arXiv:2410.05517v3 [math.ST] 15 Jun 2026

FUNCTIONAL EXTREME-PLS

STÉPHANE GIRARD[1] AND CAMBYSE PAKZAD[2]

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

[2]*Modal'X, Université Paris Nanterre, 200 Avenue de la République, 92001 Nanterre, France*

Abstract. We propose an extreme dimension reduction method extending the Extreme-PLS approach to the discretized functional framework, where the covariate lies in the infinite-dimensional Hilbert space $L^2([0,1])$ but is partially observed on a dense time grid. The ideas are partly borrowed from both Partial Least-Squares (PLS) and Sliced Inverse Regression (SIR) techniques. Accordingly, the method relies on the projection of the covariate onto a subspace and maximizes the covariance between its projection and the response conditionally on an extreme event capturing the tail-information. The covariate and the heavy-tailed response are supposed to be linked through a non-linear inverse single-index model and our goal is to infer the index in this regression framework. We propose a new family of estimators and show its asymptotic consistency with convergence rates under the model. Assuming mild conditions on the noise, most of the assumptions are stated in terms of regular variation unlike the standard literature on SIR and single-index regression. In addition, we expand the theoretical analysis with a model-free almost sure consistency result for the empirical tail-moments in a general separable Hilbert space. Finally, our results are illustrated on a finite-sample study with synthetic functional data as well as high-frequency financial data, highlighting the effectiveness of the dimension reduction for capturing tail dependence and for extreme risk management.

1. Introduction

Among the popular methods for dimension reduction, the Partial Least Squares (PLS) method, initiated by Wold (1975) and adapted to $L^2([0,1])$ by Preda and Saporta (2005), has proven to be successful in both theoretical and practical sides; especially in chemometrics (Martens and Naes, 1992) where it takes root. It combines characteristics of Principal Component Analysis (PCA) and regression. Its core idea is to find a projection of the covariate/predictor X having high variance and correlation with the response Y. Another standard line of research is Sliced Inverse Regression (SIR), originally introduced by Li (1991) and then by Dauxois et al. (2001) in the functional framework. It hinges on the inverse

E-mail address: [1]stephane.girard@inria.fr, [2]cambyse.pakzad@parisnanterre.fr.

2020 *Mathematics Subject Classification.* 60G70, 62G08, 62G32, 62R10.

Key words and phrases. Extreme value theory; extremal stochastic processes; nonparametric regression and quantile regression; statistics of extreme values and tail-inference; functional data analysis.

regression 'X against Y' that happens to be simpler than the forward regression, usually at the price of a certain condition of linearity and constant variance. This somewhat departs from the Fisherian viewpoint, in which inference is carried out conditionally on the covariate X; see Cook (2007) for a discussion advocating inverse regression.

In any case, the aforementioned techniques require the full sample and are not tailored for extreme-value analysis. Hence, one may wonder what happens in the extreme regime, *i.e.,* when the focus is on the conditional tail of Y given a high-dimensional covariate X. This question belongs to an active line of research combining dimension reduction and extreme-value statistics. Here, a major challenge is a twofold scarcity, in the sense that both the curse of dimensionality and the rarity of tail observations operate. Such hindrance may apply in the estimation of certain statistics or risk measures in the functional extreme regime (see *e.g.*, Gardes and Girard (2012); Girard et al. (2022)). To circumvent this issue, one would build up the inference on the existence of a projection that captures all the information on the extreme values of Y. Next, instead of using standard estimation methods on X, one would replace X by its projection as the covariate, opting for a lower-dimensional method expected to yield comparable or even better performances. The pioneering work may be identified as Gardes (2018), later followed up by Gardes (2020), which proposes a new notion of tail conditional independence. Applying this to the extremal quantile regression, the author then studies the inference under a single and multi-index model when the dimension is large but finite and when the dimension reduction subspace is known. Still, some theoretical guarantees remain open and the computational cost is substantial. In this direction, Xu et al. (2020) develop a three-step estimation procedure for such large conditional quantiles and show $\sqrt{n}$-consistency under a new tail single-index model with unknown indices, at a reduced computational cost. Alternatively, focusing on the prediction of tail events, Aghbalou et al. (2024) propose a new framework which introduces a revised definition of tail conditional independence and provides theoretical guarantees for estimating a low-dimensional projection of covariates that is sufficient for predicting extremes, using inverse regression under standard SIR assumptions.

Recently, a novel method based on the PLS in the extreme regime was introduced by Bousebata et al. (2023). Therein, the authors introduce an inverse single-index model. This means that the covariate is written in terms of the response and some unknown link function (instead of the converse) which results in a reduction of dimension. In particular, under some conditions and within the regular variation framework, this inverse point of view may also be interpreted as a forward single-index model. Next, they propose an estimator of the index which is shown to be consistent with explicit rate and under a deterministic threshold level. The nature of their assumptions is quite different from the standard literature of dimension reduction as neither conditions on the inverse conditional mean $\mathbb{E}(X|Y)$, such as linearity, nor independence or centring of the noise are required. We also refer to Arbel et al. (2024) for an adaptation to the Bayesian framework allowing the introduction of prior information on the index.

In this work, we take up the methodology introduced in Bousebata et al. (2023) for the supervised branch of extreme dimension reduction and convert it to the functional covariate setting. Specifically, we project the predictor/covariate onto

a one-dimensional subspace while capturing most of the information explaining the extreme values of the response variable. To this end, we consider covariates that lie in the infinite-dimensional Hilbert space $H = L^2([0,1])$ and the projection is expressed in terms of the associated inner product. In addition, we assume the more realistic case where the functional covariate is discretized as in Cardot et al. (2003); Crambes et al. (2009), *i.e.*, partially observed on a deterministic time grid. Our methodology would equally apply when the functional covariate is ideally/fully observed (in this case, H could be any separable Hilbert space) and also when $H = \mathbb{R}^d$ with d large but finite, thereby encompassing Bousebata et al. (2023) as a particular case. Following the philosophy of PLS, we maximize the covariance between the projected predictor/covariate and the response conditionally on the response being larger than an increasing threshold. This corresponds to the extreme framework where the response is assumed to be heavy-tailed. As for the covariate, we echo SIR by modelling its relationship to the response through a non-linear inverse regression model which relies on a single index, *i.e.,* a deterministic vector in H. One may again recover a forward single-index model interpretation. Thereby, the method at hand induces an effective reduction of dimension as we provide an estimation of the aforementioned index based on the observations. Using a simple geometrical interpretation, we in fact propose a new family of estimators that generalizes the one in Bousebata et al. (2023) and hence gives the statistician more freedom in choosing which part of the extreme regime to take into account. The consistency (with rate of convergence) of these estimators is established under similar conditions as in Arbel et al. (2024); Bousebata et al. (2023) while extending these two works to the more realistic case of a random threshold given by an extreme order statistic. The assumptions are expressed in the language of regular variation and require minimal conditions about the noise, unlike the standard literature on SIR and forward single-index model. In practice, the finite-sample properties of the method show good results on numerical simulations with synthetic functional data and on high-frequency financial data.

The remainder of the paper is organized as follows. First, we introduce in Section 2 the mathematical framework for the Functional Extreme Partial Least-Squares (FEPLS for short) and, more specifically, the inverse single-index model. The estimators are presented in Section 3 with the associated consistency results within and outside the inverse regression model. Next, the FEPLS finite-sample properties are illustrated in Section 4 on simulated data. Beyond simulations, we illustrate in Section 5 the method's practical value in a financial risk assessment context, using high-frequency asset data to estimate conditional extreme quantiles and compare tail dependence capture against standard dimension-reduction techniques. The corresponding `Python` code is publicly available on the website of the authors. The mathematical details needed for the derivation of our main result are gathered in A.

2. Theoretical framework

Let $H = L^2([0,1],\mathbb{R})$ endowed with its usual scalar product $\langle\cdot,\cdot\rangle$ and Borel σ-field $\mathcal{B}(H)$, and let $(\Omega,\mathcal{A},\mathbb{P})$ be a complete probability space. Consider a random pair (Y,X), where $Y:\Omega\to\mathbb{R}$ and $X:\Omega\to H$ are random variables, *i.e.*, Y is

Borel measurable and X is strongly $\mathbb{P}$-measurable in the sense of Bochner, see Hytonen et al. (2016, Definition 1.1.3).

Let us first focus on the direction defined as

$$w(y) := \underset{\|w\|=1}{\arg\max}\, \mathrm{Cov}\left(\langle w, X\rangle, Y \mid Y \geq y\right), \quad y \in \mathbb{R}, \tag{1}$$

where $\|\cdot\|$ is the norm induced by the inner product $\langle\cdot,\cdot\rangle$ and y is a threshold, ultimately taken large. The expectation underlying the covariance in (1) is understood in the sense of Lebesgue. The interpretation of (1) follows the core idea of PLS but in the extreme regime. Using an optimization framework, we seek the element in H which maximizes the tail-information, by considering the covariance conditionally on the tail event of Y, shared between Y and the projection of X on a single direction.

In the sequel, the cumulative distribution function (cdf) of Y is denoted by F and the survival function by $\bar{F} = 1 - F$. In addition, let us define the tail-moment of a random object W by $m_W(y) := \mathbb{E}(W1_{\{Y\geq y\}})$ whenever it exists. In the case where $W \in H$, we interpret the expectation at hand as the Bochner expectation denoted by $\mathbb{E}_{\mathcal{B}}$. Note that, since we consider a separable space, we could equivalently use the Pettis integral instead. We refer to Hytonen et al. (2016) for more details about the notion.

A key fact is that the solution of the optimization problem (1) is explicit:

Proposition 1. *Suppose that* $\mathbb{E}\left((1+\|X\|)(1+|Y|)\right) < +\infty$. *Then, the unique solution of the optimization problem* (1) *is given by, for any* $y \in \mathbb{R}$ *such that* $\bar{F}(y) > 0$ *and* $v(y) \neq 0$,

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

It readily appears that $w(y)$ is a linear combination of $m_{XY}(y) \in H$ and $m_X(y) \in H$, two tail-moments of the form $m_{X\varphi(Y)}(y)$ associated with the respective weights $\varphi(y) = y$ and $\varphi \equiv 1$. Under the inverse regression model introduced below, both directions asymptotically align with a common index. We thus investigate the use of more general tail-moments by considering

$$w_\varphi(y) = m_{X\varphi(Y)}(y)/\|m_{X\varphi(Y)}(y)\|, \tag{3}$$

where $\varphi : \mathbb{R} \to \mathbb{R}^+$ is some test function such that

$$\varphi \in \mathcal{RV}_\tau(+\infty), \quad \tau \in \mathbb{R}. \tag{4}$$

Let us recall that $\varphi \in \mathcal{RV}_\tau(+\infty)$ is a standard notation meaning that φ is regularly varying at infinity with index τ. The definition is recalled below for the sake of completeness, see Bingham et al. (1989) for further details.

Definition 1 ($\mathcal{RV}_\tau(+\infty)$)**.** A positive measurable function V defined on some neighbourhood $[x_0, +\infty)$ of infinity and such that, for all $x > 0$,

$$\lim_{t\to+\infty} \frac{V(tx)}{V(t)} = x^\tau$$

is said to be regularly-varying at infinity with index τ.

