Extreme Partial Least-Squares

Meryem Bousebata[a,b], Geoffroy Enjolras[b], Stéphane Girard[⋆,a]

[a] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.*
[b] *Univ. Grenoble Alpes, CERAG, 38000 Grenoble, France.*

Abstract

We propose a new approach, called Extreme-PLS, for dimension reduction in conditional extreme values settings. The objective is to find linear combinations of covariates 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.
2020 MSC: 62G32, 62H25, 62H12, 62E20.

1. Introduction

One of the main goals of statistical analysis is to seek a relationship between a response variable Y and a p-dimensional vector X of covariates starting from a n-sample. A common way to describe the possible link is to use the regression function $\mathbb{E}(Y|X)$. However, in many situations, the entire conditional distribution of Y given X may be of interest, rather than the only central part. This has led to the development of conditional quantiles, or regression quantiles, as an alternative to the conditional mean. Quantile regression was introduced by [42] in a parametric framework. Since then, non-parametric regression methods have been considered in the literature. Among others, [5] studied kernel and nearest neighbour estimators of conditional quantiles, while [36] focused on spline methods.

A complementary 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. One motivating example in agricultural risk management is the study of the influence of meteorological parameters (temperature, humidity,...), agricultural inputs (pesticides, fertilisers,...) and risk management tools (insurance premiums, subsidies,...) on low values of crop yields (see Section 6). Other motivations can be found in finance [49], climatology [41], hydrology [27] and environmental sciences [54], to name a few. In these applications, the estimation of extreme conditional quantiles is a crucial issue that has been studied from various points of view. A first approach is based on a parametric modelling, see for instance [11] 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 [10, 22, 54] using extreme-value techniques to model exceedances above a high threshold. A second line of work relies on non-parametric approaches that can be split into three main categories: fixed design, random design, and functional covariates. Fixed design methods aim at estimating conditional extreme quantiles depending on a non-random covariate, see [27] for a nearest neighbors technique. In the random design setting, one can cite [20, 31] who studied the estimation of extreme quantiles under a conditional heavy-tail model, later extended in [19] to conditional distributions belonging to any maximum domain of attraction. Finally, see [28] for the functional covariate situation. Semi-parametric approaches have also been considered for trend modelling in extreme values. Local polynomial fitting of extreme-value models is investigated in [2, 21], a non-parametric estimation of the temporal trend is combined with parametric models for extreme values in [33], and a location-dispersion regression model for heavy-tailed distributions is introduced in [1].

[∗]Corresponding author. Email address: `stephane.girard@inria.fr`

In a high dimensional context, *i.e.*, when the dimension p of X is large, the above-mentioned estimation methods may suffer from the so-called "curse of dimensionality". This phenomenon results in an exploding variance of the estimators, thus impeding the inference in practice. A number of statistical approaches to dimension reduction have been introduced to circumvent this issue and to exhibit the most relevant directions in the high-dimensional covariate space. One of the most popular ones is Partial least squares (PLS), introduced by [60], which combines characteristics of Principal component analysis (PCA) for dimension reduction and multiple regression, see also [47] for an interpretation as a Krylov space method. The development of PLS has been initiated within the chemometrics field, see the reference book [48]. Since then, PLS has also received attention in the statistical literature. For example, [37] discusses the statistical properties of the PLS procedure under a factor analysis model, while [25] provides a comparison between PLS and Principal component regression (PCR) from various perspectives. See also [16] for a connection between PLS approach and envelopes and [13] for a sparse version of PLS. The basic idea of PLS is to seek directions, *i.e.*, linear combinations of X coordinates having both high variance and high correlation with Y, unlike the PCR method which only takes into account high variance components [25, 56]. Sliced inverse regression (SIR) [45] is an alternative method that takes advantage 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 [29] such as Partial inverse regression handling the $p > n$ situation [46], Kernel sliced inverse regression allowing the estimation of a nonlinear subspace [61], Student sliced inverse regression dealing with heavy-tailed errors [12], Sliced inverse regression for multivariate response [18], 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 [40, 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 ones are the average derivative estimation method in the context of kernel smoothing [34, 52], and the M-estimation technique based on spline regression [59, 64]. One can also mention [43, 62] who considered single-index models to estimate conditional quantiles. In [50], it is proved that PLS provides a consistent estimator of the direction when Y given X follows a single-index model and when $n \to \infty$ with a fixed dimension p, under the additional assumption of independence between noise and covariates. From the practical point of view, PLS can perform better than SIR even though the link function is non-linear. This result is extended in [13] to the multiple-index situation and when both $n \to \infty$ and $p \to \infty$ with $p/n \to 0$. It is also shown that, in contrast, the PLS estimator is no more consistent in the case where $p/n \to k > 0$, and a sparse version of PLS is introduced to avoid this vexing effect. More recently, [15] tempered this conclusion by exhibiting some designs under which single-index PLS is consistent in the linear regression situation, when both $n \to \infty$ and $p \to \infty$, regardless of the alignment between n and p.

Finally, the curse of dimensionality may also be tackled using shrinkage methods which aim at reducing the complexity of the inference by variable selection. As an example, Lasso method [57] penalizes regression coefficients similarly to Ridge regression [39] but replacing the L_2 penalty by the L_1 counterpart. Some extensions include Fused lasso [58] and Elastic net [66] to deal with the case where p is larger than n. Many other shrinkage and variable selection methods are discussed in [35, Chapter 3].

Dimension reduction dedicated to conditional extremes is limited in the literature, and only a few recent works have been devoted to it. One can mention [26] where a dimension reduction framework adapted to conditional tail distributions is developed assuming that, when Y is large, X and Y are independent conditionally on a linear combination of the covariates. Another approach [17, 23] consists in adapting the (unsupervised) dimension reduction method PCA to the extreme setting. In [63], a semi-parametric approach is introduced for the estimation of extreme conditional quantiles based on a tail single-index model. The authors propose to estimate the dimension reduction direction β using local linear quantile regression. The method is developed under the tail single-index model and a conditional mean linearity assumption, which is satisfied, for instance, when X is elliptically distributed (the method is described in further detail in Section 5).

We introduce a new approach, referred to as extreme-PLS (EPLS), for dimension reduction in an extreme conditional setting. The underlying idea is to look 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. 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 properties of the EPLS estimator are established under an inverse single-index model and a heavy-tail assumption, without recourse to linearity as in [63] nor independence assumptions as in [26]. It appears on simulated data that the EPLS estimator provides promising results in high-dimensional settings and outperforms the estimator proposed in [63] in a wide range of situations.

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 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 method's performance is 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. A Supplementary material is also provided to complete the simulation study. Data and R code are available at `https://github.com/meryembst/EPLS`.

2. Single-index EPLS approach

Let Y be a real random response variable and X a p-dimensional random covariate. We denote by $w(y)$ the unit vector maximizing the covariance between $w^t X$ and Y given that Y exceeds a large threshold y:

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

where $\|\cdot\|$ denotes the Euclidean norm. 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. For all $y \in \mathbb{R}$, introduce $\bar{F}(y) = \mathbb{P}(Y \geq y)$ the survival function of Y and consider the three tail-moments, whenever they exist, $m_Y(y) = \mathbb{E}(Y \mathbb{1}_{\{Y \geq y\}}) \in \mathbb{R}$, $m_X(y) = \mathbb{E}(X \mathbb{1}_{\{Y \geq y\}}) \in \mathbb{R}^p$, $m_{XY}(y) = \mathbb{E}(XY \mathbb{1}_{\{Y \geq y\}}) \in \mathbb{R}^p$.

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

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

Let us note that solution (2) is invariant with respect to the scaling and location of X. In the following, we aim at investigating the behaviour of $w(y)$ for large thresholds, *i.e.*, as $y \to \infty$. To this end, consider the following single-index non linear inverse regression model:

(M_1) $X = g(Y)\beta + \varepsilon$, where X and ε are p-dimensional random vectors, Y is a real random variable, $g : \mathbb{R} \to \mathbb{R}$ is an unknown link function, $\beta \in \mathbb{R}^p$ is an unknown unit vector.

Let us highlight that no independence assumption is made on the pair (X, ε). However, in the particular case where 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 $y \in \mathbb{R}$. Model (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, 14]. Under model (M_1), 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 $Y \simeq g^{-1}(\beta^t X)$. Our first goal is to establish the convergence of $w(y)$ towards β, as $y \to \infty$, without resorting neither to a linear conditional expectation assumption as in [63] nor to a conditional independence assumption as in [26, 53]. In contrast, additional assumptions on the link function g and the distribution tail of Y are considered:

(A_1) 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 $t > 0$,

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

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

(A$_2$) $g \in RV_c$ with $c > 0$.

(A$_3$) There exists $q > 1/(\gamma c)$ such that $\mathbb{E}(\|\varepsilon\|^q) < \infty$.

Let us note that (A$_1$) implies that $\bar{F} \in RV_{-1/\gamma}$ in view of Karamata's theorem [7, Theorem 1.5.8]. In other words, (A$_1$) 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 [32, 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 $\mathbb{E}(|Y|\mathbb{1}_{\{Y \geq y\}})$ exists for all $y \in \mathbb{R}$. Assumption (A$_2$) means that the link function ultimately behaves like a power function. Finally, (A$_3$) 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 (A$_3$) thus imposes that $\gamma_{g(Y)} > \gamma_{\|\varepsilon\|}$, meaning that $g(Y)$ has an heavier right tail than $\|\varepsilon\|$. In model (M$_1$), 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 $y \to \infty$, we let

$$\Delta(w(y), \beta) := 1 - \cos^2(w(y), \beta) = 1 - (w(y)^t \beta)^2. \tag{3}$$

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 (M$_1$), (A$_1$) , (A$_2$) and (A$_3$) hold with $\gamma(c+1) < 1$. Then, as $y \to \infty$,

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

It should be noted that, since $\|w(y) - \beta\| \to 0$ as $y \to \infty$, the EPLS axis has asymptotically the same direction as β (without sign issue). Besides, in view of assumptions (A$_1$) and (A$_2$), the function $y \mapsto g(y)\bar{F}^{1/q}(y)$ is regularly-varying with index $c - 1/(q\gamma) > 0$ from (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 inference from data distributed from model (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 (M$_1$) and let $y_n \to \infty$ as the sample size n tends to infinity. The solution (2) is estimated by its empirical counterpart introducing

$$\hat{v}(y_n) = \hat{\bar{F}}(y_n)\hat{m}_{XY}(y_n) - \hat{m}_X(y_n)\hat{m}_Y(y_n), \tag{4}$$

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

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

For all $j \in \{1, \ldots, p\}$, let us denote by $X_{.,j}$ the jth coordinate of X. It readily follows that the jth coordinates of $m_X(y_n)$ and $m_{XY}(y_n)$ are respectively given by $m_{X_{.,j}}(y_n)$ and $m_{X_{.,j}Y}(y_n)$. Let us note that $\hat{m}_Y(y_n)$ can be interpreted as the expected shortfall [24] of Y while $\hat{m}_X(y_n)$ corresponds to the marginal expected shortfall [8] of X given Y.

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 $y_n \to \infty$ and $n\bar{F}(y_n) \to \infty$. 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 (M_1), (A_1), (A_2) and (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 $j \in \{1, \dots, d\}$ and $\beta_{d+1} = \cdots = \beta_p = 0$. Let $y_n \to \infty$ such that $n\bar{F}(y_n) \to \infty$ and introduce the $\mathbb{R}^{2(d+1)}-$ random vector

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

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

$$B = \left(\begin{array}{c|c|ccc|ccc}
1 & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ \hline
1 & b_{22} & b_{23} & \dots & b_{23} & b_{24} & \dots & b_{24} \\ \hline
1 & b_{23} & b_{33} & \dots & b_{33} & b_{34} & \dots & b_{34} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & b_{23} & b_{33} & \dots & b_{33} & b_{34} & \dots & b_{34} \\ \hline
1 & b_{24} & b_{34} & \dots & b_{34} & b_{44} & \dots & b_{44} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & b_{24} & b_{34} & \dots & b_{34} & b_{44} & \dots & b_{44}
\end{array} \right)$$

and where

$$b_{22} = \frac{(1-\gamma)^2}{1-2\gamma}, \qquad b_{33} = \frac{(1-\gamma c)^2}{1-2\gamma c}, \qquad b_{44} = \frac{(1-\gamma(c+1))^2}{1-2\gamma(c+1)},$$

$$b_{23} = \frac{(1-\gamma)(1-\gamma c)}{1-\gamma(c+1)}, \qquad b_{24} = \frac{(1-\gamma(c+1))(1-\gamma)}{1-\gamma(c+2)}, \qquad b_{34} = \frac{(1-\gamma c)(1-\gamma(c+1))}{1-\gamma(2c+1)}.$$

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 (M_1), (A_1), (A_2) and (A_3) hold with $2\gamma(c+1) < 1$. Let $y_n \to \infty$ such that $n\bar{F}(y_n) \to \infty$. 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

$$\lambda(c,\gamma) = a_1^2(3 + b_{44}) + a_2^2(2b_{23} + b_{22} + b_{33}) - 2a_1a_2(2 + b_{24} + b_{34}), \tag{5}$$

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

The asymptotic variance $\lambda(c, \gamma)$ is plotted on Fig. 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 [53] 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 (M_1), (A_1), (A_2) and (A_3) hold with $2\gamma(c+1) < 1$. Let $y_n \to \infty$ such that $n\bar{F}(y_n) \to \infty$ and $n\bar{F}(y_n)^{1-2/q}/g^2(y_n) \to 0$ as $n \to \infty$. Then,

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

with $\xi \sim \mathcal{N}(0, \lambda(c, \gamma))$ and where $\lambda(c, \gamma)$ is defined in (5).

