763–771.
- [42] Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.
- [43] Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 1403–1430.

- [44] Saracco, J. (1997). An asymptotic theory for sliced inverse regression. *Communications in statistics-Theory and methods*, 26(9), 2141–2171.
- [45] Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, 367–377.
- [46] Smith, V. H., & Goodwin, B. K. (1996). Crop insurance, moral hazard, and agricultural chemical use. *American Journal of Agricultural Economics*, 78(2), 428–438.
- [47] Stone, M., & Brooks, R. J. (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society: Series B*, 52(2), 237–258.
- [48] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288.
- [49] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B*, 67(1), 91–108.
- [50] Wang, H. J., & Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, 108(503), 1062–1074.
- [51] Wang, H. J., Li, D., & Xu, W. (2020). Extreme quantile estimation based on the tail single-index model. *Statistica Sinica*, to appear, 10.5705/ss.202020.0051
- [52] Wang, L., & Yang, L. (2009). Spline estimation of single-index models (Vol. 19). *Statistica Sinica*, 765–783.
- [53] Wold, H. (1975). Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. *Journal of Applied Probability*, 12(S1), 117–142.
- [54] Wu, H. M. (2008). Kernel sliced inverse regression with applications to classification. *Journal of Computational and Graphical Statistics*, 17(3), 590–610.
- [55] Wu, T. Z., Yu, K., & Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis*, 101(7), 1607–1621.
- [56] Yu, Y., & Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97(460), 1042–1054.
- [57] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2), 301–320.