The influence of τ (and thus of φ) on the FEPLS direction $w_\varphi(y)$ is discussed in Remark 4 and illustrated on simulations in Section 4. Before undertaking an inference procedure based on the empirical tail-moments in (3), let us introduce

the four model assumptions (5), (6), (7) and (8). An extreme-value framework is considered where the response variable Y is heavy-tailed to the second order. To this end, denote the generalized inverse function or quantile function of the cdf F by $u \mapsto F^-(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$. The tail quantile function is then $t \geq 1 \mapsto U(t) := F^-(1-1/t)$ and is assumed to belong to the class of second-order regularly-varying functions:

$$U \in 2\mathcal{RV}_{\gamma,\rho}(+\infty), \quad \gamma \in (0,1) \text{ and } \rho \leq 0. \tag{5}$$

Definition 2 ($2\mathcal{RV}_{\gamma,\rho}(+\infty)$)**.** A positive measurable function V defined on some neighbourhood $[x_0, +\infty)$ of infinity is said to be second-order regularly-varying with parameters ($\gamma \in \mathbb{R}, \rho \leq 0$), if, for all $x > 0$,

$$\lim_{t\to+\infty} \frac{1}{A(t)} \left(\frac{V(tx)}{V(t)} - x^\gamma \right) = x^\gamma H_\rho(x) := x^\gamma \int_1^x u^{\rho-1} \mathrm{d}u,$$

where A is an auxiliary function ultimately of constant sign with $A(t) \to 0$ as $t \to +\infty$.

Note that for $y > 0$, $H_\rho(y) = \log(y) 1_{\{\rho=0\}} + \frac{y^\rho - 1}{\rho} 1_{\{\rho<0\}}$. Clearly, $2\mathcal{RV}_{\gamma,\rho}(+\infty) \subset \mathcal{RV}_\gamma(+\infty)$. Besides, by de Haan and Ferreira (2006, Theorem 2.3.9), the following correspondence in terms of regular variation indices holds: $U \in 2\mathcal{RV}_{\gamma,\rho}(+\infty) \iff \bar{F} \in 2\mathcal{RV}_{-1/\gamma,\rho/\gamma}(+\infty)$. A direct consequence of (5) is that $U \in \mathcal{RV}_\gamma(+\infty)$ while $\bar{F} \in \mathcal{RV}_{-1/\gamma}(+\infty)$. The latter first-order condition is equivalent to assuming that the distribution of Y is in the Fréchet maximum domain of attraction with positive tail-index γ, see Bingham et al. (1987, Theorem 1.5.8) and de Haan and Ferreira (2006, Theorem 1.2.1). This domain of attraction consists of heavy-tailed distributions, such as Pareto, Burr and Student distributions, see Beirlant et al. (2004) for further examples. The larger γ is, the heavier the tail. Going beyond the first order, the control of some tail-moment as in Stupfler (2019) requires (5) which is expressed in terms of tail quantile functions. The restriction to $\gamma < 1$ ensures that the first-order moment $\mathbb{E}(Y 1_{\{Y \geq y\}})$ exists for all $y \geq 0$. On the auxiliary functions level, de Haan and Ferreira (2006, Theorem 2.3.3) provides $|A| \in \mathcal{RV}_\rho(+\infty)$ and $|A| \circ (1/\bar{F}) \in \mathcal{RV}_{\rho/\gamma}(+\infty)$.

Next, in order to provide theoretical guarantees on the inference method described in the next section and based on (3), we consider the following (non-linear) inverse single-index functional model,

$$X = g(Y)\,\beta + \varepsilon, \quad \beta \in H, \quad \|\beta\| = 1, \tag{6}$$

with $\varepsilon : \Omega \to H$ a random variable being the noise and an unknown deterministic link function $g : \mathbb{R} \to \mathbb{R}^+$ ultimately behaving like a power function *i.e.*

$$g \in \mathcal{RV}_\kappa(+\infty), \quad \kappa > 0. \tag{7}$$

Model (6) is referred to as an inverse regression model since the functional covariate X is written as a function of the response variable Y. The interest of such inverse models is highlighted in Cook (2007) to compare the theoretical properties of PCA, principal fitted components and SIR, among others. The appeal of model (6) stems from the fact that, for instance when ε is independent of Y and centred, the FEPLS direction (3) coincides with the true index: $w_\varphi(y) = \beta$ for any test function $\varphi \in \mathcal{RV}_\tau(+\infty)$ and any $y \geq 0$ such that $m_{g\varphi(Y)}(y) \in (0, +\infty)$. Heuristically, assuming more generally that Y and ε are dependent but ε has a

small contribution in the tail regime of Y, *i.e.* $X \simeq g(Y)\,\beta$, one expects that $w_\varphi(y) \to \beta$ as $y \to +\infty$. Moreover, in such a situation, model (6) yields the approximate single-index forward model $Y \simeq g^{-1}(\langle \beta, X\rangle)$, where g^{-1} denotes an asymptotic inverse of g, which exists since $\kappa > 0$.

To deal with the conditional expectation involving the tail event $\{Y \geq y\}$ for large y, as well as with the distribution of order statistics, we need some regularity on the distribution of Y. So let us assume that the density $f := F'$ of Y exists and satisfies the von Mises condition

$$f \in \mathcal{RV}_{-1/\gamma-1}(+\infty). \tag{8}$$

By Karamata's theorem, (8) implies $\bar{F} \in \mathcal{RV}_{-1/\gamma}(+\infty)$, consistently with (5).

3. Inference and consistency

Let $(X_i, Y_i)_{1\leq i\leq n}$ be independent copies of $(X, Y) \in H\times\mathbb{R}$. Let $k := k_n \to +\infty$ be an intermediate sequence, *i.e.* integer deterministic such that $k/n \to 0$ as $n \to +\infty$, and consider a deterministic sequence $y_{n,k}$ such that $y_{n,k} \sim U(n/k)$ *i.e.* $y_{n,k}/U(n/k) \to 1$ as $n \to +\infty$. Denoting the order statistics of the sample $(Y_i)_{1\leq i\leq n}$ by $Y_{1,n} \leq \cdots \leq Y_{n,n}$, one may see that the empirical version of $U(n/k)$ is $Y_{n-k+1,n}$. In the sequel and for $\alpha > 0$, we denote by $\mathcal{C}^{0,\alpha}([0,1],\mathbb{R})$ the space of α-Hölder continuous functions, *i.e.*, the set of $f : [0,1] \to \mathbb{R}$ such that there exists a finite constant $c > 0$ satisfying

$$|f(x) - f(y)| \leq c|x - y|^\alpha \qquad \text{for all } x, y \in [0,1].$$

For $h \in \mathcal{C}^{0,\alpha}([0,1],\mathbb{R})$, the Hölder norm of h is defined as $\|h\|_{0,\alpha} := \sup_{x\in[0,1]} |h(x)| + \sup_{x\neq y} |h(x) - h(y)|/|x - y|^\alpha$, with the convention $\|h\|_{0,\alpha} = +\infty$ whenever $h \notin \mathcal{C}^{0,\alpha}([0,1],\mathbb{R})$. Concerning the noise, denoting by $\varepsilon_i := X_i - g(Y_i)\beta$, $1 \leq i \leq n$, the noise associated with the i-th observation under model (6), we require the existence of $q > 2$ and $\alpha_e > 0$ independent of n such that

$$\limsup_{n\to+\infty} \sup_{y\geq 0} \mathbb{E}\left(\|\varepsilon_1\|_{0,\alpha_e}^q \mid Y_{n-k+1,n} = y\right) < +\infty. \tag{9}$$

This uniform control of the q-mean of $\|\varepsilon_1\|_{0,\alpha_e}$ given the extreme order statistic $Y_{n-k+1,n}$ implies that $\mathbb{E}(\|\varepsilon\|^q) < +\infty$. Since $q > 2$, one has $\mathbb{E}(\|\varepsilon\|) < +\infty$ and, by Hytonen et al. (2016, Proposition 1.2.2), ε is also Bochner-integrable since ε is strongly $\mathbb{P}$-measurable by definition.

For any random variable W, the empirical counterpart of the tail-moment $m_W(y)$ (whenever it exists) is defined by

$$\hat{m}_W(y) := \frac{1}{n}\sum_{i=1}^n W_i 1_{\{Y_i\geq y\}},$$

based on the n-sample $(W_i, Y_i)_{1\leq i\leq n}$. Two situations are investigated. When the full functions $X_1, \ldots, X_n$ are observed, it is then possible to estimate $w_\varphi(y)$ by replacing the tail moment in (3) by its empirical counterpart defined as

$$\hat{\beta}_\varphi(y) := \frac{\hat{m}_{X\varphi(Y)}(y)}{\|\hat{m}_{X\varphi(Y)}(y)\|} \text{ with } \hat{m}_{X\varphi(Y)}(y) = \frac{1}{n}\sum_{i=1}^n X_i\varphi(Y_i)1_{\{Y_i\geq y\}}.$$

The FEPLS estimator $\hat{\beta}_\varphi(y)$ is thus a linear combination of the X_i's whose associated Y_i's are located in the tail of $\bar{F}$.

Let us now consider the more realistic situation where each function X_i, $i \in \{1,\dots,n\}$, is observed on a finite deterministic grid $0 \le t_1 < \cdots < t_N \le 1$ only. Let $\{I_1,\dots,I_N\}$ be a partition of $[0,1]$ into intervals such that $t_j \in I_j$ for any $j \in \{1,\dots,N\}$ and

$$\max_{1\le j\le N} |I_j| \le c_0/N, \tag{10}$$

for some constant $c_0 \ge 1$ independent of N, where $|I_j|$ denotes the length of I_j; the regular design $t_j = (2j-1)/(2N)$ with $I_j = [(j-1)/N, j/N)$ is the prototypical example. The discrete record $(X_i(t_1),\dots,X_i(t_N))$ is then summarized by the piecewise-constant reconstruction

$$\tilde{X}_i := \sum_{j=1}^{N} X_i(t_j)\, 1_{I_j} \in H.$$

Note that, under model (6) with $\beta \in \mathcal{C}^{0,\alpha_b}([0,1],\mathbb{R})$ and $\varepsilon \in \mathcal{C}^{0,\alpha_e}([0,1],\mathbb{R})$ a.s., the trajectories of X are continuous, so that the pointwise evaluations $X_i(t_j)$ are well defined. In this context, the estimator of β is given by

$$\tilde{\beta}_\varphi(y) := \frac{\hat{m}_{\tilde{X}\varphi(Y)}(y)}{\|\hat{m}_{\tilde{X}\varphi(Y)}(y)\|} \text{ with } \hat{m}_{\tilde{X}\varphi(Y)}(y) = \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i \varphi(Y_i) 1_{\{Y_i \ge y\}}.$$

3.1. **Consistency under the inverse single-index model.** Let us now state our consistency result under the inverse single-index model.

Theorem 1. *Assume $\beta \in \mathcal{C}^{0,\alpha_b}([0,1],\mathbb{R})$ for some $\alpha_b \in (0,1]$ and that (6) holds. Assume*

$$0 < 2(\kappa+\tau)\gamma < 1, \tag{11}$$

$$q\kappa\gamma > 1, \tag{12}$$

along with (4), (5), (7), (8), (9) and the grid condition (10). Assume that φ and g are continuously differentiable in a neighbourhood of infinity such that $t(\varphi g)'(t)/(\varphi g)(t) \to \tau+\kappa$ as $t \to +\infty$. Then, for any intermediate sequence $k := k_n \to +\infty$ such that $\sqrt{k}A(n/k) = O(1)$, any deterministic sequence $y_{n,k} \sim U(n/k)$ and any deterministic sequence $N := N_n \to +\infty$, it holds that:

$$\|\tilde{\beta}_\varphi(Y_{n-k+1,n}) - \beta\| = O_{\mathbb{P}}\big(\delta_{n,k} \vee N^{-\alpha_b}\big) \xrightarrow[n\to+\infty]{\mathbb{P}} 0,$$

where $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$.

Let us highlight that the sub-rate $\delta_{n,k}$ is a regularly-varying function of k/n with index $\gamma\kappa - 1/q > 0$. It only depends on the tail distribution of Y, the integrability order of the noise ε and the asymptotic behaviour of the deterministic link function g.

We now state some comments about the set of assumptions and their implications.