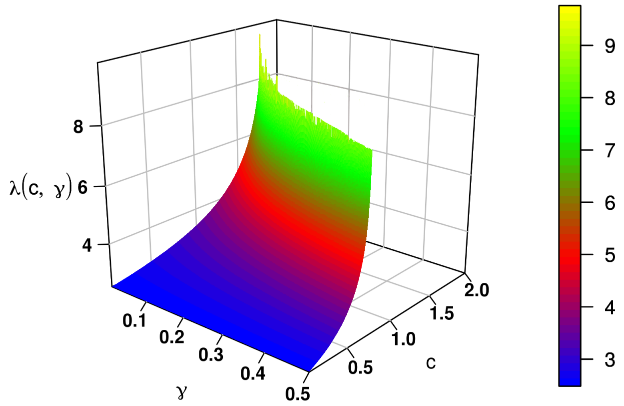

Fig. 1: Asymptotic variance $(c, \gamma) \in [1/2, 2] \times [0, 1/2] \mapsto \lambda(c, \gamma)$ given in Proposition 4, Equation (5), on a logarithmic scale.

Assumption $n\bar{F}(y_n) \to \infty$ ensures that the variance of the estimator tends to zero while condition $n\bar{F}(y_n)^{1-2/q}/g^2(y_n) \to 0$ 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 (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 (M_1) can be extended to a multi-index setting by considering, for some $K \geq 1$:

(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, $g_\ell : \mathbb{R} \to \mathbb{R}$ are unknown link functions, $\beta^{(\ell)} \in \mathbb{R}^p$ are unknown orthogonal unit vectors.

Denoting by $\mathcal{Y}$ a set of candidate values for y_n, the following iterative procedure is considered to estimate $\beta^{(1)}, \dots, \beta^{(K)}$:

1. Initialization: Set $R_i^{(0)} := X_i$ for all $i \in \{1, \dots, n\}$.
2. For $\ell \in \{1, \dots, p\}$,
   - Estimation of the ℓth direction for all $y_n \in \mathcal{Y}$:

$$\hat{v}^{(\ell)}(y_n) = \hat{\bar{F}}(y_n)\hat{m}_{R^{(\ell-1)}Y}(y_n) - \hat{m}_{R^{(\ell-1)}}(y_n)\hat{m}_Y(y_n).$$

   - Computation of the threshold maximizing the conditional correlation:

$$y^{(\ell)} = \arg\max_{y_n \in \mathcal{Y}} \rho\left(\left(R^{(\ell-1)}\right)^t \hat{v}^{(\ell)}(y_n), Y | Y \geq y_n\right) = \arg\max_{y_n \in \mathcal{Y}} \frac{\operatorname{cov}\left(\left(R^{(\ell-1)}\right)^t \hat{v}^{(\ell)}(y_n), Y | Y \geq y_n\right)}{\sigma\left((R^{(\ell-1)})^t \hat{v}^{(\ell)}(y_n) | Y \geq y_n\right) \sigma\left(Y | Y \geq y_n\right)}, \tag{6}$$

     and recording of the optimal value: $\Xi_\ell = \rho\left(\left(R^{(\ell-1)}\right)^t \hat{v}^{(\ell)}(y^{(\ell)}), Y | Y \geq y^{(\ell)}\right)$.
   - Update of the residuals: for all $i \in \{1, \dots, n\}$:

$$R_i^{(\ell)} := R_i^{(\ell-1)} - \frac{\hat{v}^{(\ell)}(y^{(\ell)}) \left(\hat{v}^{(\ell)}(y^{(\ell)})\right)^t}{\|\hat{v}^{(\ell)}(y^{(\ell)})\|^2} R_i^{(\ell-1)}.$$

The idea is the following. At the first iteration, $\hat{v}^{(1)}(y_n)$, $y_n \in \mathcal{Y}$, corresponds to the estimator associated with the single-index model computed by (4). Then, the threshold $y^{(1)}$ maximizing w.r.t. $y_n \in \mathcal{Y}$ the correlation between the projected covariate $X^t\hat{v}^{(1)}(y_n)$ and the response variable Y given $Y \geq y_n$ is computed and the maximum correlation $\Xi^{(1)}$ is recorded. From model (M_K), in view of the orthogonality of the directions, one has $X^t\hat{v}^{(1)}(y^{(1)}) \simeq \|\hat{v}^{(1)}(y^{(1)})\|X^t\beta^{(1)} \simeq \|\hat{v}^{(1)}(y^{(1)})\|g_1(Y)$ and thus the residual

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

approximately satisfies the same inverse regression model with $K-1$ directions. It is then natural to iterate the process and estimate $\beta^{(2)}$ from (4) computed on the residuals $R^{(1)}$. Moreover, since these residuals are by construction orthogonal to $\hat{v}^{(1)}(y^{(1)})$, one necessarily has $\hat{v}^{(2)}(y^{(2)}) \perp \hat{v}^{(1)}(y^{(1)})$. Thanks to the above orthogonality property, the estimated number of directions can be upper bounded by p. We refer to [37] for a similar result on the original PLS method. The estimation of the number K of directions can be achieved by a visual inspection of the scree plot $\ell \in \{1, \ldots, p\} \mapsto \Xi(\ell)$. The estimated $\hat{K}$ is defined as an elbow in the above graph, which is detected using Cattell's method [9], see Fig. 4 for an illustration on the real data experiment (Section 6).

5. Validation on simulations

Let us consider a sample of size $n = 1000$ and dimension p from model (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 [63] referred to as SIMEXQ (single-index model extreme quantile) in the sequel. SIMEXQ method is an extension of the global single-index quantile regression model developed in [65] where β is estimated by the slope obtained by fitting a misspecified linear quantile regression model to the data. In the SIMEXQ methodology, it is shown that β can be similarly estimated under the weaker assumption of a tail single-index model and a conditional mean linearity assumption. In practice, it is sufficient to narrow the fit of the misspecified linear quantile regression model to the exceedances.

Two heavy-tailed distributions 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.

Let us stress that, in both cases, the tail-index of Y is $\gamma = 1/5$ and does not depend on the covariate. Two dimensions of the covariate are considered:

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

Each coordinate $\varepsilon^{(j)}$, $j \in \{1, \ldots, p\}$ of the error ε is simulated from the $\mathcal{N}(0, \sigma^2)$ distribution and depending on Y using a copula. Two copula models are investigated:

- the Frank copula defined for all $(u_1, u_2) \in [0, 1]^2$ by

$$C_\theta(u_1, u_2) = -\frac{1}{\theta}\log\left(1 + \frac{(\mathrm{e}^{-\theta u_1} - 1)(\mathrm{e}^{-\theta u_2} - 1)}{\mathrm{e}^{-\theta} - 1}\right),$$

  where $\theta \in \mathbb{R}$ is a parameter tuning the dependence between the margins. Frank copula is an Archimedean copula, see [51, Section 4.2], able to model the full range of dependence: $\theta \to -\infty$ yields the counter-monotonicity copula, $\theta \to +\infty$ yields the co-monotonicity copula while $\theta = 0$ corresponds to independence. Here, we choose $\theta \in \{0, 10, 20\}$ leading to the association measure Kendall's $\tau \in \{0, 0.67, 0.82\}$, see [51, Section 5.1].

- the Gaussian copula defined for all $(u_1, u_2) \in [0, 1]^2$ by

$$C_\theta(u_1, u_2) = \Phi_{R_\theta}(\Phi^{-1}(u_1) + \Phi^{-1}(u_2)) \text{ with } R_\theta = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix},$$

  where Φ and Φ_{R_θ} are respectively the cumulative distribution functions of the standard univariate Gaussian distribution and bivariate centered Gaussian distribution with covariance matrix R_θ, $\theta \in (-1, 1)$. Here $\theta \to -1$ yields the counter-monotonicity copula, $\theta \to 1$ yields the co-monotonicity copula while $\theta = 0$ corresponds to independence. We choose $\theta \in \{0, 0.87, 0.96\}$ leading to the same Kendall's $\tau \in \{0, 0.67, 0.82\}$ as above.

The standard deviation σ is selected such that the Signal to Noise Ratio (SNR) defined as 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:

$$\mathrm{PC}(y_n) = \frac{1}{N}\sum_{r=1}^{N} \cos^2(\hat{v}(y_n)^{[r]}, \beta) = 1 - \frac{1}{N}\sum_{r=1}^{N} \Delta(\hat{v}(y_n)^{[r]}, \beta), \tag{7}$$

where $\Delta(\cdot,\cdot)$ is defined in (3) and $\hat{v}(y_n)^{[r]}$ denotes the estimator (4) computed on the rth replication. The closer PC(y_n) is to 1, the better the estimator is. The performance of EPLS and SIMEXQ methods are compared by computing (7) in the $4 \times 2 \times 2 \times 2 \times 3 = 96$ considered situations. To this end, denoting by $Y_{1,n} \le Y_{2,n} \le \cdots \le Y_{n,n}$ the order statistics of the sample $(Y_1, \ldots, Y_n)$, the quality measure PC($Y_{n-k+1,n}$) is plotted in Fig. 2 as a function of the number of exceedances $k \in \{1, \ldots, 200\}$ in the Pareto + Frank case with $p = 3$. Other situations including the Student distribution, the Gaussian copula and a larger dimension $p = 30$ are reported in Fig. 1–7 of the Supplementary material.

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 PC ≥ 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 PC ≥ 0.8) for a wide range of choices of k when $c \ge 1$ for all dependence situations. Good results (PC ≥ 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). The selection of k using the procedure described in Section 4 is illustrated 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 network, an annual database of commercial-sized farm holdings. This dataset of size $n = 949$ contains significant accounting and financial information about French farm incomes in 2014. Our goal is to investigate the relationship between low yields and various factors. The response variable Y is the inverse of the wheat yield (in quintals/hectare), as we focus on 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 average temperature (degree Celsius). We first carry out, in Fig. 3, 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 drawing a Hill plot:

$$\left(k, \hat{\gamma}_k = \frac{1}{k}\sum_{i=1}^{k} Z_{i,n}\right), \ k \in \{1, \ldots, 150\},$$

where $Z_{i,n} := \log(Y_{n-i+1,n}/Y_{n-k,n})$, $i \in \{1, \ldots, k\}$ denote the log-excesses computed from the consecutive top order statistics. The Hill statistics $\hat{\gamma}_k$ [38] aims at estimating the tail index γ under the semi-parametric model (A_1). In this

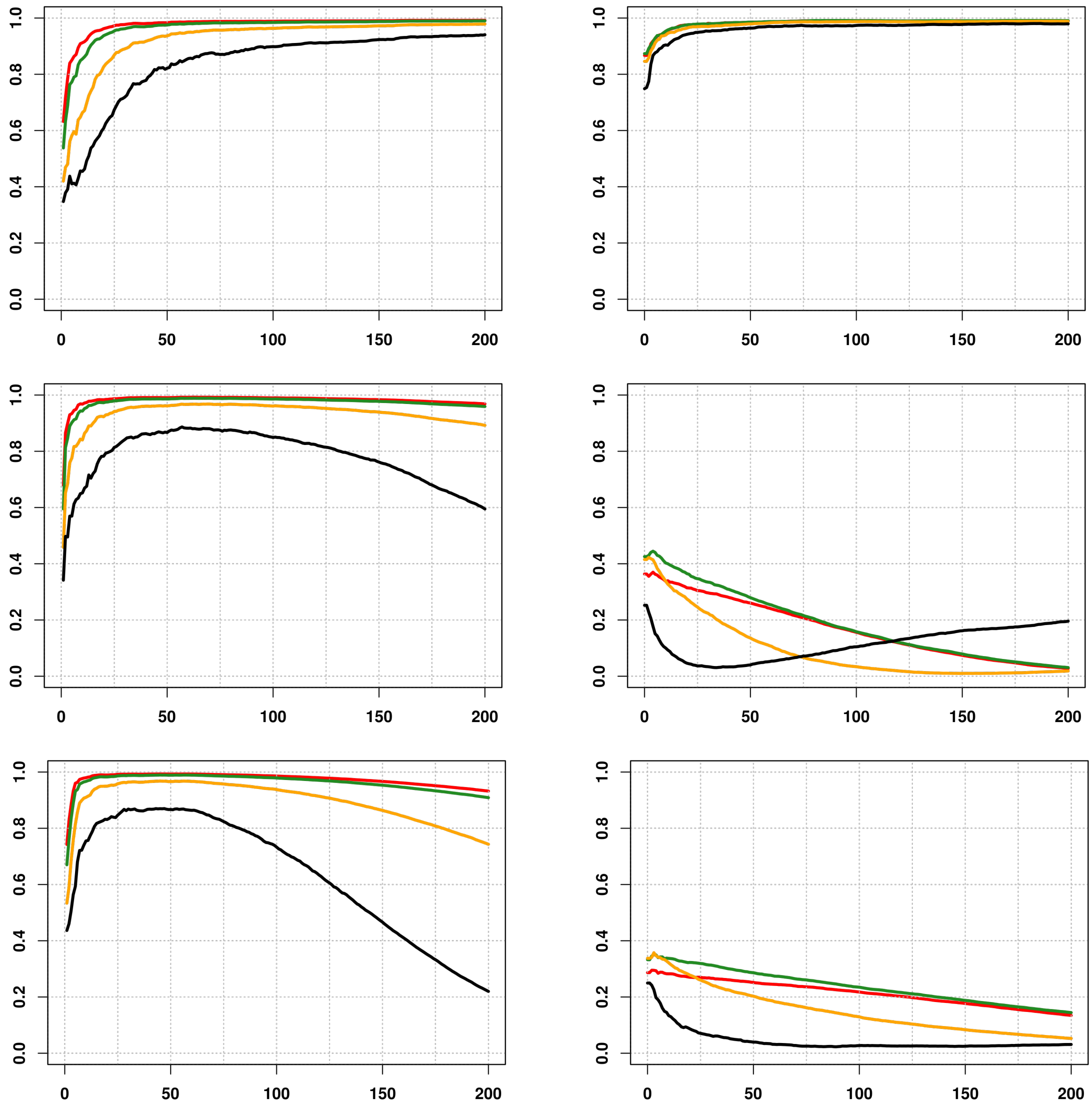

Fig. 2: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ [63] (right) estimators, on simulated data from a Pareto distribution, Frank copula, dimension $p = 3$. Horizontally: number $k \in \{1, \ldots, 200\}$ of exceedances, vertically: $\mathrm{PC}(Y_{n-k+1,n})$ quality measure. 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}.

situation, 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], and $\hat{\gamma}_k$ thus estimates γ using the empirical mean. The resulting graph (top right panel) shows a nice stability of $\hat{\gamma}_k$ as a function of $k \in \{50, \dots, 150\}$ 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 (bottom panel of Fig. 3) 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 thus set $\mathcal{Y} := \{Y_{n-k+1,n}, k \in \{50, \dots, 150\}\}$ and compute $\hat{v}^{(\ell)}(y_n)$ for $y_n \in \mathcal{Y}$ using the procedure described in Section 4. The top left panel of Fig. 4 displays the conditional correlation (6) between the projected residuals and the high values of the response variable Y. All graphs benefit from a stable behaviour with respect to the threshold $y_n \in \mathcal{Y}$, confirming together with the previous Hill-plot and quantile-quantile plot that the associated range of number of exceedances is well-suited to the dataset. It also appears that the first index captures about $\Xi_1 = 59\%$ of the correlation (with $k^{(1)} = 97$, blue curve), while the second index fails to represent a significant correlation ($\Xi_2 = 13\%$, $k^{(2)} = 58$, red curve). The second direction is thus discarded in the sequel. The top right panel of Fig. 4 represents the conditional correlation between the projected covariate $X^t \hat{v}^{(1)}(y_n)$ on the first direction and each coordinate $X^{(j)}$ of the covariate as a function of $y_n \in \mathcal{Y}$. 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). Besides, the effect of crop insurance purchase could be explained by moral hazard that leads insured farmers to use fewer agricultural inputs [55].

The projected scatter plot $(Y_i, X_i^t \hat{v}^{(1)}(y^{(1)}))$, $i \in \{1, \dots, n\}$ is displayed in a logarithmic scale for the visualization sake on the top panel of Fig. 5 together with two estimations (linear and non-linear) of the conditional mean $\mathbb{E}(X^t \hat{v}^{(1)}(y^{(1)}) \,|\, Y)$. A positive trend appears for large values of Y in accordance to the inverse regression model (M_1). Let us now focus on the conditional quantiles $\hat{q}(\alpha \,|\, X^t \hat{v}^{(1)}(y^{(1)}))$ computed through a kernel estimator of the conditional survival function [20]. The results are reported in the bottom panel of Fig. 5 together with the scatter plot $(X_i^t \hat{v}^{(1)}(y^{(1)}), Y_i)$, $i \in \{1, \dots, n\}$. 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 $\log(X^t \hat{v}^{(1)}(y^{(1)})) \leq 9.5$: Lowest yields are (mainly) linked to high operating costs. The interpretation of the results for $\log(X^t \hat{v}^{(1)}(y^{(1)})) > 9.5$ 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 [44].

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 neither to linearity nor to independence assumptions. An extension to the multi-index setting is proposed together with a data-driven method for selecting the number of directions to estimate. The proposed method can then be used to facilitate the estimation of extreme conditional quantiles or expectiles, 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, leveraging the theoretical tools introduced in [30].

Acknowledgments. The authors would like to thank two anonymous reviewers for their remarks which led to an improved presentation of their results. 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.

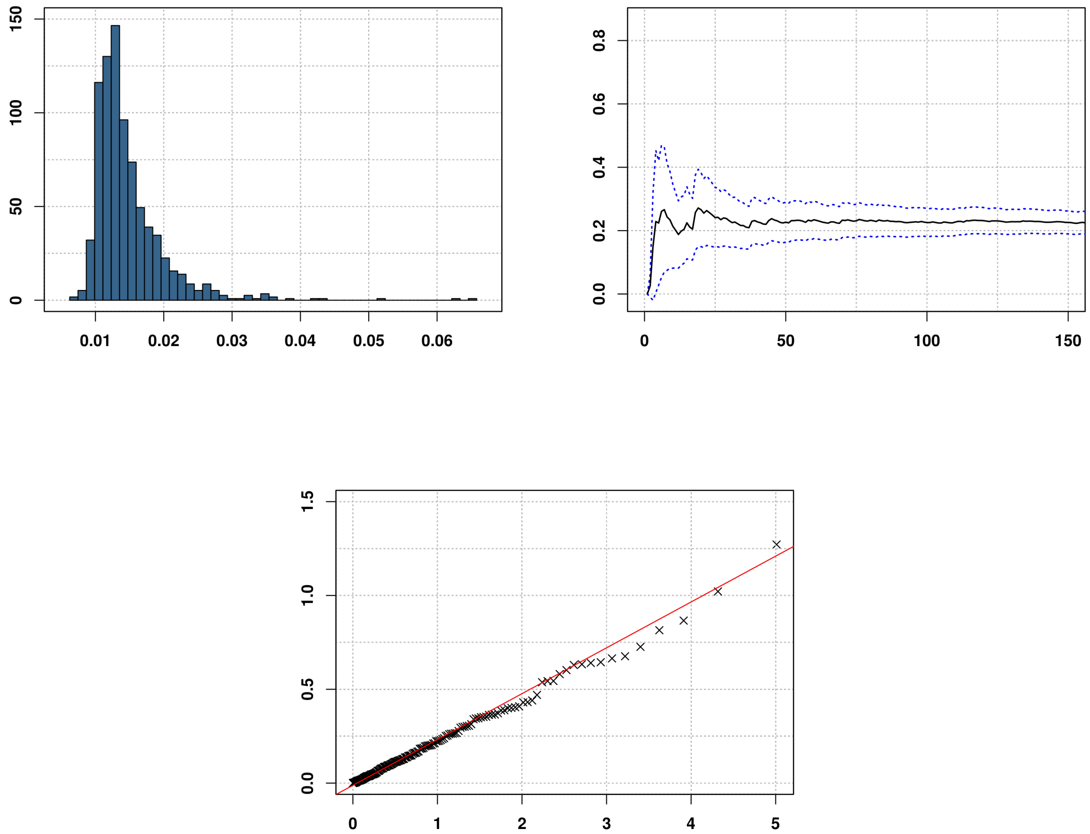

Fig. 3: Farm income data. Top left panel: Histogram of the inverse yields Y_i, $i \in \{1, \ldots, n\}$. Top right panel: Hill plot (Horizontally: $k \in \{1, \ldots, 150\}$, vertically: Hill estimator $\hat{\gamma}_k$, dashed blue line: empirical 95% confidence interval). Bottom panel: quantile-quantile plot (horizontally: $\log(k/i)$, vertically: $\log(Y_{n-i+1,n}/Y_{n-k,n})$ for $i \in \{1, \ldots, k = 150\}$, red: regression line).

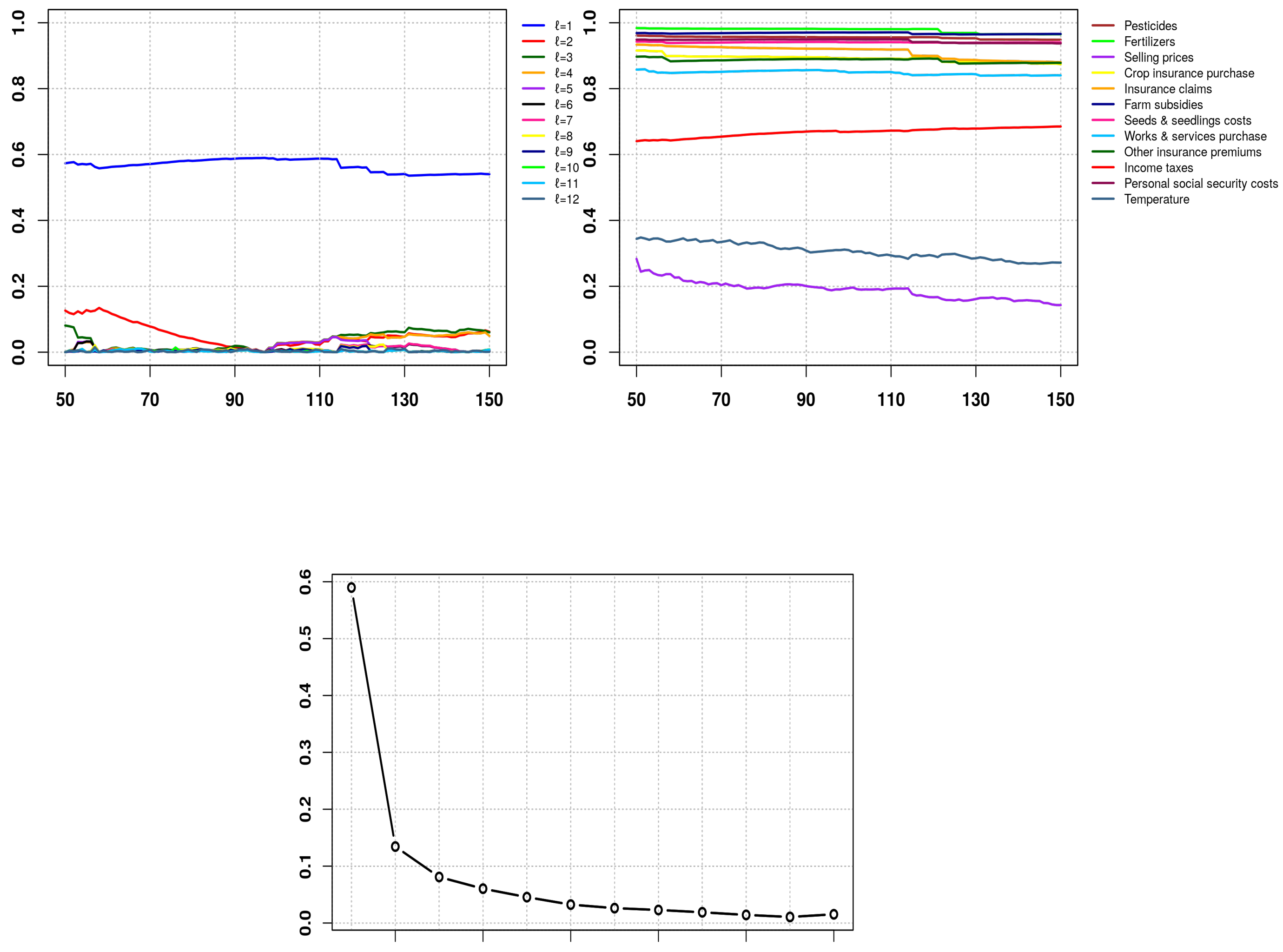

Fig. 4: Farm income data. Top left panel: Graph of the estimated conditional correlation function $y_n \in \mathcal{Y} \mapsto \rho((R^{(\ell-1)})^t \hat{\nu}^{(\ell)}(y_n), Y|Y \geq y_n)$ for $\ell \in \{1, \ldots, 12\}$ and top right panel: Graph of the estimated conditional correlation function $y_n \in \mathcal{Y} \mapsto \rho(X^t \hat{\nu}^{(1)}(y_n), X^{(j)}|Y \geq y)$ for $j \in \{1, \ldots, 12\}$ (horizontally: number of exceedances k, vertically: conditional correlation estimated by its empirical counterpart using the threshold $y_n = Y_{n-k+1,n}$). Bottom panel: Scree plot of $\ell \in \{1, \ldots, 12\} \mapsto \Xi_\ell$ (horizontally: iteration ℓ, vertically: maximum correlation Ξ_ℓ.