Remark 1 (The noise ε does not contribute to the extremes). Assumption (7) combined with $U \in \mathcal{RV}_\gamma(+\infty)$ implies that $g(Y)$ is heavy-tailed with tail-index $\gamma_{g(Y)} := \kappa\gamma$. Similarly, when $\tau > 0$, (4) and $U \in \mathcal{RV}_\gamma(+\infty)$ yield that $\varphi(Y)$ is heavy-tailed with tail-index $\gamma_{\varphi(Y)} := \tau\gamma$. Recall that (9) entails $\mathbb{E}(\|\varepsilon\|^q) < +\infty$ which may be interpreted as an assumption on the tail of $\|\varepsilon\|$. It is satisfied, for

instance, by distributions with exponential-like tails such as Gaussian, Gamma or Weibull distributions. Heuristically, it would imply that the tail-index associated with $\|\varepsilon\|$ is such that $\gamma_{\|\varepsilon\|} < 1/q$. Condition (12) thus imposes that $\gamma_{g(Y)} > \gamma_{\|\varepsilon\|}$, meaning that $g(Y)$ has a heavier right tail than $\|\varepsilon\|$. Under model (6), one has $\gamma_{\|X\|} = \gamma_{g(Y)}$ and thus the tail behaviours of $\langle \beta, X\rangle$ and $\|X\|$ are driven by $g(Y)$, which is the desired property.

Remark 2 (Rate of convergence). (i) The regime $y_{n,k} \sim U(n/k)$ entails $n\bar{F}(y_{n,k}) \sim k \to \infty$ as $n \to \infty$ and it still holds when $y_{n,k}$ is replaced by its empirical counterpart, namely $Y_{n-k+1,n}$. Hence the expected number of tail observations, *i.e.* of observations above the (random or deterministic) threshold, grows with the sample size. This roughly ensures that the threshold does not grow too fast so that there are enough data points in the inference scheme. The number of such extreme points is ruled by the usual condition $\sqrt{k}A(n/k) = O(1)$ in extreme-value theory, which is fulfilled when $k \sim cn^{-2\rho/(1-2\rho)}$, $\rho < 0$, $c > 0$, provided $|A(t)|$ is asymptotically proportional to t^ρ, see de Haan and Ferreira (2006, Equation (3.2.10)).

(ii) Condition (11) implies the existence of the second moment of $\varphi g(Y)\mathbf{1}\{Y \geq y\}$ for all $y \geq 0$. To this end, one should pick $\tau < 1/(2\gamma) - \kappa$, this upper bound being either positive or negative.

(iii) The constraint (12) is the translation in the regular variation language of the fact that $\delta_{n,k} \to 0$ as $n \to \infty$ regardless of the underlying slowly-varying part of g and $\bar{F}$. Under the choice of k in (i), the rate $\delta_{n,k}$ is of order $n^{(1/q-\gamma\kappa)/(1-2\rho)}$, up to slowly-varying factors.

Remark 3 (Comparison with empirical-process inverse regression). The authors in Portier (2016) and Aghbalou et al. (2024) also study the conditional mean of the covariate given the response, but through an empirical-process lens. One may roughly distinguish three essential differences. First, their setting is finite-dimensional, whereas our work is functional. Next and most importantly, they describe the entire quantile-indexed process and prove a functional central limit theorem, at the parametric rate in the central regime of Portier (2016) and at the tail rate in the extreme regime of Aghbalou et al. (2024). We do not target the whole process: the estimator is evaluated at a single data-driven extreme level, where the effective information is carried by the largest observations rather than by the full sample. Our conclusion is accordingly a convergence in probability with an explicit rate and not weak convergence. Finally, concerning the model and tools, the empirical-process route rests on a linearity condition on the covariate and on bounded index weights. Our model assumes no linearity, centring or independence of the noise, and replaces them by a heavy-tail, regularly-varying structure, at the price of a constraint linking the admissible growth of the weight to the tail index. The machinery is correspondingly that of extreme-value theory, regular variation and order statistics, rather than Donsker classes and the functional delta method.

Example 1. Let Y be Burr distributed, with $\bar{F}(y) = (1+y^{-\rho/\gamma})^{1/\rho}$ for $y \geq 0$, $\gamma \in (0,1)$ and $\rho < 0$, so that (5) and (8) hold. Let $g(y) = y^\kappa$ with $\kappa > 0$ and $\varphi(y) = y^\tau$ with $\tau \in \mathbb{R}$, so that (7) and (4) hold. Finally, let ε be independent of Y and distributed as $\sigma B^H + \mu$, where $\sigma > 0$, $\mu : [0,1] \to \mathbb{R}$ is a deterministic Lipschitz function and B^H is a fractional Brownian motion on $[0,1]$ with Hurst parameter

$H \in (0,1)$. The trajectories of B^H are almost surely α_e-Hölder continuous for any $\alpha_e < H$ and, by Fernique's theorem, $\mathbb{E}\left(\|B^H\|_{0,\alpha_e}^q\right) < +\infty$ for all $q > 0$, so that, by independence, condition (9) holds for every $q > 2$. Conditions (11) and (12) then reduce to $-\kappa < \tau < 1/(2\gamma) - \kappa$ and $q > 1/(\kappa\gamma)$: for instance, with $\gamma = 1/3$ and $\kappa = 1$, every $\tau \in (-1, 1/2)$ is admissible. Consequently, for any $\beta \in \mathcal{C}^{0,\alpha_b}([0,1],\mathbb{R})$ with $\|\beta\| = 1$ and $\alpha_b \in (0,1]$, all the assumptions of Theorem 1 are satisfied, and its conclusion holds. In this setting, model (6) also admits the forward interpretation $Y \simeq \langle \beta, X\rangle^{1/\kappa}$ in the tail regime. A variant of this example, in which the noise scale is allowed to depend on Y, underlies the simulation design of Section 4.

3.2. **A model-free asymptotic result on the empirical tail-moment.** The consistency established in Theorem 1 and in Proposition 3 of A ultimately relies only on the behaviour of the empirical tail-moment $\hat{m}_W$, the inverse single-index model (6) being used solely to identify the limiting direction β. We now isolate this ingredient for a generic Hilbert space H and $W \in H$, and drop (6) together with any assumption on the direction of the conditional mean.

In the following, $\mathbb{E}_{\mathcal{F}}$ denotes the conditional expectation given a σ-field $\mathcal{F}$ and, if moreover the integrand is Banach-valued, we denote by $\mathbb{E}_{\mathcal{F},\mathcal{B}}$ the Bochner version of the conditional expectation, see (Hytonen et al., 2016) for the technical details.

Let $W : \Omega \to H$ be Bochner-integrable. By construction, W may be decomposed along its conditional mean as

$$W = \mu(Y) + \xi, \qquad \mu(Y) := \mathbb{E}_{\mathcal{F},\mathcal{B}}(W) \text{ with } \mathcal{F} = \sigma(Y), \qquad \xi := W - \mu(Y), \tag{13}$$

where $\mu : \mathbb{R} \to H$ denotes the associated regression function and $\mathbb{E}_{\mathcal{F},\mathcal{B}}(\xi) = 0$, so that $m_W(y) = m_{\mu(Y)}(y)$ by conditioning. The decomposition (13) is merely the definition of the conditional mean: it imposes no model on (W, Y), in particular no linearity, and μ is the forward regression of W given Y. We denote by $s(y) := \|\mu(y)\|$ the norm of the conditional mean and by $a(y) := \mu(y)/s(y)$ its direction, for y large enough so that $s(y) > 0$. No assumption is made on $a(y)$: it may rotate arbitrarily and need not converge. Accordingly, the natural target is the tail-moment $m_W(y)$ itself rather than a fixed direction, and the limiting direction is expressed through

$$\bar{a}(y) := \frac{m_W(y)}{\|m_W(y)\|}, \qquad \text{whenever } m_W(y) \neq 0. \tag{14}$$

Here $\bar{a}(y)$ is the direction of the tail-moment, a tail average of the pointwise directions $a(t)$, $t \geq y$, that need not coincide with any of them. In the spirit of (7), we assume that $s : \mathbb{R} \to \mathbb{R}^+$ is ultimately positive with

$$s \in \mathcal{RV}_\lambda(+\infty), \quad \lambda > 0, \qquad 2\lambda\gamma < 1, \tag{15}$$

and, concerning the residual ξ, that there exists $q > 2$ such that

$$\sup_y \mathbb{E}\left(\|\xi\|^q \mid Y = y\right) < +\infty, \tag{16}$$

where the supremum is over the support of Y. By Lemma 2, this implies $\sup_y \mathbb{E}\left(\|\xi_1\|^q \mid Y_{n-k+1,n} = y\right) < +\infty$ uniformly in n and k, which is the form

used in the proofs, together with the index condition, in the same vein as (12),

$$q\lambda\gamma > 1. \tag{17}$$

Note that (15) and (17) are compatible for any $q > 2$ since they read $1/(q\gamma) < \lambda < 1/(2\gamma)$. The upper bound $2\lambda\gamma < 1$ ensures that the second tail-moment $m_{s^2(Y)}$ is finite, which controls the fluctuation of $\hat{m}_W$; it is also the index condition required by Stupfler (2019, Theorem 2) with $a = \lambda$, used in regime (i) of Proposition 2 below, where the norm s is moreover assumed continuously differentiable in a neighbourhood of infinity. Beyond the conditional centring $\mathbb{E}_{\mathcal{F},\mathcal{B}}(\xi) = 0$, which holds by construction, and the moment control (16), no assumption is made on the residual ξ; in particular, it is not assumed independent of Y. Finally, we set $c_W(y) := \|m_W(y)\|/m_{s(Y)}(y) \in [0,1]$ and consider the non-degeneracy condition

$$\liminf_{y\to+\infty} c_W(y) > 0; \tag{18}$$

since $U(n/k) \to +\infty$, this implies $\liminf_{n\to+\infty} c_W(U(n/k)) > 0$, the form used in the proof. The quantity $c_W(y)$ measures the loss of norm due to the averaging of the directions of μ over the tail: $c_W \equiv 1$ when μ keeps a constant direction on $[y,+\infty)$, while $c_W(y) \to 0$ only when the directions $\{a(t) : t \geq y\}$ cancel asymptotically. Condition (18) rules out this cancellation, and nothing more.

Proposition 2. *Under* (13)*, assume* (15)*,* (16) *and* (17)*, and that* (8) *holds with* $\gamma \in (0,1)$*, so that* $\bar{F} \in \mathcal{RV}_{-1/\gamma}(+\infty)$*. Let* $k := k_n \to +\infty$ *be a deterministic integer sequence with* $k/n \to 0$*. Then* $\|m_W(Y_{n-k+1,n})\| \leq m_{s(Y)}(Y_{n-k+1,n})$ *and*

$$\frac{\hat{m}_W(Y_{n-k+1,n}) - m_W(Y_{n-k+1,n})}{m_{s(Y)}(Y_{n-k+1,n})} \longrightarrow 0 \quad \text{in } H,$$

in either of the two following alternative regimes:

(i) *if* (5) *holds,* $\sqrt{k}A(n/k) = O(1)$ *and* s *is continuously differentiable in a neighbourhood of infinity with* $ts'(t)/s(t) \to \lambda$ *as* $t \to +\infty$*, then the convergence holds in probability, at rate* $O_{\mathbb{P}}(1/\sqrt{k})$*;*

(ii) *if* $k_n \geq n^\theta$ *eventually for some* $\theta > 2\lambda\gamma$*, with none of the assumptions of regime* (i) *required, then the convergence holds almost surely.*

If moreover (18) *holds, then* $\mathbb{P}\big(m_W(Y_{n-k+1,n}) \neq 0\big) \to 1$ *in regime* (i)*, while* $m_W(Y_{n-k+1,n}) \neq 0$ *for* n *large enough almost surely in regime* (ii)*. Moreover,*

$$\frac{\hat{m}_W(Y_{n-k+1,n})}{\|\hat{m}_W(Y_{n-k+1,n})\|} - \bar{a}(Y_{n-k+1,n}) \longrightarrow 0 \qquad \text{and} \qquad \frac{\|\hat{m}_W(Y_{n-k+1,n})\|}{\|m_W(Y_{n-k+1,n})\|} \longrightarrow 1,$$

in the same mode as the first claim.