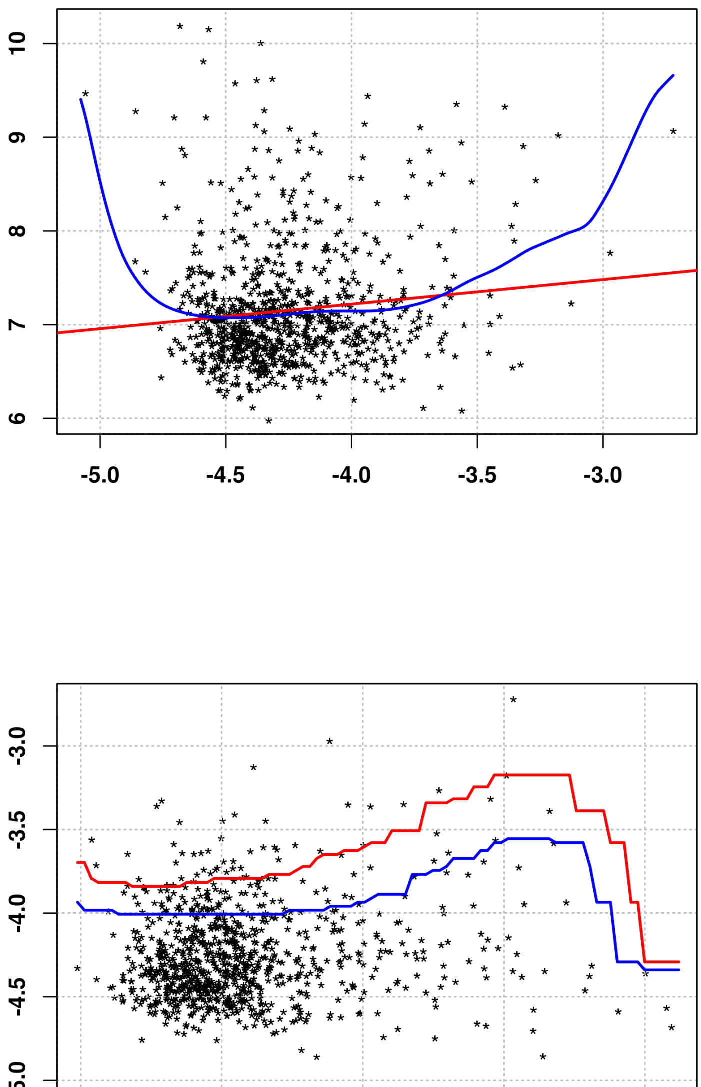

Fig. 5: Farm income data. Top: scatter-plot $(Y_i, X_i^t\,\hat{v}^{(1)}(y^{(1)}))$, $i \in \{1, \dots, n\}$ in log scale (horizontally: Y_i, vertically: $X_i^t\,\hat{v}^{(1)}(y^{(1)})$). The regression line (red) and a kernel estimate of the link function (blue) are superimposed. Bottom: scatter-plot $(X_i^t\,\hat{v}^{(1)}(y^{(1)}), Y_i)$, $i \in \{1, \dots, n\}$ in log scale (horizontally: $X_i^t\,\hat{v}^{(1)}(y^{(1)})$, vertically: Y_i). The estimated conditional quantiles $x \mapsto \hat{q}(\alpha | x^t\,\hat{v}(y^{(1)}))$ are superimposed ($\alpha = 0.15$: blue line, $\alpha = 0.05$: red line).

8. Appendix: Proofs

Some preliminary lemmas are first provided in Paragraph 8.1. They will reveal useful in the proofs of the main results collected in Paragraph 8.2.

8.1. 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 (A_1) with $\gamma q_2 < 1$. Let $\phi(\cdot)$ be a continuous function on $[1, \infty)$ such that $\phi(t) \to \kappa > 1$ as $t \to \infty$. Then,

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

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

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

see for example [32, 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: $|x - \phi(t)|^{q_2} \leq (2\kappa - 1)^{q_2} x^{q_2}$. Collecting the latter inequality with (8) 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 $x \mapsto x^{q_2 - \frac{1}{\gamma} - 1 + \epsilon}$ is integrable on $[1, \infty)$. Then, Lebesgue's dominated convergence theorem together with the regular variation property of f conclude the proof. □

The next lemma is an adaptation of [7, 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 (A_1) with $\gamma\mu < 1$. Then, as $y \to \infty$,

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

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, [7, Proposition 1.5.10] shows that, as $y \to \infty$,

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

Recalling that $f \in RV_{-1/\gamma-1}$ and using again [7, Proposition 1.5.10] prove that $\gamma y f(y) \sim \bar{F}(y)$ as $y \to \infty$. Finally, $\mathbb{E}[\rho(Y)|Y \geq y] \sim \rho(y)/(1 - \gamma\mu)$, as $y \to \infty$ and the conclusion follows. □

The following lemma establishes sufficient conditions such that the moment conditions of Proposition 1 hold in the context of the inverse regression model (M_1).

Lemma 3. Assume (M_1), (A_1), (A_2) and (A_3) hold with $\gamma(c+1) < 1$. Then, $\mathbb{E}(|Y|\mathbb{1}_{\{Y\geq y\}}) < \infty$, $\mathbb{E}(\|XY\|\mathbb{1}_{\{Y\geq y\}}) < \infty$ and $\mathbb{E}(\|X\|\mathbb{1}_{\{Y\geq y\}}) < \infty$ for all $y \in \mathbb{R}$.

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

$$\mathbb{E}(\|X\|\mathbb{1}_{\{Y\geq y\}}) < \mathbb{E}(|g(Y)|\mathbb{1}_{\{Y\geq y\}}) + \mathbb{E}\|\varepsilon\|.$$

Let us note that $\mathbb{E}(|g(Y)|\mathbb{1}_{\{Y\geq y\}}) < \infty$ 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 (A_3). Hence, $\mathbb{E}\|\varepsilon\| < \infty$ and $\mathbb{E}(\|X\|\mathbb{1}_{\{Y\geq y\}}) < \infty$. Third,

$$\mathbb{E}(\|XY\|\mathbb{1}_{\{Y\geq y\}}) < \mathbb{E}(|Yg(Y)|\mathbb{1}_{\{Y\geq y\}}) + \mathbb{E}(\|Y\varepsilon\|\mathbb{1}_{\{Y\geq y\}}),$$

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

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

for all $q_2 \geq 1$ such that $1/q + 1/q_2 = 1$. As already remarked, $\mathbb{E}(\|\varepsilon\|^q) < \infty$ from (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 $\mathbb{E}(|Y|^{q_2}\mathbb{1}_{\{Y\geq y\}}) < \infty$ as well. As a conclusion, $\mathbb{E}(\|XY\|\mathbb{1}_{\{Y\geq y\}}) < \infty$ and the result is proved. □

Lemma 4 provides, in the framework of model (M_1), an alternative expression of $v(y)$ defined in Proposition 1.

Lemma 4. Assume (M_1) and the assumptions of Proposition 1 hold. Then, for all $y \in \mathbb{R}$, $v(y)$ can be rewritten as

$$v(y) = \bar{F}(y)\mathbb{E}[g(Y)\Psi_y(Y)]\,(\beta + \eta(y))\,, \tag{9}$$

where

$$\eta(y) := \frac{\mathbb{E}[\varepsilon\Psi_y(Y)]}{\mathbb{E}[g(Y)\Psi_y(Y)]} \quad , \quad \Psi_y(Y) := \left(Y - \frac{m_Y(y)}{\bar{F}(y)}\right)\mathbb{1}_{\{Y\geq y\}}. \tag{10}$$

Proof. Proposition 1 states that $v(y) = \bar{F}(y)\mathbb{E}[XY\mathbb{1}_{\{Y\geq y\}}] - \mathbb{E}[X\mathbb{1}_{\{Y\geq y\}}]m_Y(y)$. Recalling that $X = g(Y)\beta + \varepsilon$ from model (M_1) yields

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

Hence the result. □

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 (10) in Lemma 4.

Lemma 5. Let $q_2 > 0$ and suppose Y is a random variable satisfying (A_1) with $\gamma q_2 < 1$. Then,

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

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

Proof. From (10), 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

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

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

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

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

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 (10) in Lemma 4.

Lemma 6. Let Y be a random variable satisfying (A_1) and (A_2) with $\gamma(c+1) < 1$. Then,

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

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

Proof. From (10), we have:

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

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 $y \to \infty$:

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

which concludes the proof. □

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 (M_1), (A_1), (A_2) and (A_3) hold with $\gamma(c+1) < 1$. Let $\zeta \in \{0,1\}$. Then,

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

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

as $y \to \infty$.

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

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

as $y \to \infty$.

Remark that, in view of the above lemma, condition (A_3) implies that, for all $j \in \{1, \ldots, p\}$ 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 (M_1), we have:

$$\begin{aligned}
\mathbb{E}(X_{.,j}Y^{\zeta}\mathbb{1}_{\{Y\geq y\}}) &= \mathbb{E}(Y^{\zeta}g(Y)\mathbb{1}_{\{Y\geq y\}})\beta_j + \mathbb{E}(Y^{\zeta}\varepsilon_{.,j}\mathbb{1}_{\{Y\geq y\}}) = \mathbb{E}(Y^{\zeta}g(Y)|Y\geq y)\bar{F}(y)\beta_j + \mathbb{E}(Y^{\zeta}\varepsilon_{.,j}\mathbb{1}_{\{Y\geq y\}}) \\
&= \frac{\beta_j}{1-\gamma(c+\zeta)}\, y^{\zeta}g(y)\bar{F}(y)(1+o(1)) + \mathbb{E}(Y^{\zeta}\varepsilon_{.,j}\mathbb{1}_{\{Y\geq y\}}),
\end{aligned} \tag{11}$$

in view of Lemma 2. Let $q_2 \geq 1$ such that $1/q + 1/q_2 = 1$. Combining Hölder inequality with (A_3) yields:

$$\begin{aligned}
\mathbb{E}|\varepsilon_{.,j}Y^{\zeta}\mathbb{1}_{\{Y\geq y\}}| &\leq [\mathbb{E}|\varepsilon_{.,j}|^q]^{1/q}\, [\mathbb{E}(|Y|^{\zeta q_2}\mathbb{1}_{\{Y\geq y\}})]^{1/q_2} = [\mathbb{E}|\varepsilon_{.,j}|^q]^{1/q}\, [\mathbb{E}(|Y|^{\zeta q_2}|Y\geq y)]^{1/q_2}\bar{F}(y)^{1/q_2} \\
&= [\mathbb{E}|\varepsilon_{.,j}|^q]^{1/q}\left(\frac{|y|^{\zeta q_2}}{1-q_2\zeta\gamma}\right)^{1/q_2}\bar{F}(y)^{1/q_2}(1+o(1)) = O\left(y^{\zeta}\bar{F}(y)^{1-1/q}\right),
\end{aligned} \tag{12}$$

as $y \to \infty$, 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 (11) and (12),

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

As a consequence of (A_1) and (A_2), $\bar{F}(\cdot)^{1/q}g(\cdot)$ is a regularly-varying function with index $c - 1/(q\gamma) > 0$. Therefore, $\bar{F}(y)^{1/q}g(y) \to \infty$ when $y \to \infty$ and (i) is proved. □

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

Lemma 8. Assume (M_1), (A_1), (A_2) and (A_3) hold with $2\gamma(c+1) < 1$. For all $\zeta \in \{0,1\}$ let

$$T^{(\zeta)}_{.,n} = \sqrt{\bar{F}(y_n)}\left(\sum_{\beta_j\neq 0}\frac{\alpha^{(\zeta)}_j\varepsilon_{.,j}}{m_{X_{.,j}Y^{\zeta}}(y_n)}\right)Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}},$$

where $\alpha^{(\zeta)}_j \in \mathbb{R}$ for all $j = 1,\dots,p$. Then, $\chi^{(\zeta)}_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n (T^{(\zeta)}_{i,n} - \mathbb{E}(T^{(\zeta)}_{i,n})) \xrightarrow{P} 0$.

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

$$\begin{aligned}
\mathrm{var}(\chi^{(\zeta)}_n) &= \mathrm{var}(T^{(\zeta)}_{.,n}) = \sum_{\beta_j\neq 0}\sum_{\beta_\ell\neq 0}\frac{\alpha^{(\zeta)}_j\alpha^{(\zeta)}_\ell\bar{F}(y_n)}{m_{X_{.,j}Y^{\zeta}}(y_n)m_{X_{.,\ell}Y^{\zeta}}(y_n)}\mathrm{cov}(\varepsilon_{.,j}Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}},\varepsilon_{.,\ell}Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}}) \\
&\sim \frac{(1-\gamma(c+\zeta))^2}{y_n^{2\zeta}g(y_n)^2\bar{F}(y_n)}\sum_{\beta_j\neq 0}\sum_{\beta_l\neq 0}\frac{\alpha^{(\zeta)}_j\alpha^{(\zeta)}_\ell}{\beta_j\beta_\ell}\mathrm{cov}(\varepsilon_{.,j}Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}},\varepsilon_{.,\ell}Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}}),
\end{aligned}$$

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

$$\mathrm{cov}(\varepsilon_{.,j}Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}},\varepsilon_{.,\ell}Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}}) = \mathbb{E}(\varepsilon_{.,j}\varepsilon_{.,\ell}Y^{2\zeta}\mathbb{1}_{\{Y\geq y_n\}}) - \mathbb{E}(\varepsilon_{.,j}Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}})\mathbb{E}(\varepsilon_{.,\ell}Y^{\zeta}\mathbb{1}_{\{Y\geq y_n\}}).$$

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

$$\begin{aligned}
\mathbb{E}(\varepsilon_{.,j}\varepsilon_{.,\ell}Y^{2\zeta}\mathbb{1}_{\{Y\geq y_n\}}) &\leq [\mathbb{E}|\varepsilon_{.,j}|^q]^{1/q}[\mathbb{E}|\varepsilon_{.,\ell}|^q]^{1/q}[\mathbb{E}|Y^{2\zeta}\mathbb{1}_{\{Y\geq y_n\}}|^{q_3}]^{1/q_3} \\
&\leq [\mathbb{E}|\varepsilon_{.,j}|^q]^{1/q}[\mathbb{E}|\varepsilon_{.,\ell}|^q]^{1/q}[\mathbb{E}|Y^{2\zeta q_3}|Y\geq y_n|]^{1/q_3}\bar{F}(y_n)^{1/q_3} \\
&= [\mathbb{E}|\varepsilon_{.,j}|^q]^{1/q}[\mathbb{E}|\varepsilon_{.,\ell}|^q]^{1/q}\left(\frac{y_n^{2\zeta q_3}}{1-2\zeta\gamma p'}\right)^{1/q_3}\bar{F}(y_n)^{1/q_3}(1+o(1)) = O\left(y_n^{2\zeta}\bar{F}(y_n)^{1-2/q}\right),
\end{aligned} \tag{13}$$

in view of Lemma 2 and (A_3). Indeed, condition $2\gamma(c+1) < 1$ is equivalent to $\gamma c < 1/2 - \gamma$. Besides, from (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:

$$|\mathbb{E}(\varepsilon_{\cdot,j} Y^{\zeta} \mathbb{1}_{\{Y \geq y_n\}}) \mathbb{E}(\varepsilon_{\cdot,\ell} Y^{\zeta} \mathbb{1}_{\{Y \geq y_n\}})| = O(y_n^{2\zeta} \bar{F}(y_n)^{2-2/q}) = o\left(y_n^{2\zeta} \bar{F}(y_n)^{1-2/q}\right). \tag{14}$$

Taking account of (13) and (14) yields

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

and therefore

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

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

8.2. Proofs of main results

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

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

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 $\lambda = \|v(y)\|$ and $w = v(y)/\|v(y)\|$. □

Proof of Proposition 2. From Lemma 3, $\mathbb{E}(|Y|\mathbb{1}_{\{Y \geq y\}})$, $\mathbb{E}(\|XY\|\mathbb{1}_{\{Y \geq y\}})$ and $\mathbb{E}(\|X\|\mathbb{1}_{\{Y \geq y\}})$ exist for all $y \in \mathbb{R}$. We may then apply Lemma 4 to get:

$$\cos(w(y), \beta) = w(y)^t \beta = \operatorname{sign}(\mathbb{E}[g(Y)\Psi_y(Y)]) \frac{1 + \beta^t \eta(y)}{\|\beta + \eta(y)\|}, \tag{16}$$

with $\eta(y) = \mathbb{E}[\varepsilon \Psi_y(Y)]/\mathbb{E}[g(Y)\Psi_y(Y)]$, see (10), and where $\operatorname{sign}(u) = 1$ is $u \geq 0$ and $\operatorname{sign}(u) = -1$ 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 $\|\eta(y)\| \to 0$ as $y \to \infty$ to get

$$1 - \cos^2(w(y), \beta) = O(\|\eta(y)\|^2), \tag{17}$$

as $y \to \infty$. Under assumption (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 $y \to \infty$, with $\lambda_1(\gamma, q_2) > 0$. Therefore,

$$\|\mathbb{E}[\varepsilon \Psi_y(Y)]\| \leq [\mathbb{E}\|\varepsilon\|^q]^{1/q} (\lambda_1(\gamma, q_2))^{1/q_2} y^{1+1/q_2} f(y)^{1/q_2}. \tag{18}$$

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

$$\mathbb{E}[g(Y)\Psi_y(Y)] \sim \lambda_2(\gamma, c) y g(y) \bar{F}(y), \tag{19}$$

with $\lambda_2(\gamma, c) > 0$. Collecting (18) and (19) thus yields:

$$\|\eta(y)\| \leq \frac{[\mathbb{E}\|\varepsilon\|^q]^{1/q}(\lambda_1(\gamma, q_2))^{1/q_2}}{\lambda_2(\gamma, c)} \frac{(yf(y))^{1/q_2}}{g(y)\bar{F}(y)}(1 + o(1)) = O\left(\frac{1}{g(y)\bar{F}^{1/q}(y)}\right). \tag{20}$$

Under assumptions (A_1) and (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, $\|\eta(y)\| \to 0$ as $y \to \infty$ and the first part of the result is proved. Second, (19) 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 (16), and then (17) entails that $\cos(w(y), \beta) \to 1$ as $y \to \infty$ 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 $y \to \infty$, from (20). The result is proved. □

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 $\alpha = (\alpha_1, \alpha_2, \alpha_{3,1}, \ldots, \alpha_{3,d}, \alpha_{4,1}, \ldots, \alpha_{4,d})^t \in \mathbb{R}^{2(d+1)}$ and let us investigate the asymptotic distribution of χ_n defined as follows:

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

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

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

Under model (M_1), $X_{i,j} = g(Y_i)\beta_j + \varepsilon_{i,j}$, for $j \in \{1, \ldots, d\}$ and $i \in \{1, \ldots, n\}$, we get the following decomposition:

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

where

$$T_{1,i,n} = \frac{\alpha_1}{\sqrt{\bar{F}(y_n)}}\mathbb{1}_{\{Y_i \geq y_n\}}, \qquad T_{2,i,n} = \sqrt{\bar{F}(y_n)}\frac{\alpha_2}{m_Y(y_n)}Y_i\mathbb{1}_{\{Y_i \geq y_n\}},$$

$$T_{3,i,n} = \sqrt{\bar{F}(y_n)}\left(\sum_{j=1}^{d}\frac{\alpha_{3,j}\beta_j}{m_{X_{\cdot,j}}(y_n)}\right)g(Y_i)\mathbb{1}_{\{Y_i \geq y_n\}}, \qquad T'_{3,i,n} = \sqrt{\bar{F}(y_n)}\left(\sum_{j=1}^{d}\frac{\alpha_{3,j}\varepsilon_{i,j}}{m_{X_{\cdot,j}}(y_n)}\right)\mathbb{1}_{\{Y_i \geq y_n\}},$$

$$T_{4,i,n} = \sqrt{\bar{F}(y_n)}\left(\sum_{j=1}^{d}\frac{\alpha_{4,j}\beta_j}{m_{X_{\cdot,j}Y}(y_n)}\right)Y_i g(Y_i)\mathbb{1}_{\{Y_i \geq y_n\}}, \qquad T'_{4,i,n} = \sqrt{\bar{F}(y_n)}\left(\sum_{j=1}^{d}\frac{\alpha_{4,j}\varepsilon_{i,j}}{m_{X_{\cdot,j}Y}(y_n)}\right)Y_i\mathbb{1}_{\{Y_i \geq y_n\}}.$$

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

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

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

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

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

$$\mathbb{E}(Y^{\zeta} g(Y)^{\omega} \mathbb{1}_{\{Y\ge y_n\}}) \sim \frac{1}{1-\gamma(\omega c+\zeta)}\, y_n^{\zeta} g(y_n)^{\omega} \bar{F}(y_n), \tag{21}$$

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

$$m_Y(y_n) \sim \frac{1}{1-\gamma} y_n \bar{F}(y_n), \tag{22}$$

$$m_{X_{\cdot,j}}(y_n) \sim \frac{\beta_j}{1-\gamma c} g(y_n)\bar{F}(y_n), \tag{23}$$

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

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:

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

Combining (21) and (22) leads to:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 [6, 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

$$
n\mathbb{E}\left|\sum_{\ell=1}^4 \frac{T_{\ell,1,n}-\mathbb{E}(T_{\ell,1,n})}{\sqrt{n}}\right|^{2+\delta} \to 0 \text{ as } n\to\infty. \tag{25}
$$

Using both triangle and Jensen inequalities, we get

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

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

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

leading to

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

which proves (25). As a conclusion, $\chi'_{0,n}\xrightarrow{d}\mathcal{N}(0,\alpha^tB\alpha)$ and $\sqrt{n\bar{F}(y_n)}\Lambda_n\xrightarrow{d}\mathcal{N}(0,B)$. □

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

$$\sigma_n^{-1}\left(\frac{\hat{v}_j(y_n)-v_j(y_n)}{\|v(y_n)\|}\right)_{1\leq j\leq d}\xrightarrow{d}\xi\left(\beta_j\right)_{1\leq j\leq d} \text{ with } \xi\sim\mathcal{N}\left(0,\lambda(c,\gamma)\right), \tag{26}$$

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

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

for all $j\in\{1,\dots,d\}$ and with $\vartheta_n\xrightarrow{d}\mathcal{N}(0,B)$. Substituting in (4), we get

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

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

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

From (22)–(24) in the proof of Proposition 3, it follows that, for all $j\in\{1,\dots,D\}$:

$$\frac{\sigma_n^{-1}}{y_ng(y_n)\bar{F}(y_n)^2}(\hat{v}_j(y_n)-v_j(y_n)) = \frac{\beta_j}{1-\gamma(c+1)}\left[\vartheta_{1,n}+\vartheta_{j+2+d,n}+o_P(1)\right]-\frac{\beta_j}{(1-\gamma)(1-\gamma c)}\left[\vartheta_{2,n}+\vartheta_{j+2,n}+o_P(1)\right]. \tag{27}$$

Besides, Lemma 4 yields $\|v(y_n)\| = \bar{F}(y_n)\mathbb{E}[g(Y)\Psi_{y_n}(Y)]\|\beta+\eta(y_n)\|$, with $\mathbb{E}[g(Y)\Psi_{y_n}(Y)]\sim\lambda_2(\gamma,c)\ y_ng(y_n)\bar{F}(y_n)$ and $\|\beta+\eta(y_n)\|\to\|\beta\|=1$ as $y_n\to\infty$, from (19) and (20) in the proof of Proposition 2. It follows that

$$\|v(y_n)\| = \lambda_2(\gamma,c)y_ng(y_n)\bar{F}^2(y_n)(1+o(1)). \tag{28}$$

Collecting (27) and (28) entails, for all $j\in\{1,\dots,d\}$:

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

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

$$\frac{\sigma_n^{-1}}{\|v(y_n)\|}\left(\hat{v}_j(y_n) - v_j(y_n)\right)_{1\le j\le d} \xrightarrow{d} \mathcal{N}\left(0, \operatorname{diag}(\beta_1,\dots,\beta_d)ABA^t\operatorname{diag}(\beta_1,\dots,\beta_d)\right), \tag{29}$$

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

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

Straightforward algebra shows that $\operatorname{diag}(\beta_1,\dots,\beta_d)ABA^t\operatorname{diag}(\beta_1,\dots,\beta_d) = \lambda(c,\gamma)(\beta_1,\dots,\beta_d)^t(\beta_1,\dots,\beta_d)$, so that the limiting Gaussian distribution is non-degenerated in the only direction $(\beta_1,\dots,\beta_d)^t$. Therefore, the convergence in distribution (29) can be rewritten as in (26). The second step consists in proving that

$$\sigma_n^{-1}\left(\frac{\hat{v}_j(y_n) - v_j(y_n)}{\|v(y_n)\|}\right)_{d+1\le j\le p} \xrightarrow{\mathbb{P}} 0. \tag{30}$$

For all $j\in\{d+1,\dots,p\}$ and $\zeta\in\{0,1\}$, the inverse model (M_1) shows that

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

Moreover, from (15) in the proof of Lemma 8, we have

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

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

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

Substituting in (4) and taking into account Lemma 7(ii) yield

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

Therefore, in view of (28), 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 $n\to\infty$ 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 (30) holds, hence the result. □