First, Proposition 2 contains the inverse single-index model by taking $W = X\varphi(Y)$. When the noise is conditionally centred, $\mathbb{E}_{\mathcal{F},\mathcal{B}}(\varepsilon) = 0$, the regression function is $\mu(Y) = \varphi(Y)g(Y)\beta$, so that $s = \varphi g \in \mathcal{RV}_{\tau+\kappa}$ (hence $\lambda = \tau + \kappa$), $\xi = \varphi(Y)\varepsilon$, $\bar{a} \equiv \beta$ and $c_W \equiv 1$; assumptions (15), (16) and (17) then read $2(\tau+\kappa)\gamma < 1$, the conditional moment bound on $\|\varphi(Y)\varepsilon\|$ (a joint condition on (φ,ε) which, for $\tau > 0$, is not implied by (9) alone) and $q(\tau+\kappa)\gamma > 1$, the first being exactly (11), while (18) holds automatically, and the proposition recovers the consistency of $\hat{\beta}_\varphi$ towards β. The unweighted choice $W = X$ ($\varphi \equiv 1$) is the special case $s = g$, $\lambda = \kappa$. In parallel, the model-free statement neither centres

nor constrains the noise beyond (16), it holds for an arbitrary Bochner-integrable W in any separable Hilbert space, and it lets the conditional-mean direction $a(\cdot)$ vary in the tail, the limiting direction $\bar{a}$ being then a tail average rather than a fixed vector. Next, the rate in regime (i) is that of estimating the tail-moment of the conditional mean alone: since ξ is conditionally centred by construction, its contribution $\hat{m}_\xi(Y_{n-k+1,n})$, measured relative to $m_{s(Y)}$, is of order $o_{\mathbb{P}}(k^{-1/2})$, hence negligible against the fluctuation of $\hat{m}_{\mu(Y)}$, and the residual enters the rate only through the noise-free term. This contrasts with Lemma 5, where the noise of (6) is not conditionally centred: its tail-moment then carries a non-vanishing bias of order $\delta_{n,k}$, which governs the rate whenever $\delta_{n,k} \gg k^{-1/2}$, that is, when k grows fast enough. Finally, the residual ξ is required to be lighter-tailed than the conditional mean: $s(Y)$ is heavy-tailed with index $\lambda\gamma$, the moment bound (16) caps the tail index of $\|\xi\|$ at $1/q$, and (17) places $1/q$ below $\lambda\gamma$. The heavier the conditional-mean norm, the fewer residual moments are needed; this is the model-free counterpart of the constraint $\gamma_{g(Y)} > \gamma_{\|\varepsilon\|}$ discussed after Theorem 1.

Remark 4 (On the test function φ). The optimisation problem of Proposition 1 naturally leads to the weight $\varphi(y) = y$, but there is no reason to restrict to it, so we replace it by a general $\varphi \in \mathcal{RV}_\tau$ in (3). The function φ tunes the weight applied to each selected tail observation. It is not entirely free: the conditions of Theorem 1 restrict τ to $-\kappa < \tau < 1/(2\gamma) - \kappa$. In particular, $\tau = 0$, and thus the unweighted choice $\varphi \equiv 1$, is admissible as soon as $2\kappa\gamma < 1$. The main point is that the choice of φ does not affect the estimation of β at first order. Under the inverse single-index model, the limiting direction is β for every admissible φ, because the noise does not contribute to the extremes and the conditional mean is aligned with β in the tail. The rate $\delta_{n,k}$ is not affected either and in parallel, condition (12) only involves κ. Hence no admissible φ gives a wrong direction or a slower rate. The function φ still keeps two effects. First, it can change the direction, but only outside the single-index model, as in Proposition 2. The element $w_\varphi(y)$ points in the direction of the average of $\mathbb{E}(X \mid Y = t)$ over the tail $t \geq y$, each level t being weighted by $\varphi(t)$, and this average depends on φ only when the direction of $\mathbb{E}(X \mid Y = t)$ varies with t. In contrast, under the single-index model this direction is constant and equal to β. Second, φ changes the finite-sample behaviour, since it controls the weight given to the most extreme observations, hence the variance and the effective number of exceedances. In this respect φ is a tuning parameter like the number k of order statistics, without changing the rate. Finally, this invariance gives a simple check of the model: since the limiting direction does not depend on φ under the single-index model, a direction that changes substantially with φ indicates that this model does not hold.

Remark 5 (Bivariate functional covariates). All our results only require H to be a separable Hilbert space. A bivariate functional covariate $X = (X^{(1)}, X^{(2)})$ is thus a special case, with $H = L^2([0,1], \mathbb{R}^2)$; so is a covariate on a two-dimensional domain, with $H = L^2([0,1]^2)$. Propositions 1, 2 and 3 still hold, as their proofs use only the inner product of H and the noise through its norm $\|\varepsilon\|$. Proposition 2 is moreover already stated for an arbitrary separable Hilbert space. Theorem 1 carries over by the same arguments, the reconstruction on a grid being carried out component-wise, with the rate then being the worst of the two component

rates. Component-wise, the inverse single-index model (6) writes as $X^{(j)} = g(Y)\beta^{(j)} + \varepsilon^{(j)}$, with one link g, one index $\beta = (\beta^{(1)}, \beta^{(2)})$ and a common tail-index $\kappa\gamma$. With distinct tail-indices $g_j \in \mathcal{RV}_{\kappa_j}$, the model no longer holds and Proposition 2 applies: $\hat{\beta}_\varphi$ still converges but to the direction of the heavier component, the other not contributing to the extremes. Estimating both, or several indices, is a multi-index problem beyond model (6).

4. Illustration on simulated data

The performance of our method is assessed by a Monte Carlo simulation experiment with 500 independent repetitions. The ambient space is $H = L^2([0,1])$ endowed with its usual inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$. Functional observations are discretized on a regular grid of length $N \in \{101, 1001\}$ with nodes $t_j = (j-1)/(N-1)$ for $1 \le j \le N$. To implement the L^2-inner product in the discretized setting, we consider the diagonal positive definite matrix $\Psi \in \mathbb{R}^{N\times N}$ given by $\Psi_{j,j'} = |I_j|\, 1_{\{j=j'\}}$ for all $(j, j') \in \{1, \dots, N\}^2$, so that $\langle \tilde{X}_i, \tilde{X}_\ell\rangle = \mathcal{X}_i^\top \Psi\, \mathcal{X}_\ell$ exactly, where $\mathcal{X}_i := (X_i(t_1), \dots, X_i(t_N))^\top$, $1 \le i \le n$. For any discretized functions $u = (u(t_j))_{1\le j\le N}, v = (v(t_j))_{1\le j\le N} \in \mathbb{R}^N$, the quadratic form $\langle u, v\rangle_\Psi := u^\top \Psi v$ induces an inner product and all subsequent computations use $\|\cdot\|_\Psi := \sqrt{\langle\cdot,\cdot\rangle_\Psi}$ as the induced norm.

4.1. Experimental design.

We draw a data sample of $n = 500$ independent replications (X_i, Y_i) of (X, Y) using the following scheme:

- Y has a Burr distribution, *i.e.*, $\bar{F}(y) = (1+y^{-\rho/\gamma})^{1/\rho} \in 2\mathcal{RV}_{-1/\gamma,\rho/\gamma}(+\infty)$ with $y \ge 0$, $\gamma \in (0,1)$ being the tail-index and $\rho < 0$ being the second-order parameter. Here, $\gamma \in \{1/3, 1/2, 9/10\}$ while $\rho \in \{-2\gamma, -\gamma/2\}$.
- $\varepsilon \mid Y = y$ has the same distribution as $\sigma(y)B_y^H + \mu(y)$ where B_y^H is the fractional Brownian motion (fBm) on $[0,1]$ with conditional Hurst parameter $H(y) \in (0,1)$, $\sigma(y)$ is the conditional noise deviation and $\mu(y)$ the conditional mean.
- $\beta : t \in [0,1] \mapsto \sqrt{2}\sin(2\pi t)$ and $g : y \in \mathbb{R}^+ \mapsto y^\kappa$ with $\kappa \in \{1, 3/2, 2\}$.

Hence, conditionally on $Y = y$, the noise process based on B_y^H is centered Gaussian with conditional covariance function

$$(s,t) \mapsto \frac{\sigma^2(y)}{2}\left(t^{2H(y)} + s^{2H(y)} - |t-s|^{2H(y)}\right).$$

Assumption (9) thus holds for all $q > 0$. Note that $H(y) = 1/2$ yields the Brownian motion B_y with mean $\mu(y)$ and variance $\sigma^2(y)$.

Once both the response variable Y and the noise ε are simulated, and the deterministic functions β and g are chosen, the covariate sampling $\{\tilde{X}_1, \dots, \tilde{X}_n\}$ is readily derived from (6). The choice of σ is guided by the following reasonable requirement: One should ensure that the contribution in (6) of ε does not overwhelm $g(Y)$, but also conversely, in order to avoid a trivial case where the noise is not impacting. To this end, we propose to pick $\sigma(y) = g(y)/10$. Moreover, we fix $(H, \mu) \equiv (1/3, 200)$ regardless of $Y = y$.

The FEPLS estimator is implemented with $\varphi(y) = y^\tau$ as test function. The selected values of τ depend on the pair (γ, κ) in order to fulfil condition (11). They are given in Table 1.

	$\gamma = 1/3$	$\gamma = 1/2$	$\gamma = 9/10$
$\kappa = 1$	$\{0, -1, -2\}$	$\{-1, -2, -3\}$	$\{-1, -2, -3\}$
$\kappa = 3/2$	$\{-1, -2, -3\}$	$\{-1, -2, -3\}$	$\{-1, -2, -3\}$
$\kappa = 2$	$\{-1, -2, -3\}$	$\{-2, -3, -4\}$	$\{-2, -3, -4\}$

TABLE 1. Selected values of τ in the nine simulated situations.

4.2. **Selection of the threshold.** We consider the following data-driven selection rule of $k \in \{1, \dots, n\}$ based on the maximization of the sample covariance between Y and $\langle \tilde{\beta}_\varphi(Y_{n-k+1,n}), \tilde{X} \rangle$ when $Y \geq Y_{n-k+1,n}$:

$$\tilde{k} := \underset{5 \leq k \leq n/5}{\arg\max}\, r(k),$$

with $r(k)$ being

$$\frac{1}{k}\sum_{i=1}^{k} Y_{n-i+1,n} \langle \tilde{\beta}_\varphi(Y_{n-k+1,n}), \tilde{X}_{(n-i+1,n)} \rangle - \frac{1}{k}\sum_{i=1}^{k} Y_{n-i+1,n} \frac{1}{k}\sum_{i=1}^{k} \langle \tilde{\beta}_\varphi(Y_{n-k+1,n}), \tilde{X}_{(n-i+1,n)} \rangle,$$

and where $\tilde{X}_{(n-i+1,n)}$ denotes the concomitant of $Y_{n-i+1,n}$, *i.e.*, the random variable $\tilde{X}_s$ with $s \in \{1, \dots, n\}$ being the unique index such that $Y_s = Y_{n-i+1,n}$. Let us highlight that the above selection rule imposes $\tilde{k} \geq 5$ in order to prevent instabilities of estimates built on too few data points.