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

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

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

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

as $y_n \to \infty$, and the result is proved. □

References

[1] A. A. Ahmad, A. Diop, S. Girard, A. Usseglio-Carleve, Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model, Electron. J. Stat. 14 (2020) 4421–4456.
[2] J. Beirlant, Y. Goegebeur, Local polynomial maximum likelihood estimation for Pareto-type distributions, J. Multivariate Anal. 89 (2004) 97–118.
[3] J. Beirlant, Y. Goegebeur, J. Segers, J. L. Teugels, Statistics of extremes: theory and applications, volume 558, John Wiley & Sons, New-York, 2004.
[4] C. Bernard-Michel, L. Gardes, S. Girard, Gaussian regularized sliced inverse regression, Stat. Comput. 19 (2009) 85–98.
[5] P. K. Bhattacharya, A. K. Gangopadhyay, Kernel and nearest-neighbor estimation of a conditional quantile, Ann. Stat. 18 (1990) 1400–1415.
[6] P. Billingsley, Measure and Probability (third edition), John Wiley & Sons, New-York, 1995.
[7] N. Bingham, C. Goldie, J. Teugels, Regular Variation, volume 27 of Encyclopedia of Mathematics and its application, Cambridge University Press, 1987.
[8] J.-J. Cai, J. H. Einmahl, L. de Haan, C. Zhou, Estimation of the marginal expected shortfall: the mean when a related variable is extreme, J. Roy. Stat. Soc. B 77 (2015) 417–442.
[9] R. B. Cattell, The scree test for the number of factors, Multivar. Behav. Res. 1 (1966) 245–276.
[10] V. Chavez-Demoulin, A. C. Davison, Generalized additive modelling of sample extremes, J. Roy. Stat. Soc. C 54 (2005) 207–222.
[11] V. Chernozhukov, Extremal quantile regression, Ann. Stat. 33 (2005) 806–839.
[12] A. Chiancone, F. Forbes, S. Girard, Student sliced inverse regression, Comput. Stat. Data An. 113 (2017) 441–456.
[13] H. Chun, S. Keleş, Sparse partial least squares regression for simultaneous dimension reduction and variable selection, J. Roy. Stat. Soc. B 72 (2010) 3–25.
[14] R. D. Cook, Fisher lecture: Dimension reduction in regression, Stat. Sci. 22 (2007) 1–26.
[15] R. D. Cook, L. Forzani, Big data and partial least-squares prediction, Can. J. Stat. 46 (2018) 62–78.
[16] R. D. Cook, I. Helland, Z. Su, Envelopes and partial least squares regression, J. Roy. Stat. Soc. B 75 (2013) 851–877.
[17] D. Cooley, E. Thibaud, Decompositions of dependence for high-dimensional extremes, Biometrika 106 (2019) 587–604.
[18] R. Coudret, S. Girard, J. Saracco, A new sliced inverse regression method for multivariate response, Comput. Stat. Data An. 77 (2014) 285–299.
[19] A. Daouia, L. Gardes, S. Girard, On kernel smoothing for extremal quantile regression, Bernoulli 19 (2013) 2557–2589.
[20] A. Daouia, L. Gardes, S. Girard, A. Lekina, Kernel estimators of extreme level curves, Test 20 (2011) 311–333.
[21] A. C. Davison, N. Ramesh, Local likelihood smoothing of sample extremes, J. Roy. Stat. Soc. B 62 (2000) 191–208.
[22] A. C. Davison, R. L. Smith, Models for exceedances over high thresholds, J. Roy. Stat. Soc. B 52 (1990) 393–425.
[23] H. Drees, A. Sabourin, Principal component analysis for multivariate extremes, Electron. J. Stat. 15 (2021) 908–943.
[24] J. El Methni, L. Gardes, S. Girard, Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions, Scand. J. Stat. 41 (2014) 988–1012.
[25] L. E. Frank, J. H. Friedman, A statistical view of some chemometrics regression tools, Technometrics 35 (1993) 109–135.
[26] L. Gardes, Tail dimension reduction for extreme quantile estimation, Extremes 21 (2018) 57–95.
[27] L. Gardes, S. Girard, Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels, Extremes 13 (2010) 177–204.
[28] L. Gardes, S. Girard, Functional kernel estimators of large conditional quantiles, Electron. J. Stat. 6 (2012) 1715–1744.
[29] S. Girard, H. Lorenzo, J. Saracco, Advanced topics in sliced inverse regression, J. Multivariate Anal. 188 (2022) 104852.
[30] S. Girard, G. Stupfler, A. Usseglio-Carleve, Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models, Ann. Stat. 49 (2021) 3358–3382.
[31] Y. Goegebeur, A. Guillou, A. Schorgen, Nonparametric regression estimation of conditional tails: the random covariate case, Statistics 48 (2014) 732–755.
[32] L. de Haan, A. Ferreira, Extreme value theory: an introduction, Springer Science & Business Media, New-York, 2007.
[33] P. Hall, N. Tajvidi, Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, Stat. Sci. 15 (2000) 153–167.

[34] W. Härdle, T. M. Stoker, Investigating smooth multiple regression by the method of average derivatives, J. Am. Stat. Assoc. 84 (1989) 986–995.
[35] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Springer Series in Statistics, New-York, 2001.
[36] X. He, P. Ng, S. Portnoy, Bivariate quantile smoothing splines, J. Roy. Stat. Soc. B 60 (1998) 537–550.
[37] I. S. Helland, Partial least squares regression and statistical models, Scand. J. Stat. 17 (1990) 97–114.
[38] B. Hill, A simple general approach to inference about the tail of a distribution, Ann. Stat. 3 (1975) 1163–1174.
[39] A. E. Hoerl, R. W. Kennard, Ridge regression: applications to nonorthogonal problems, Technometrics 12 (1970) 69–82.
[40] J. L. Horowitz, Semiparametric and nonparametric methods in econometrics, volume 12, Springer, New-York, 2009.
[41] T. H. Jagger, J. B. Elsner, Modeling tropical cyclone intensity with quantile regression, Int. J. Climatol. 29 (2009) 1351–1361.
[42] R. Koenker, G. Bassett Jr, Regression quantiles, Econometrica 46 (1978) 33–50.
[43] E. Kong, Y. Xia, A single-index quantile regression model and its estimation, Economet. Theor. 28 (2012) 730–768.
[44] C. Kyung-Joon, W. R. Schucany, Nonparametric kernel regression estimation near endpoints, J. Stat. Plan. Infer. 66 (1998) 289–304.
[45] K.-C. Li, Sliced inverse regression for dimension reduction, J. Am. Stat. Assoc. 86 (1991) 316–327.
[46] L. Li, R. D. Cook, C.-L. Tsai, Partial inverse regression, Biometrika 94 (2007) 615–625.
[47] Y. Li, D. von Rosen, Maximum likelihood estimators in a two step model for PLS, Commun. Stat. Theory Methods 41 (2012) 2503–2511.
[48] H. Martens, T. Naes, Multivariate calibration, John Wiley & Sons, New-York, 1992.
[49] L. Meligkotsidou, I. D. Vrontos, S. D. Vrontos, Quantile regression analysis of hedge fund strategies, J. Empir. Finance 16 (2009) 264–279.
[50] P. Naik, C.-L. Tsai, Partial least squares estimator for single-index models, J. Roy. Stat. Soc. B 62 (2000) 763–771.
[51] R. B. Nelsen, An introduction to copulas, Springer Science & Business Media, New-York, 2007.
[52] J. L. Powell, J. H. Stock, T. M. Stoker, Semiparametric estimation of index coefficients, Econometrica 57 (1989) 1403–1430.
[53] J. Saracco, An asymptotic theory for sliced inverse regression, Commun. Stat. Theory Methods 26 (1997) 2141–2171.
[54] R. L. Smith, Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone, Stat. Sci. 4 (1989) 367–377.
[55] V. H. Smith, B. K. Goodwin, Crop insurance, moral hazard, and agricultural chemical use, Am. J. Agr. Econ. 78 (1996) 428–438.
[56] M. Stone, R. J. Brooks, Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression, J. Roy. Stat. Soc. B 52 (1990) 237–258.
[57] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. Roy. Stat. Soc. B 58 (1996) 267–288.
[58] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, J. Roy. Stat. Soc. B 67 (2005) 91–108.
[59] L. Wang, L. Yang, Spline estimation of single-index models, Stat. Sinica 19 (2009) 765–783.
[60] H. Wold, Soft modelling by latent variables: the non-linear iterative partial least squares (nipals) approach, J. Appl. Probab. 12 (1975) 117–142.
[61] H.-M. Wu, Kernel sliced inverse regression with applications to classification, J. Comput. Graph. Stat. 17 (2008) 590–610.
[62] T. Z. Wu, K. Yu, Y. Yu, Single-index quantile regression, J. Multivariate Anal. 101 (2010) 1607–1621.
[63] W. Xu, D. Li, H. Wang, Extreme quantile estimation based on the tail single-index model, Stat. Sinica 32 (2020) 1–22.
[64] Y. Yu, D. Ruppert, Penalized spline estimation for partially linear single-index models, J. Am. Stat. Assoc. 97 (2002) 1042–1054.
[65] L. Zhu, M. Huang, R. Li, Semiparametric quantile regression with high-dimensional covariates, Stat. Sinica 22 (2012) 1379–1401.
[66] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. Roy. Stat. Soc. B 67 (2005) 301–320.

Supplementary material for
“Extreme Partial Least-Squares”
by M. Bousebata, G. Enjolras & S. Girard

This Supplementary material includes additional results on $n = 1000$ simulated data (Section 5 of the main paper). The finite sample behavior of the EPLS estimator $\hat{v}(y_n)$ is illustrated and compared to SIMEXQ (single-index model extreme quantile) estimator. Seven cases are investigated:

- dimension $p = 3$, Frank copula and Student distribution (Fig. 1),
- dimension $p = 30$, Frank copula and Pareto distribution (Fig. 2),
- dimension $p = 30$, Frank copula and Student distribution (Fig. 3),
- dimension $p = 3$, Gaussian copula and Pareto distribution (Fig. 4),
- dimension $p = 3$, Gaussian copula and Student distribution (Fig. 5),
- dimension $p = 30$, Gaussian copula and Pareto distribution (Fig. 6),
- dimension $p = 30$, Gaussian copula and Student distribution (Fig. 7),

while the remaining case ($p = 3$, Frank copula and Pareto distribution) is presented in Fig. 2 of the main paper.

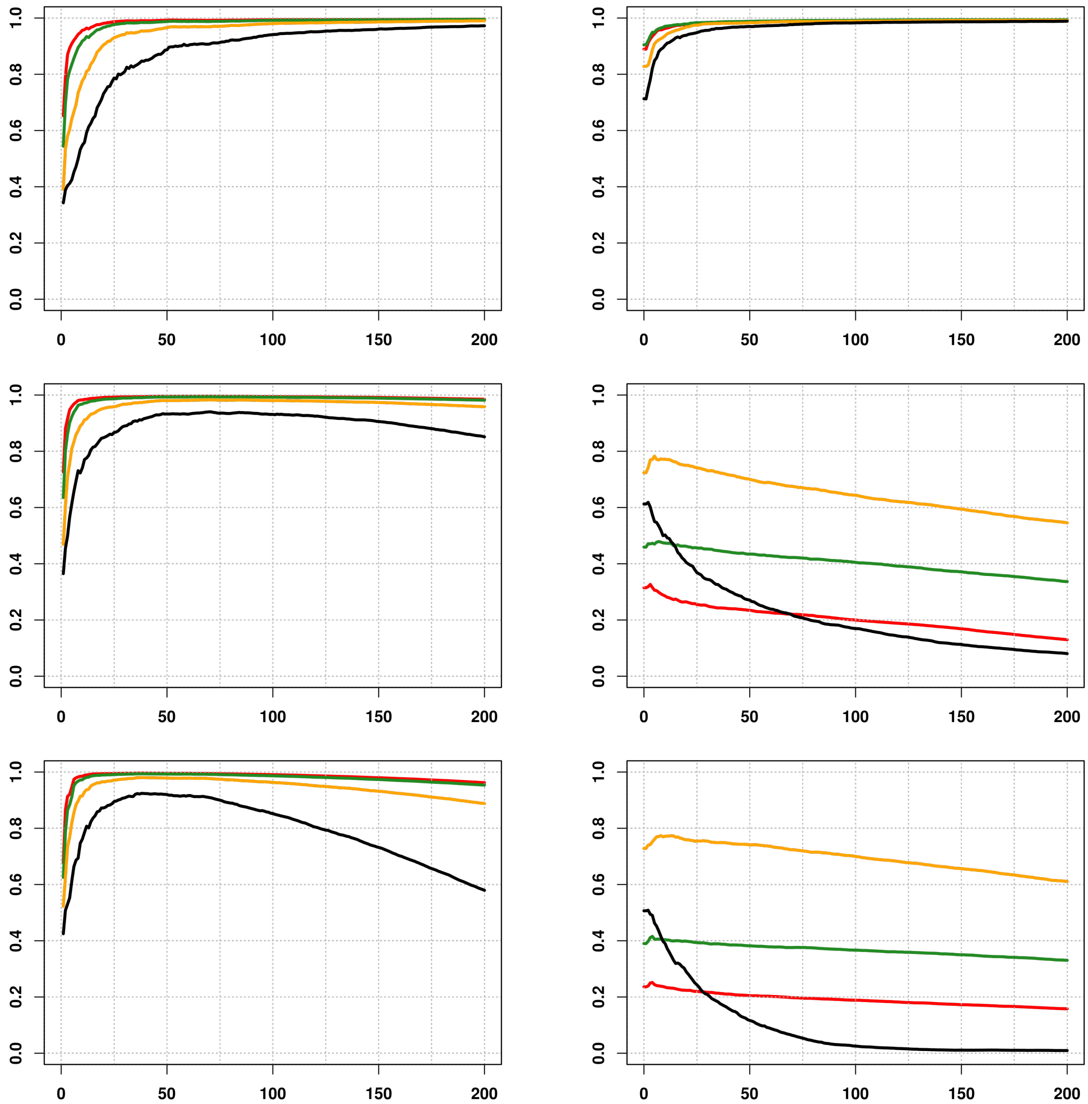

Fig. 1: Finite sample behaviour of EPLS $\hat{\nu}(Y_{n-k+1,n})$ (left) and SIMEXQ (right) estimators, on simulated data from a Student distribution, Frank copula, dimension $p = 3$. Horizontally: number $k \in \{1, \ldots, 200\}$ of exceedances, vertically: $\mathrm{PC}(Y_{n-k+1,n})$ quality measure. 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}.