Remark 6. Let us denote $\mathcal{X}_i = (X_i(t_1), \dots, X_i(t_N))^\top$. For any $i, \ell \in \{1, \dots, n\}$, one has $\langle \tilde{X}_i, \tilde{X}_\ell \rangle = \langle \mathcal{X}_i, \mathcal{X}_\ell \rangle_\Psi$ so that

$$\begin{aligned}
\|\hat{m}_{\tilde{X}\varphi(Y)}(y)\|^2 &= \frac{1}{n^2}\sum_{i,\ell=1}^{n} \langle \tilde{X}_i, \tilde{X}_\ell \rangle \varphi(Y_i)\varphi(Y_\ell) 1_{\{Y_i \geq y\}} 1_{\{Y_\ell \geq y\}} \\
&= \frac{1}{n^2}\sum_{i,\ell=1}^{n} \langle \mathcal{X}_i, \mathcal{X}_\ell \rangle_\Psi \varphi(Y_i)\varphi(Y_\ell) 1_{\{Y_i \geq y\}} 1_{\{Y_\ell \geq y\}} \\
&= \|\mathcal{V}_\varphi(y)\|_\Psi^2
\end{aligned}$$

where $\mathcal{V}_\varphi(y) := \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i \varphi(Y_i) 1_{\{Y_i \geq y\}} \in \mathbb{R}^N$ can be interpreted as a sampled version of $\hat{m}_{\tilde{X}\varphi(Y)}(y)$. Now, to compute $\langle \tilde{\beta}_\varphi(y), \tilde{X}_s \rangle$, we write, for the numerator,

$$\langle \hat{m}_{\tilde{X}\varphi(Y)}(y), \tilde{X}_s \rangle = \frac{1}{n}\sum_{i=1}^{n} \langle \tilde{X}_i, \tilde{X}_s \rangle \varphi(Y_i) 1_{\{Y_i \geq y\}} = \frac{1}{n}\sum_{i=1}^{n} \langle \mathcal{X}_i, \mathcal{X}_s \rangle_\Psi \varphi(Y_i) 1_{\{Y_i \geq y\}},$$

the denominator being already given by: $\|\hat{m}_{\tilde{X}\varphi(Y)}(y)\| = \|\mathcal{V}_\varphi(y)\|_\Psi$.

4.3. **Results.** Figure 1 displays the dot product curves $k \in \{1, \dots, 250\} \mapsto \langle \tilde{\beta}_\varphi(Y_{n-k+1,n}), \beta \rangle$ averaged over the 500 Monte Carlo replications of the nine considered simulated situations.

Remark 7. Computation of $\langle \tilde{\beta}_\varphi(y), \beta \rangle$. Let us focus on the numerator:

$$\langle \hat{m}_{\tilde{X}\varphi(Y)}(y), \beta \rangle = \frac{1}{n}\sum_{i=1}^{n} \langle \tilde{X}_i, \beta \rangle \varphi(Y_i) 1_{\{Y_i \geq y\}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N} X_i(t_j) \langle 1_{I_j}, \beta \rangle \varphi(Y_i) 1_{\{Y_i \geq y\}}$$

with

$$\langle 1_{I_j}, \beta \rangle = \int_{I_j} \beta(u) \mathrm{d}u =: \bar{\beta}_j,$$

and thus $\langle \hat{m}_{\tilde{X}\varphi(Y)}(y), \beta \rangle = \mathcal{V}_\varphi(y)^\top \bar{\beta}$.

It appears that the estimated index $\tilde{\beta}_\varphi(Y_{n-k+1,n})$ is close to the true one only for small values of k, or equivalently for large values of the threshold $Y_{n-k+1,n}$. This phenomenon is coherent with the fact that $\langle \beta, X \rangle$ is designed to capture the tail-information of Y. Let us highlight that the range of accurate values for k depends on the difficulty of the estimation problem. The higher γ and κ, the heavier the tail of $g(Y)$, and the easier the estimation of β is. In practice, the selection of k is performed using the above described selection procedure since β is unknown and therefore Figure 1 cannot be used in real data situations.

Finally, the mean over the 500 replications of the curves $t \in [0,1] \mapsto \tilde{\beta}_\varphi(Y_{n-\tilde{k}+1,n})(t)$ is compared to the true one $t \in [0,1] \mapsto \beta(t)$ on Figure 2 in the case where $\tau = -2$. It appears that the selected values of k yield very accurate estimations on all nine considered situations whatever the values of γ or κ. Pushing further, Figure 3 is analogous to Figure 2 when the underlying grid is of size $N = 1001$. This aims at better reflecting the high-dimensional regime, since the sample size is $n = 500$. One may observe a slight degradation of the estimation in comparison to Figure 2 mostly when $\kappa = 1$. It appears again that the higher γ (heavier tail), the easier the inference becomes. Similarly, the larger κ, the more accurate the inference. Indeed, in view of the inverse regression model $X = g(Y)\beta + \varepsilon$ and the composition rule for regularly-varying functions, κ also contributes to the leading tail-index.

5. Application on real data

We illustrate the FEPLS method on a financial application where covariates naturally take a functional form due to high-frequency sampling. Introduce the list of assets[1]

$$\mathcal{A} := \{\text{S\&P500}, \text{DAX30}, \text{Nikkei225}, \text{ETH}, \text{BTC}\}$$

where ETH and BTC respectively stand for Ethereum and Bitcoin, everything understood in their canonical currency. We consider the daily maximum log-return of some asset $a_1 \in \mathcal{A}$ as the response variable Y, and the intraday log-return curve of some asset $a_2 \in \mathcal{A}$, possibly the same as a_1, as the functional covariate X. Any asset at disposal is recorded at one-minute frequency over the period 2013–2017.

The principal motivation for applying FEPLS in this context is to facilitate the inference of conditional extreme risk measures. By replacing the functional covariate with a univariate projection $Z = \langle \tilde{X}, \tilde{\beta} \rangle_\Psi$, the estimation of quantities such as the conditional Value-at-Risk (VaR) or the conditional tail index becomes computationally simpler and statistically more stable in the extreme regime. The core question is then whether the one-dimensional projection retains the covariate information that is relevant for predicting extreme responses. To answer it, we compare the performance of FEPLS against natural competitors on the ability to

[1]Available at https://github.com/FutureSharks/financial-data/tree/master/pyfinancialdata/data/

capture tail dependence, based on the particular pair $(a_1, a_2) = (\text{DAX30}, \text{ETH})$ and then on all the possible asset pairs from $\mathcal{A}$. Ultimately and as a practical tool for extreme risk assessment, we include the estimation of extreme quantiles.

Concerning the tail tuning parameter of $\varphi(y) = y^\tau$, we arbitrarily set throughout this section $\tau = -1$.

5.1. **Sample Construction.** The raw price series of both a_1, a_2 are transformed into log-returns $R_t = \log(P_t/P_{t-1})$ where P_t denotes the price at time t. The two resulting series are aligned by date so that only contemporaneous observations are kept. This yields approximately the same number of one-minute observations for both indices over the common period Jan 1, 2013 - Dec 31, 2017. The discretized functional covariate sample $\{\mathcal{X}_i\}_{1\le i\le n}$ is constructed by cutting the a_1 log-return series into consecutive blocks of $N = 1440$ minutes (24 hours). Each block corresponds to one trading day, and the vector $\mathcal{X}_i = \big(X_i(t_1), \dots, X_i(t_N)\big)^\top$ contains the intraday log-returns of a_1 on day $2i-1$. The response sample $\{Y_i\}_{1\le i\le n}$ is obtained from the a_2 log-returns: for each even day $2i$, we take the maximum of the absolute log-returns over that day,

$$Y_i = \max_{t\in\text{day } 2i} |R_t^{a_2}|.$$

No missing data were detected and the final sample size n is among $\{265, 365, 433, 442\}$. It may even be equal to 901 when $a_1 = a_2 = \text{BTC}$. As for our particular study case, the pair DAX30-ETH consists of $n = 265$ observations.

Remark 8 (Temporal dependence). Although the theoretical guarantees of Section 2 assume i.i.d. observations, financial returns typically exhibit serial dependence and volatility clustering. A common remedy is to thin the data by keeping only observations spaced by several days, see Cai et al. (2015, Section 4). Here, it means that for some fixed day-gap $m \ge 3$, the value Y_i corresponds to the mi-th day and the discretized curve $\tilde{X}_i$ to the day $mi - 1$.

5.2. **Response Tail Behaviour.** Before applying FEPLS, we need to verify that the response variable Y exhibits a heavy tail. The evidence for the pair DAX30-ETH is given by the Hill plot in Figure 4, a standard diagnostic tool from extreme-value theory which represents the graph of the Hill estimator

$$\hat{\gamma}(k) = \frac{1}{k}\sum_{i=1}^{k}(\log Y_{n-i+1,n} - \log Y_{n-k,n})$$

as a function of $k \in \{15, \dots, 65\}$, see Hill (1975). As such, Figure 4 exhibits a stable behaviour over a range of intermediate values of k, with an estimated tail index around $\hat{\gamma} \approx 1/2$.

5.3. **Threshold Selection.** Applying the selection rule of Section 4.2, we plot the logarithm of the empirical absolute covariance between Y and $\langle\tilde{\beta}_\varphi, \tilde{X}\rangle$ in Figure 5, namely $k \mapsto \log|r(k)|$ for $k \in \{15, \dots, 65\}$. Selecting the argument of $|r(k)|$ that gives its maximum value hence leads to $\tilde{k} = 6$, which means that the extreme regime approximately consists of the top 2.26% of the 265 response values. This is used as the effective sample size for all subsequent extreme-tail analyses about the pair DAX30-ETH. We should mention that different τ would lead to very distinct outcomes for $\tilde{k}$, *e.g.*, for $\tau = 1$, one has $\tilde{k} = 17$.

5.4. **Tail Dependence Competition.** To assess the effectiveness of our method, we evaluate its performance in a tail covariance competition against several projection strategies on the present high-dimensional discretized functional data, *i.e.*, for the pair DAX30-ETH and then on all possible pairs. A variety of baseline competitor strategies is considered: random projections, along with standard dimension reduction methods and their extreme variants (with e-prefix) computed only on the tail region of the data sample. More precisely, fix a threshold $y > 0$ and define the exceedances index set as

$$\mathcal{I}_y \;:=\; \{\, i \in \{1,\dots,n\} : Y_i > y \,\}, \qquad n_y := |\mathcal{I}_y|.$$

Let $\tilde{X}_y \in \mathbb{R}^{n_y \times N}$ denote the submatrix with rows $\{\tilde{X}_i^\top : i \in \mathcal{I}_y\}$ and $Y_y \in \mathbb{R}^{n_y}$ the corresponding tail responses. Let $\mathcal{M} := \{\text{FEPLS}, \text{eFPCA}, \text{eFSIR}, \text{FPCA}, \text{FSIR}\}$, then each method $m \in \mathcal{M}$ produces a projection score

$$Z_{m,y} \;:=\; \tilde{X}_y \, \Psi \, \beta_m(y) \in \mathbb{R}^{n_y},$$

where the method-specific direction $\beta_m(y) = (\beta_m(t_1), \dots, \beta_m(t_N))^\top \in \mathbb{R}^N$ (or β_m if the method does not depend on y) is defined below, and $\Psi \in \mathbb{R}^{N \times N}$ is as in Section 4.

For each $m \in \mathcal{M}$, we compute the unbiased empirical covariance between the method's score and the tail responses, as the metric performance for our analysis,

$$\rho_m(y) := \frac{1}{n_y - 1} \sum_{i \in \mathcal{I}_y} \left((Z_{m,y})_i - \bar{Z}_{m,y} \right) \left(Y_i - \bar{Y}_y \right),$$

with $\bar{Z}_{m,y} = n_y^{-1} \sum_{i \in \mathcal{I}_y} (Z_{m,y})_i$ and $\bar{Y}_y = n_y^{-1} \sum_{i \in \mathcal{I}_y} Y_i$. The standard full-sample versions are recovered as special cases when $\mathcal{I}_y = \{1, \dots, n\}$. All competitors are implemented in their discretized form with the Ψ-inner product rather than the standard L^2 inner product.