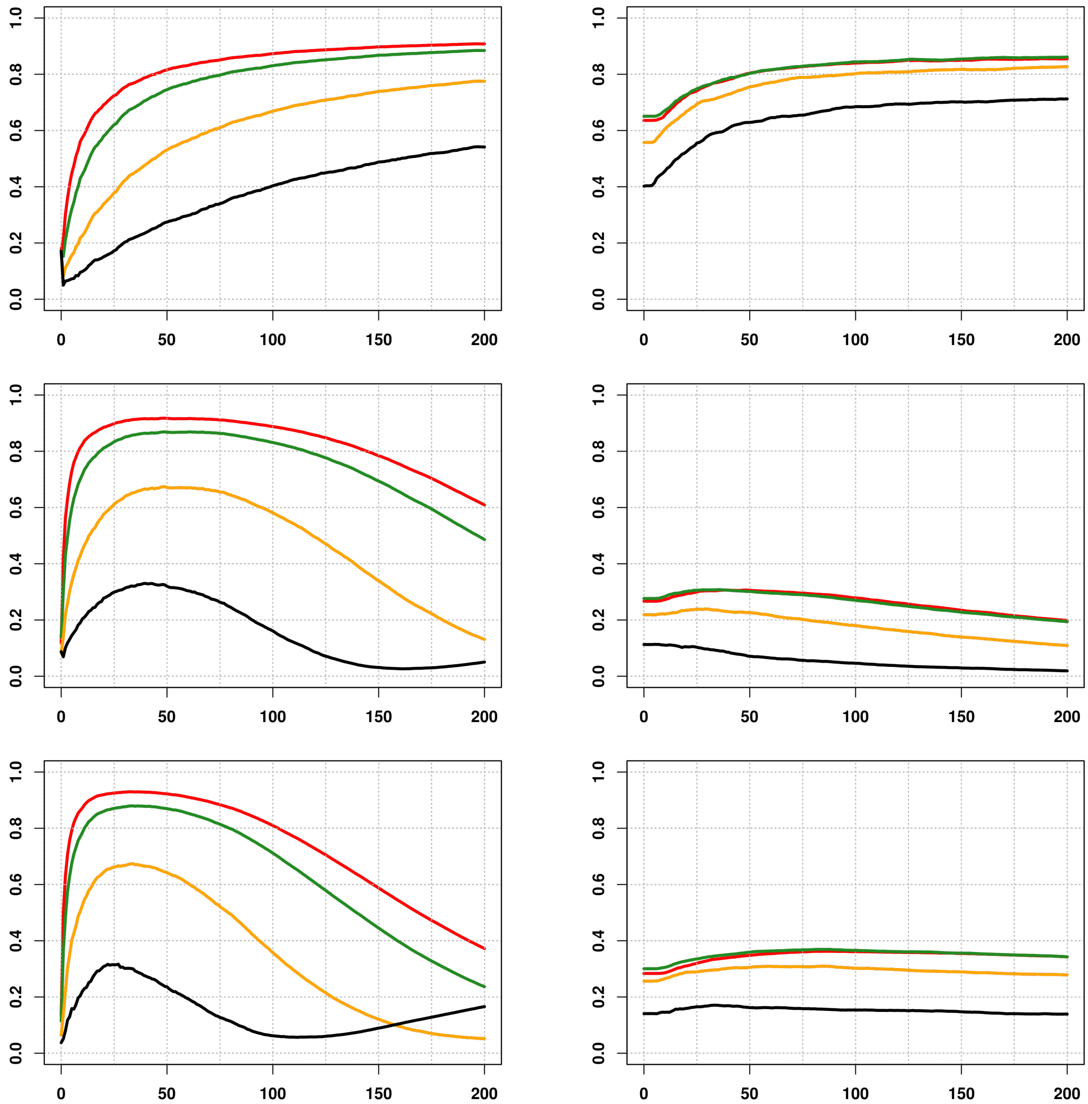

Fig. 2: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ (right) estimators, on simulated data from a Pareto distribution, Frank copula, dimension $p = 30$. Horizontally: number $k \in \{1, \dots, 200\}$ of exceedances, vertically: $\mathrm{PC}(Y_{n-k+1,n})$ quality measure. From 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}.

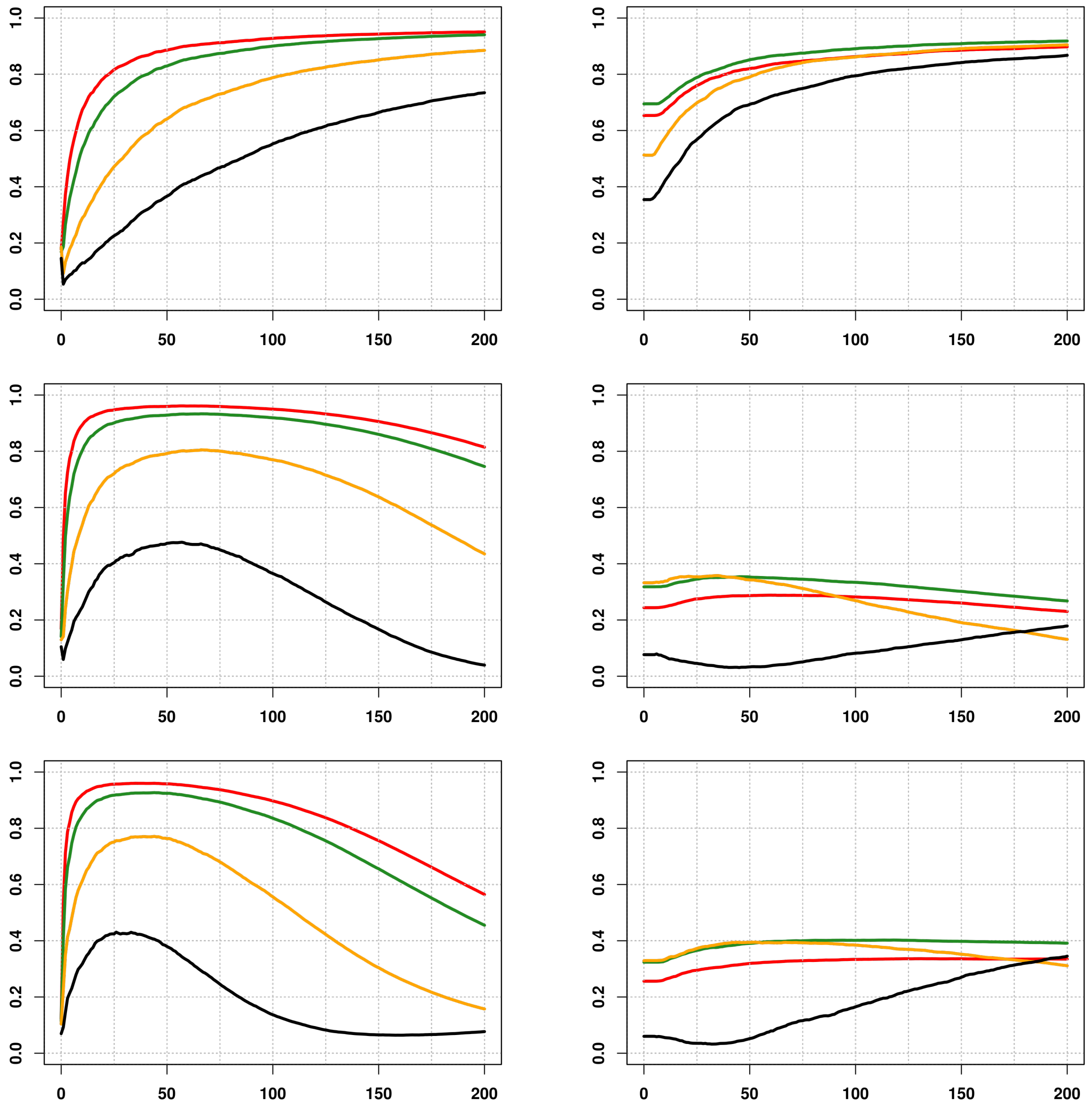

Fig. 3: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ (right) estimators, on simulated data from a Student distribution, Frank copula, dimension $p = 30$. Horizontally: number $k \in \{1, \ldots, 200\}$ of exceedances, vertically: $\mathrm{PC}(Y_{n-k+1,n})$ quality measure. 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}.

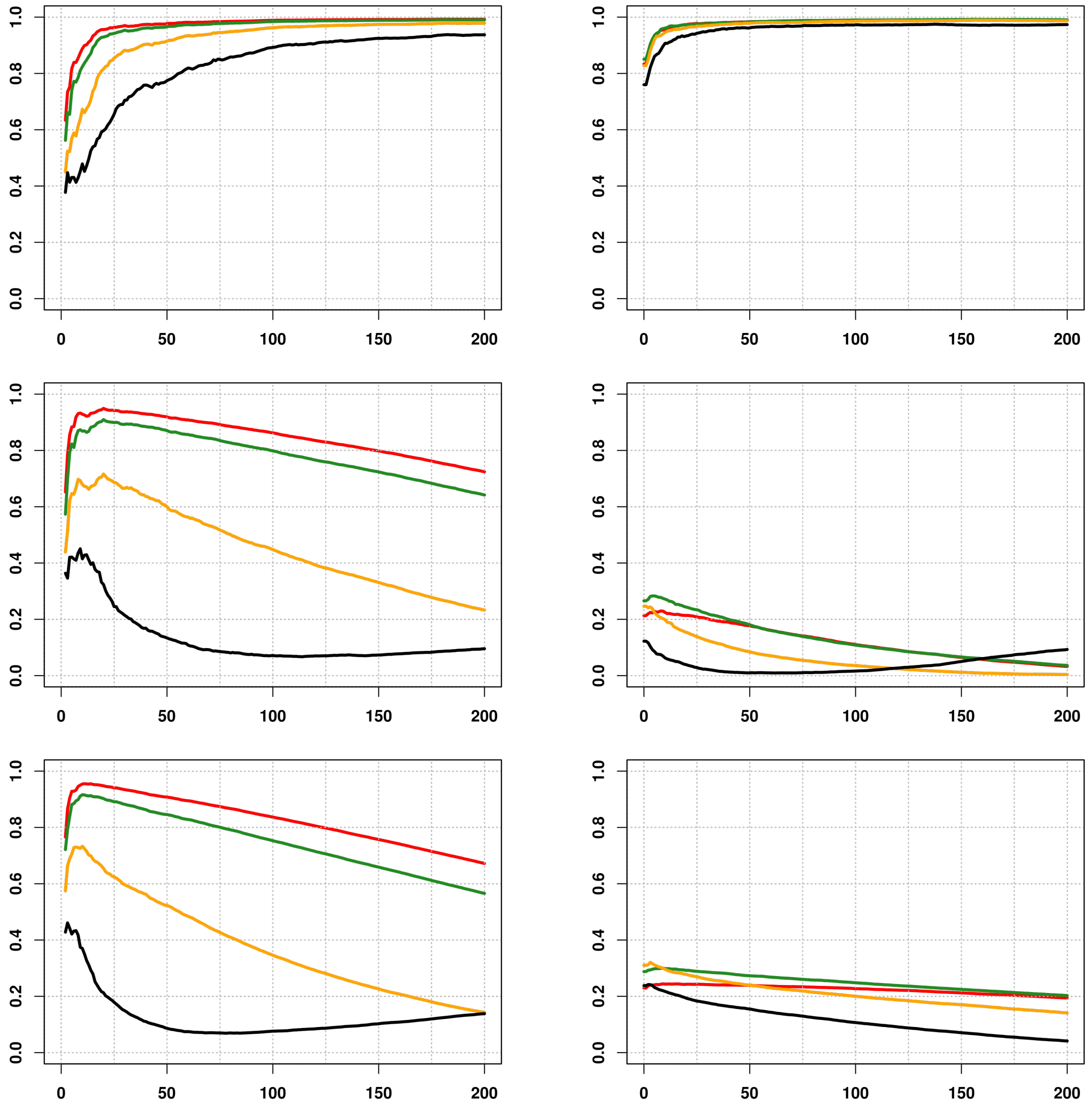

Fig. 4: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ (right) estimators, on simulated data from a Pareto distribution, Gaussian copula, dimension $p = 3$. Horizontally: number $k \in \{1, \ldots, 200\}$ of exceedances, vertically: $\mathrm{PC}(Y_{n-k+1,n})$ quality measure. From top to bottom, Gaussian copula parameter $\theta \in \{0, 0.87, 0.96\}$. 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}.