(i) Extreme Functional PCA (eFPCA). Let $\bar{\tilde{X}}_y := n_y^{-1} \sum_{i \in \mathcal{I}_y} \tilde{X}_i$ be the empirical mean over the tail. Define the empirical covariance matrix of the tail covariates:

$$\widehat{\Sigma}_y := \frac{1}{n_y} \sum_{i \in \mathcal{I}_y} (\tilde{X}_i - \bar{\tilde{X}}_y)(\tilde{X}_i - \bar{\tilde{X}}_y)^\top \in \mathbb{R}^{N \times N}.$$

The eFPCA direction $\beta_{\text{eFPCA}}(y)$ is the eigenvector corresponding to the largest eigenvalue of $\widehat{\Sigma}_y$:

$$\beta_{\text{eFPCA}}(y) := \underset{\beta \in \mathbb{R}^N, \|\beta\|_\Psi = 1}{\arg\max} \; \beta^\top \widehat{\Sigma}_y \beta,$$

where $\|\beta\|_\Psi^2 := \beta^\top \Psi \beta$.

(ii) Extreme Functional SIR (eFSIR). Partition $\mathcal{I}_y$ into 5 equal-frequency slices $\mathcal{S}_1, \dots, \mathcal{S}_5$ according to the ordered values of $\{Y_i : i \in \mathcal{I}_y\}$. For each slice $1 \le h \le 5$, define the slice mean $m_h := \frac{1}{|\mathcal{S}_h|} \sum_{i \in \mathcal{S}_h} \tilde{X}_i$. The between-slice covariance matrix is

$$\widehat{\Sigma}_{B,y} := \sum_{h=1}^{5} \frac{|\mathcal{S}_h|}{n_y} \, (m_h - \bar{\tilde{X}}_y)(m_h - \bar{\tilde{X}}_y)^\top.$$

The eFSIR direction $\beta_{\mathrm{eFSIR}}(y)$ is the solution of the generalized eigenvalue problem

$$\widehat{\Sigma}_{B,y}\beta = \lambda\,\widehat{\Sigma}_y\beta,$$

taken for the eigenvalue λ with largest absolute value. In practice we solve the regularized problem $(\widehat{\Sigma}_y + \eta I_N)\beta = \lambda^{-1}\widehat{\Sigma}_{B,y}\beta$ with $\eta = 10^{-6}\,\mathrm{tr}(\widehat{\Sigma}_y)/N$ and I_N being the identity matrix of size N. The full-sample FSIR direction β_{FSIR}, obtained by replacing $\mathcal{I}_y$ with $\{1,\dots,n\}$, uses 10 slices.

(iii) Random projections. As a naive baseline, we generate 5000 independent random directions $(\beta^{(r)}/\|\beta^{(r)}\|_\Psi)_{1\le r\le 5000}$ by drawing $\beta^{(r)} \sim \mathcal{N}(0, I_N)$.

5.5. **Competition results.** In Figure 6, we plot the graph of $k \mapsto |\rho_m(Y_{n-k+1,n})|$ for all $m \in \mathcal{M}$ and the fixed pair $(a_1, a_2) = (\mathrm{DAX30}, \mathrm{ETH})$. One may see that FEPLS dominates the competitors across almost all thresholds k. The gray shaded region represents the 5–95% inter-quantile range of covariances obtained from 5000 random directions. It highlights the fact that the observed performance of FEPLS is far above what would be expected by chance. Lastly, one may also notice that the extreme versions of the competitor strategies consistently outperform their full-sample counterpart.

Extending Figure 6 across dataset-threshold combinations, Figure 7 reports the evolution, as a function of the threshold level k, of the mean rank of the different dimension-reduction methods in terms of absolute tail covariance with the response, based upon the 25 possible pairs and all threshold levels k. For each value of k, methods are ranked from best (rank 1) to worst, and the displayed curves correspond to ranks averaged over the 25 asset pairs; hence, lower values indicate better overall performance. FEPLS consistently attains the lowest mean rank across the entire range of thresholds. Moreover, classical full-sample methods systematically exhibit higher mean ranks than their extreme counterparts, especially for smaller values of k.

Finally and as a complement to the previous plot, Figure 8 displays the win rate of FEPLS, *i.e.*, the proportion of pairs on which it outperforms all competitors, across different threshold levels k. To formally support the claim that the superior performance of FEPLS is statistically significant and not attributable to random variations, we employ a one-sided binomial test under the null hypothesis of equal performance across all five methods. Under the null hypothesis, each method has equal 20% of winning probability. With 25 (supposedly independent) datasets, the number of wins W for FEPLS follows $W \sim \mathrm{Binomial}(25, 0.20)$. The critical value for 5%-significance corresponds to 9 wins over those 25 trials, meaning a win rate of 36%, since $\mathbb{P}(W \ge 9) = 0.0468$, as visualized by the horizontal green line. The plot shows that FEPLS achieves a win rate consistently above 36% across all dataset-threshold combinations, providing evidence against the null hypothesis.

5.6. **Conditional extreme quantile inference.** To illustrate how our method can be used for risk measurement in the extreme regime, we go back to the particular case $(a_1, a_2) = (\mathrm{DAX30}, \mathrm{ETH})$ and study its related conditional quantile inference. For a generic random variable $V \in H$, let us denote by $F_{Y|V=v}$ the conditional cumulative distribution function of Y given $V = v$, where $v \in H$.

The (functional) Nadaraya-Watson estimator of $F_{Y|V=v}$ is

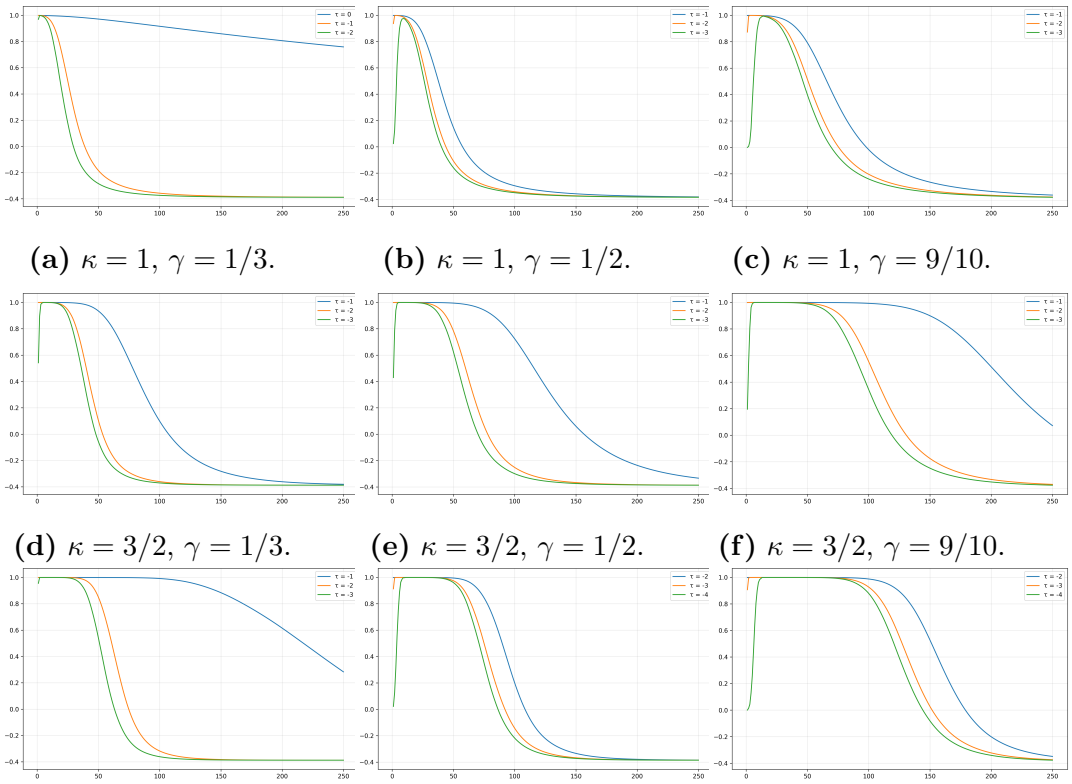
$$\hat{F}_{Y|V=v}(y) := \sum_{i=1}^{n} K\left(\frac{\|V_i - v\|}{h_n}\right) 1_{\{Y_i \leq y\}} \Bigg/ \sum_{s=1}^{n} K\left(\frac{\|V_s - v\|}{h_n}\right), \quad y \in \mathbb{R}. \quad (19)$$

Here, the kernel K is Gaussian and the parameter $h_n \downarrow 0$ is the window bandwidth. The conditional quantile function of $Y|V = v$ is defined thanks to the generalized inverse $q_{Y|V=v} := F^{-}_{Y|V=v}$ and the associated estimator is

$$\hat{q}_{Y|V=v}(\alpha) := \inf\{y > 0, \hat{F}_{Y|V=v}(y) \geq \alpha\}, \quad \alpha \in (0,1),$$

see Ferraty and Vieu (2006, Section 6.4) for details.

Figure 9 shows the scatterplot for the pair DAX30-ETH of the projected data on the FEPLS direction $(\langle \tilde{X}_i, \tilde{\beta}_\varphi \rangle, Y_i)$ for $1 \leq i \leq n$ endowed with the estimated conditional Value-at-Risk curves $t \mapsto \hat{q}_{Y|\langle \tilde{X}, \tilde{\beta}_\varphi \rangle = t}(\alpha)$ associated with risk levels $\alpha \in \{0.95, 0.98, 0.995\}$. Herein, the kernel inference relies on inverting the univariate version of (19) with bandwidth h_n arbitrarily chosen to be 5.10^{-5}. The red dots are associated with the $\tilde{k}$ pairs with highest response values. Such a graphic may be used as a visualization tool to detect the most risky situations. Indeed, given a new pair (x_0, y_0), one may determine whether y_0 given $\tilde{X} = x_0$ exceeds the conditional Value-at-Risk $\hat{q}_{Y|\langle \tilde{X}, \tilde{\beta}_\varphi \rangle = \langle x_0, \tilde{\beta}_\varphi \rangle}$ simply by comparing its projected position $(\langle x_0, \tilde{\beta}_\varphi \rangle, y_0)$ to the quantile curve on Figure 9. Additional experiments (not reported here) show that the FEPLS outputs are not very sensitive to τ in this real data application as well as to $\alpha \leq 0.98$ for the extreme quantile estimation.

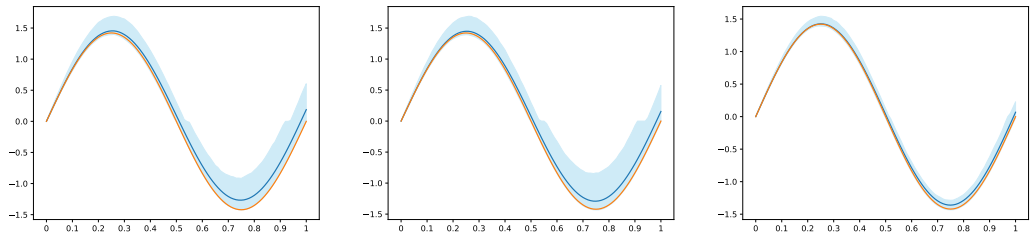


(a) $\kappa = 1$, $\gamma = 1/3$. **(b)** $\kappa = 1$, $\gamma = 1/2$. **(c)** $\kappa = 1$, $\gamma = 9/10$.

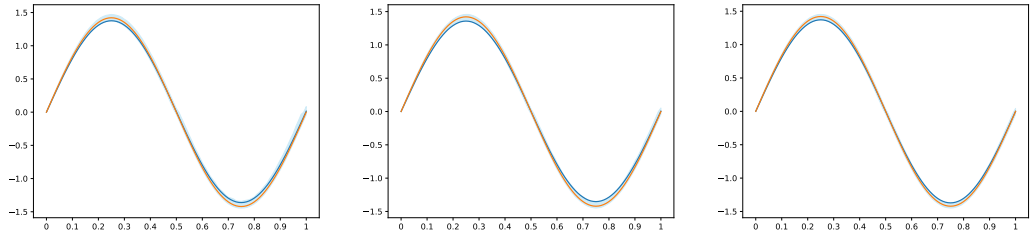
(d) $\kappa = 3/2$, $\gamma = 1/3$. **(e)** $\kappa = 3/2$, $\gamma = 1/2$. **(f)** $\kappa = 3/2$, $\gamma = 9/10$.

(g) $\kappa = 2$, $\gamma = 1/3$. **(h)** $\kappa = 2$, $\gamma = 1/2$. **(i)** $\kappa = 2$, $\gamma = 9/10$.

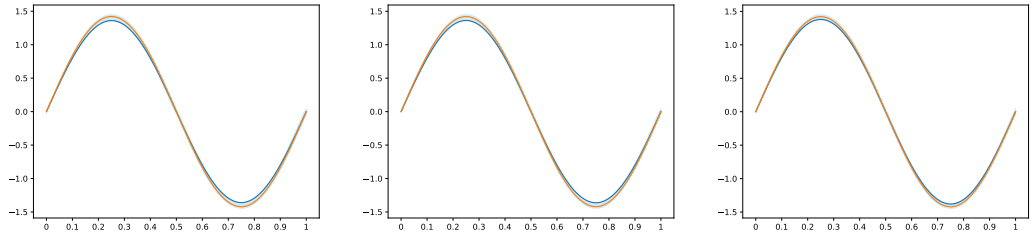
FIGURE 1. Finite sample behaviour of $\tilde{\beta}_{\varphi}(Y_{n-k+1,n})$ for different admissible values of τ (see Table 1) under the simulated inverse model with Burr response and conditional fBm noise. The x-axis represents the number k of exceedances and the y-axis is the average value of $\langle \tilde{\beta}_{\varphi}(Y_{n-k+1,n}), \beta \rangle$ over 500 Monte Carlo replications. Here, $\rho = -1/2$ and $N = 101$.



(a) $\kappa = 1$, $\gamma = 1/3$. **(b)** $\kappa = 1$, $\gamma = 1/2$. **(c)** $\kappa = 1$, $\gamma = 9/10$.

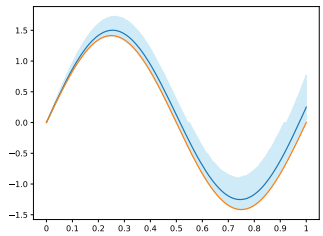


(d) $\kappa = 3/2$, $\gamma = 1/3$. **(e)** $\kappa = 3/2$, $\gamma = 1/2$. **(f)** $\kappa = 3/2$, $\gamma = 9/10$.

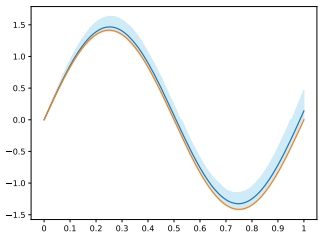


(g) $\kappa = 2$, $\gamma = 1/3$. **(h)** $\kappa = 2$, $\gamma = 1/2$. **(i)** $\kappa = 2$, $\gamma = 9/10$.

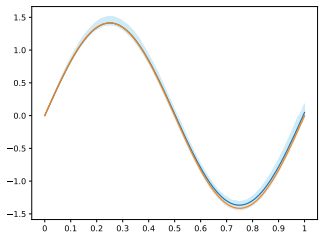
FIGURE 2. Simulation results on the inverse model with Burr response and conditional fBm noise. The orange curve is the graph of $t \in [0,1] \mapsto \beta(t)$ while the blue one is the averaged value of $t \in [0,1] \mapsto \tilde{\beta}_{\varphi}(Y_{n-\tilde{k}+1,n})(t)$ over 500 Monte Carlo replications, where $\tilde{k} = \arg\max_{5 \le k \le n/5} r(k)$. The light blue area corresponds to the pointwise 5–95% inter-quantile envelope of the Monte Carlo replications. Here, $\rho = -1/2$, $\tau = -2$ and $N = 101$.



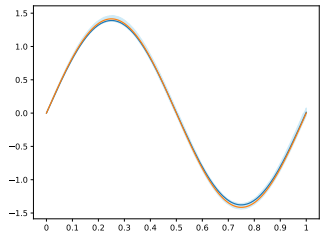
(a) $\kappa = 1$, $\gamma = 1/3$.



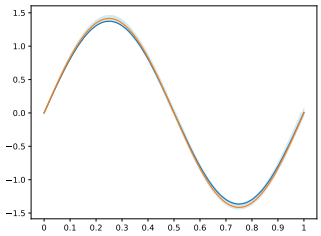
(b) $\kappa = 1$, $\gamma = 1/2$.



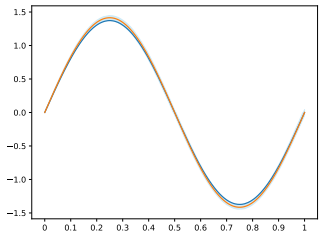
(c) $\kappa = 1$, $\gamma = 9/10$.



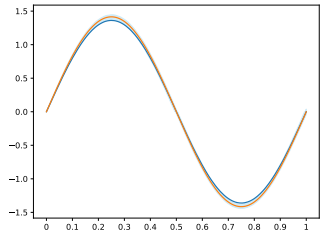
(d) $\kappa = 3/2$, $\gamma = 1/3$.



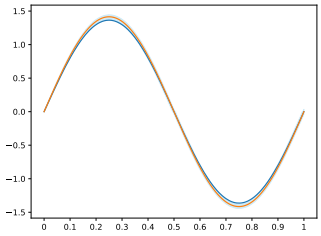
(e) $\kappa = 3/2$, $\gamma = 1/2$.



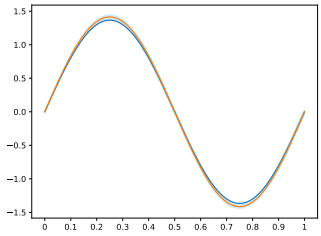
(f) $\kappa = 3/2$, $\gamma = 9/10$.



(g) $\kappa = 2$, $\gamma = 1/3$.



(h) $\kappa = 2$, $\gamma = 1/2$.



(i) $\kappa = 2$, $\gamma = 9/10$.

FIGURE 3. Simulation results akin to Figure 2 for $N = 1001$.

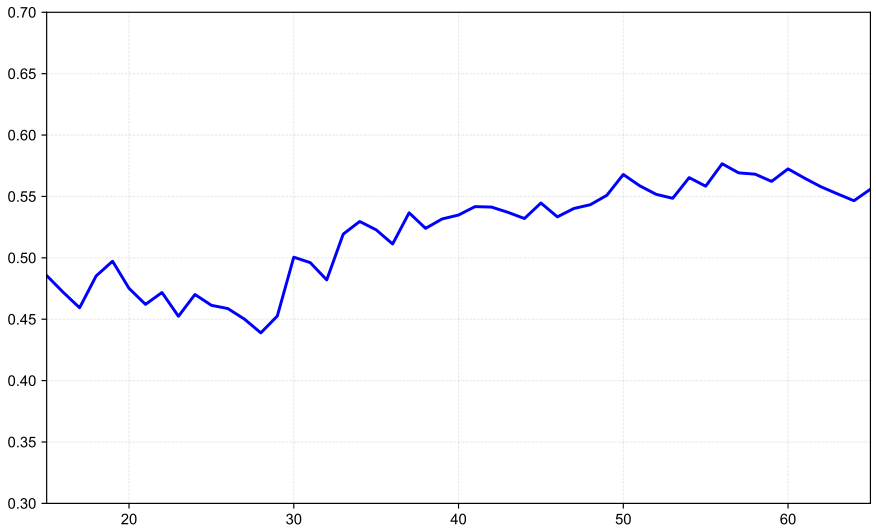


FIGURE 4. Hill plot for the response variable in the pair DAX30-ETH showing $k \mapsto \hat{\gamma}(k)$ where k is the number of extreme observations.

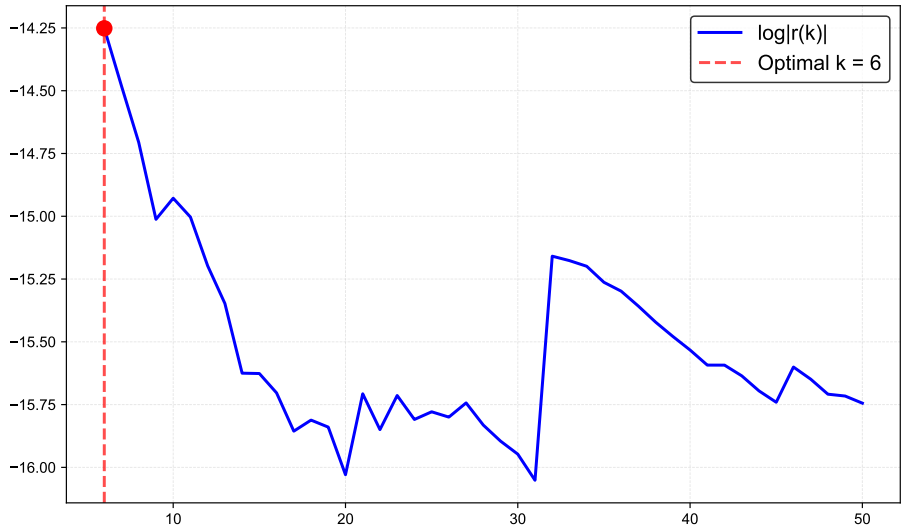


FIGURE 5. Threshold selection for the DAX30-ETH pair. The plot displays $k \mapsto \log|r(k)|$. The optimal threshold is obtained at $\tilde{k} = 6$ (red dot).

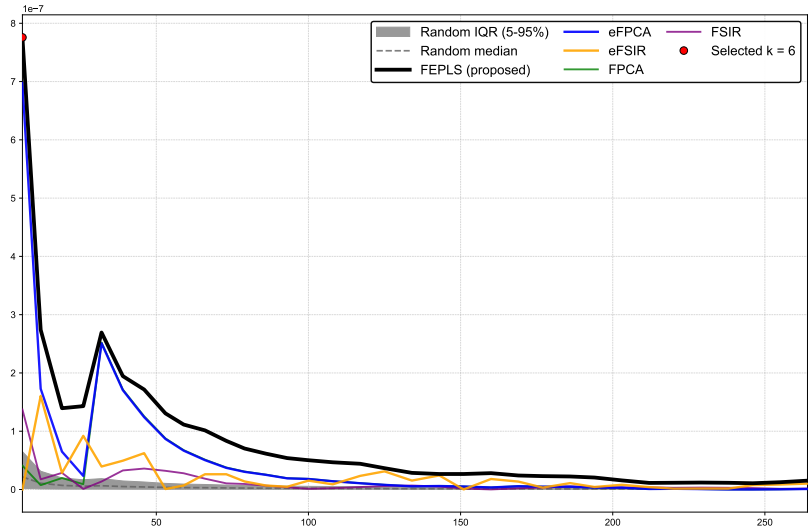


FIGURE 6. Comparison for the DAX30-ETH pair of FEPLS against four competing methods across threshold values k in the x-axis. Each curve shows $|\rho_m(Y_{n-k+1,n})|$ the absolute unbiased covariance between the method's projection and the response for the top k observations. The grey shaded region represents the 5–95% inter-quantile range of 5000 random projections.