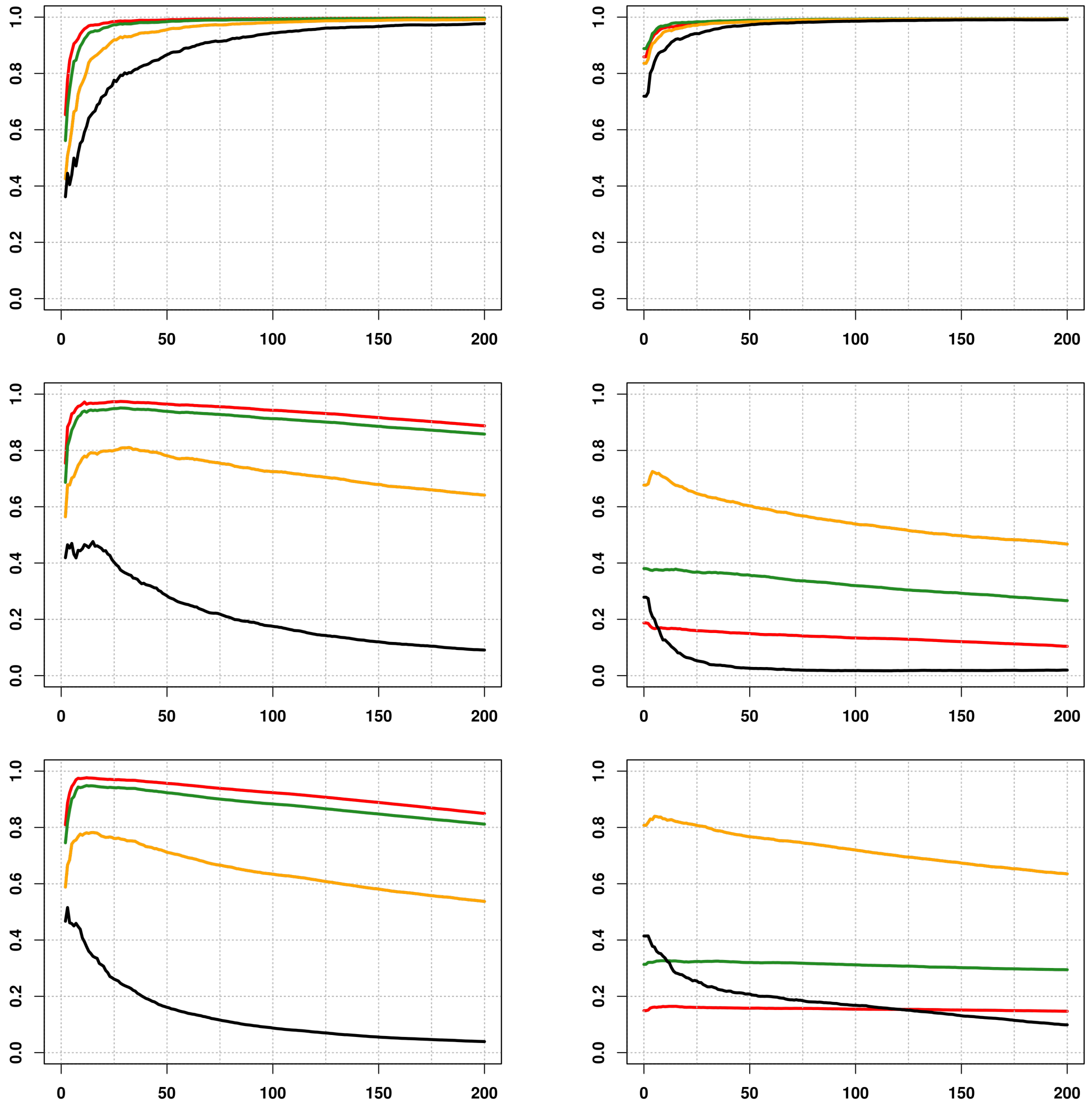

Fig. 5: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ (right) estimators, on simulated data from a Student distribution, Gaussian copula, dimension $p = 3$. Horizontally: number $k \in \{1, \ldots, 200\}$ of exceedances, vertically: $\mathrm{PC}(Y_{n-k+1,n})$ quality measure. From top to bottom, Gaussian copula parameter $\theta \in \{0, 0.87, 0.96\}$. 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}.

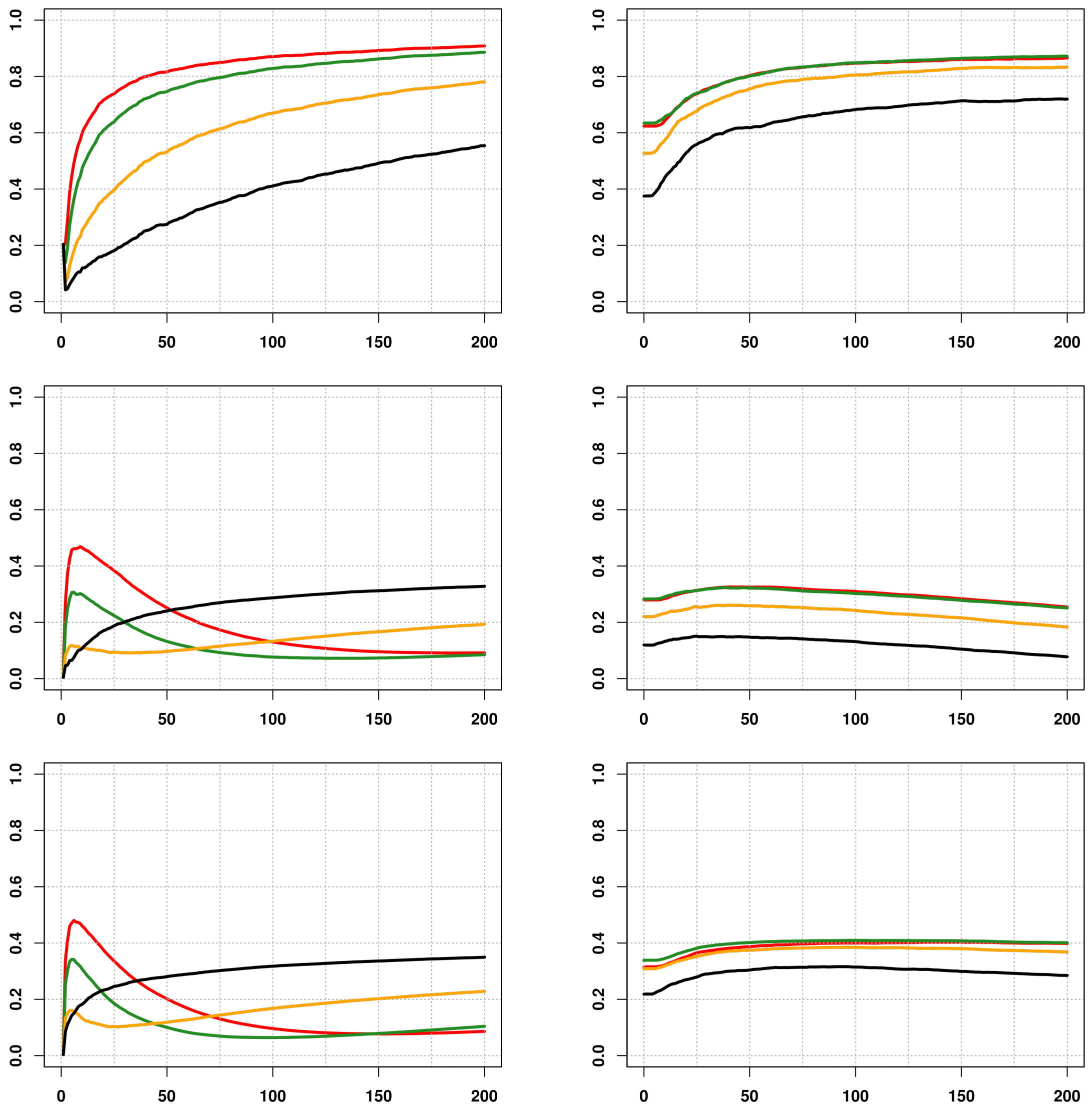

Fig. 6: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ (right) estimators, on simulated data from a Pareto distribution, Gaussian copula, dimension $p = 30$. Horizontally: number $k \in \{1, \dots, 200\}$ of exceedances, vertically: $\mathrm{PC}(Y_{n-k+1,n})$ quality measure. From top to bottom, Gaussian copula parameter $\theta \in \{0, 0.87, 0.96\}$. 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}.

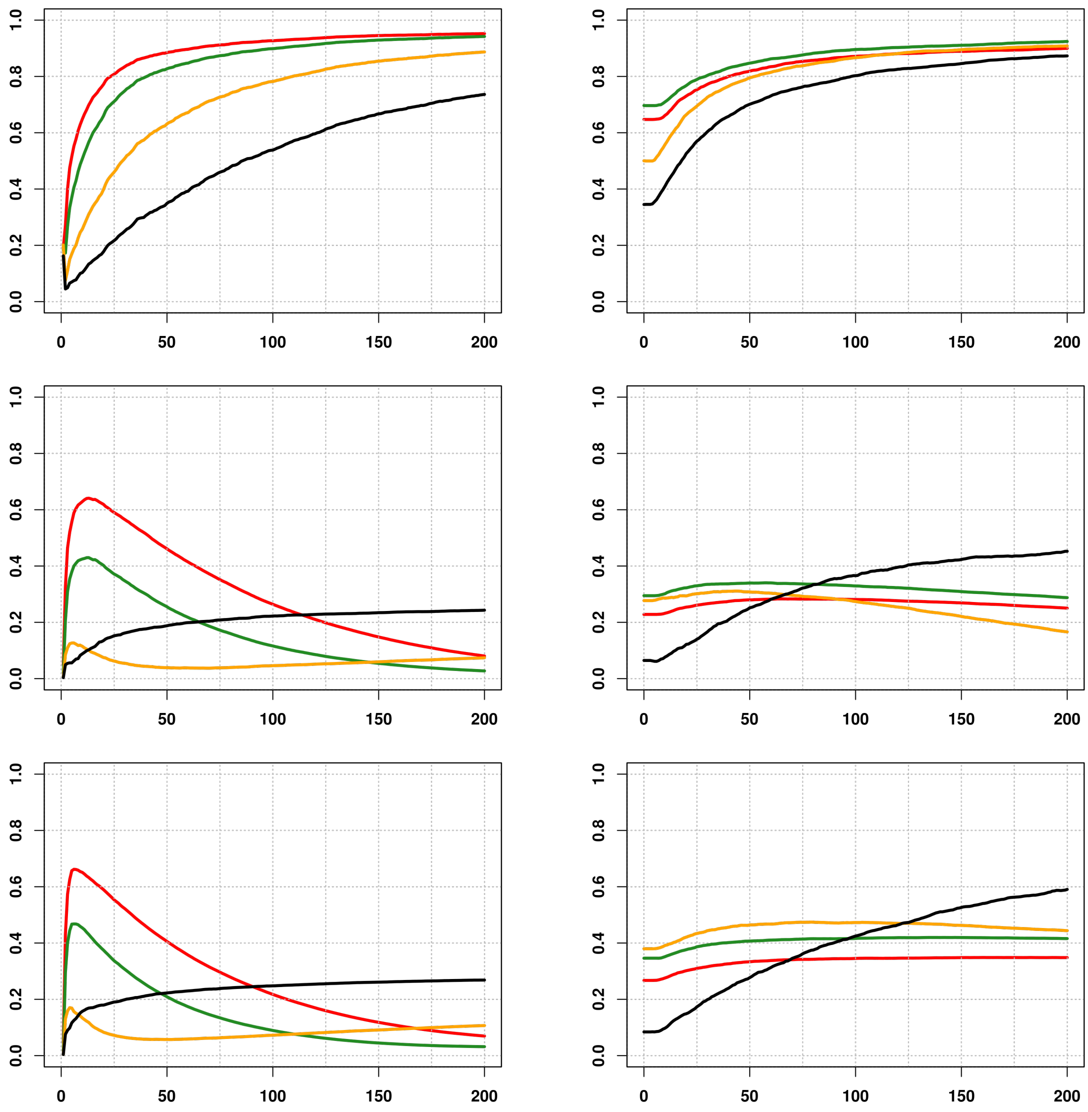

Fig. 7: Finite sample behaviour of EPLS $\hat{v}(Y_{n-k+1,n})$ (left) and SIMEXQ (right) estimators, on simulated data from a Student distribution, Gaussian copula, dimension $p = 30$. Horizontally: number $k \in \{1, \dots, 200\}$ of exceedances, vertically: $\mathrm{PC}(Y_{n-k+1,n})$ quality measure. From top to bottom, Gaussian copula parameter $\theta \in \{0, 0.87, 0.96\}$. 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}.