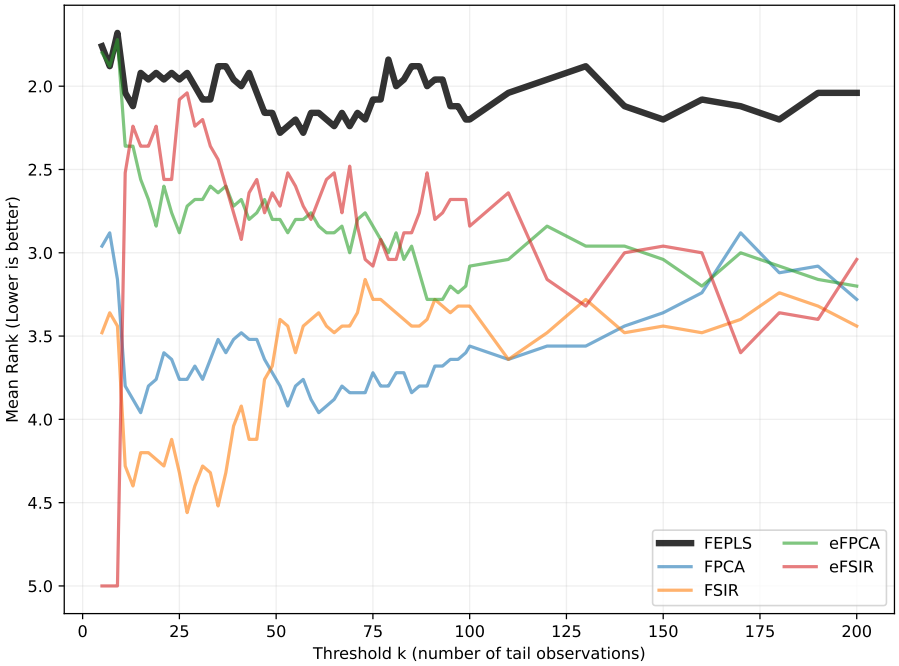


FIGURE 7. Evolution of mean rank performance across thresholds for the different methods. Results are based on the evaluation over 25 asset pairs.

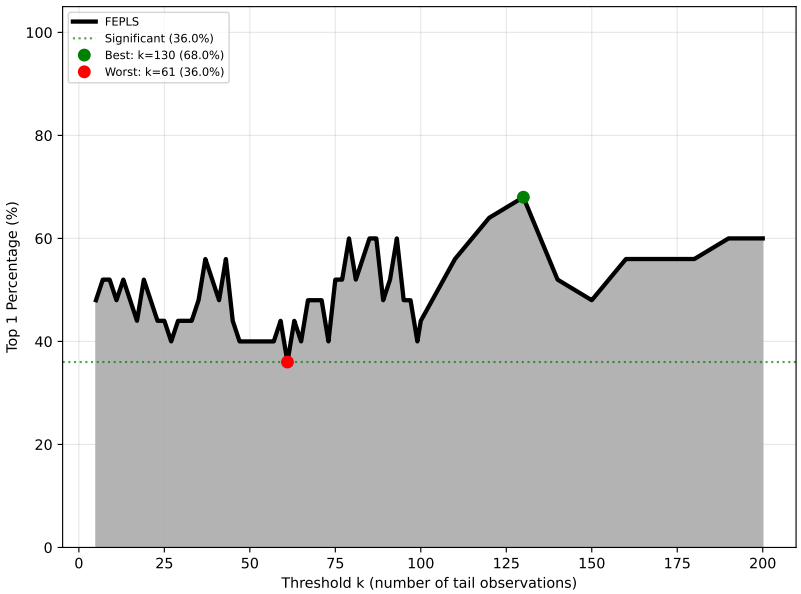


FIGURE 8. Win rate of FEPLS, *i.e.*, proportion of the 25 pairs on which it ranks first, across threshold values k. The green line shows the statistical significance threshold derived from a binomial test at level 5% with 25 datasets and 5 competing methods.

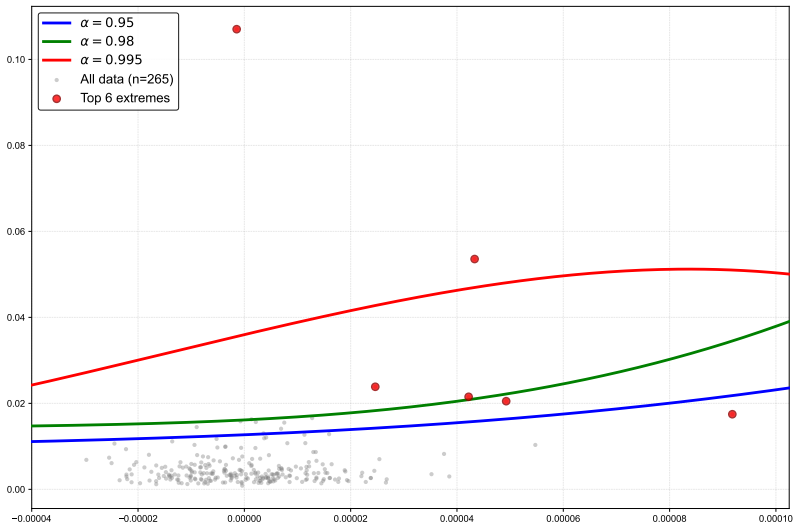


FIGURE 9. Conditional extreme quantile curves estimated via FEPLS for the DAX30-ETH pair. The plot shows the FEPLS projection score $Z = \langle \tilde{X}, \tilde{\beta}_\varphi \rangle_\Psi$ versus the response Y, with estimated conditional Value-at-Risk curves at different levels. Red dots mark the $\tilde{k} = 6$ most extreme observations; gray points show all projected data.

Appendix A. Technical results

This section gathers all the mathematical details behind Section 2 and Section 3. We start with A.1 providing useful technical tools which enable to manipulate inner products under conditional expectations, or the distribution of order statistics allowing an extension of Bousebata et al. (2023, Lemma 2) to the random threshold framework. This yields theoretical tail-moments asymptotics which will be extensively used throughout the paper. The section also contains the proof of Proposition 1. Next, the asymptotic behaviour of the relevant empirical tail-moments, involving the noise and its inner product with β, are investigated in A.2. Finally, our main Theorem 1 is established in A.3 under the inverse single-index model, as well as Proposition 2 holding for general empirical tail-moments in a model-free setup.

A.1. **Preliminaries.** A standard result in the literature of Bochner spaces, *e.g.*, Hytonen et al. (2016, Eq (1.2)), states that Bochner integrals commute with bounded linear operators such as orthogonal projections. Note that this is in fact the definition of Pettis integrals. In the Hilbert framework, projections are generally expressed in terms of the inner product. The following technical tool considers this commutation operation with the inner product when the two random variables are conditionally stochastically independent. The case of one variable being deterministic will also often occur in the proofs. Recall the conditional expectation $\mathbb{E}_{\mathcal{F}}$ and its Bochner version $\mathbb{E}_{\mathcal{F},\mathcal{B}}$, both introduced in Section 3.

Lemma 1. *Let $\mathcal{F}$ be a sub-σ-field of $\mathcal{A}$ and let W_1, W_2 be Bochner-integrable random variables having values in $(H, \langle\cdot,\cdot\rangle)$ a separable Hilbert space. Assume that W_1 and W_2 are independent and, in addition, independent conditionally on $\mathcal{F}$. Then $\langle W_1, W_2\rangle$ is integrable and*

$$\mathbb{E}_{\mathcal{F}}(\langle W_1, W_2\rangle) = \langle \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1), \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_2)\rangle \quad a.s.$$

Proof. Let us fix a countable orthonormal basis $(e_j)_{j\geq 1}$ of H. We start by observing that Cauchy-Schwarz's inequality in H, the independence of W_1 and W_2 and Tonelli's theorem entail

$$\mathbb{E}(|\langle W_1, W_2\rangle|) \leq \mathbb{E}(\|W_1\|\,\|W_2\|) = \mathbb{E}(\|W_1\|)\,\mathbb{E}(\|W_2\|) < +\infty, \tag{20}$$

so that all the conditional expectations at hand below are well defined. Denote $\mathcal{G} := \mathcal{F} \vee \sigma(W_2)$.

The first step is to establish the following pull-out property: for any sub-σ-field $\mathcal{H}$ of $\mathcal{A}$, any Bochner-integrable random variable U having values in H and any $\mathcal{H}$-measurable random variable V having values in H such that $\mathbb{E}(\|U\|\,\|V\|) < +\infty$, one has

$$\mathbb{E}_{\mathcal{H}}(\langle U, V\rangle) = \langle \mathbb{E}_{\mathcal{H},\mathcal{B}}(U), V\rangle \quad \text{a.s.} \tag{21}$$

First, when $V = \sum_{m=1}^{M} 1_{A_m} z_m$ is a simple function, with $A_1, \ldots, A_M \in \mathcal{H}$ pairwise disjoint and $z_1, \ldots, z_M \in H$, (21) follows from the linearity of the conditional expectation, the $\mathcal{H}$-measurability of the indicator functions and the fact that Bochner conditional expectations commute with bounded linear forms such as $\langle\cdot, z_m\rangle$, see Hytonen et al. (2016, Section 2.6). In the general case, the strong $\mathcal{H}$-measurability of V provides simple $\mathcal{H}$-measurable functions $\tilde{V}_M \to V$

pointwise. Replacing $\tilde{V}_M$ by $V_M := \tilde{V}_M 1_{\{\|\tilde{V}_M\| \leq 2\|V\|\}}$, which is still simple and $\mathcal{H}$-measurable, we get $V_M \to V$ with $\|V_M\| \leq 2\|V\|$ pointwise. Since $|\langle U, V_M\rangle| \leq 2\|U\|\,\|V\|$ is integrable, the conditional dominated convergence theorem applies to the left-hand side of (21) written for V_M, while the right-hand side converges almost surely by continuity of the inner product. This proves (21).

Next, let us show that

$$\mathbb{E}_{\mathcal{G},\mathcal{B}}(W_1) = \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1) \quad \text{a.s.} \tag{22}$$

For any $z \in H$, the conditional independence of W_1 and W_2 given $\mathcal{F}$ entails $\mathbb{E}_{\mathcal{G}}(\langle W_1, z\rangle) = \mathbb{E}_{\mathcal{F}}(\langle W_1, z\rangle)$ a.s., see Kallenberg (2002, Proposition 6.6). Commuting again the Bochner conditional expectation with bounded linear forms, this rewrites $\langle \mathbb{E}_{\mathcal{G},\mathcal{B}}(W_1), z\rangle = \langle \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1), z\rangle$ a.s., and letting z run through the countable basis $(e_j)_{j\geq 1}$ yields (22).

The last step is to combine (21) and (22). The random variable $\mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1)$ is $\mathcal{F}$-measurable and the conditional Jensen inequality $\|\,\mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1)\| \leq \mathbb{E}_{\mathcal{F}}(\|W_1\|)$ combined with the conditional independence of $\|W_1\|$ and $\|W_2\|$ given $\mathcal{F}$ yields

$$\begin{aligned}\mathbb{E}(\|\,\mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1)\|\,\|W_2\|) &\leq \mathbb{E}\left(\mathbb{E}_{\mathcal{F}}(\|W_1\|)\,\mathbb{E}_{\mathcal{F}}(\|W_2\|)\right) = \mathbb{E}\left(\mathbb{E}_{\mathcal{F}}(\|W_1\|\,\|W_2\|)\right)\\ &= \mathbb{E}(\|W_1\|\,\|W_2\|) < +\infty\end{aligned}$$

in view of (20). Whence, the tower property, (21) applied with $(U, V, \mathcal{H}) = (W_1, W_2, \mathcal{G})$, identity (22), and again (21), now applied with $\mathcal{H} = \mathcal{F}$, $U = W_2$ and $V = \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1)$, entail

$$\begin{aligned}\mathbb{E}_{\mathcal{F}}(\langle W_1, W_2\rangle) = \mathbb{E}_{\mathcal{F}}\left(\mathbb{E}_{\mathcal{G}}(\langle W_1, W_2\rangle)\right) &= \mathbb{E}_{\mathcal{F}}\left(\langle \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1), W_2\rangle\right)\\ &= \langle \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1), \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_2)\rangle \quad \text{a.s.},\end{aligned}$$

which ends the proof. □

We may now establish the explicit solution of the optimization problem (1) under the inverse single-index model (6). The following proof is a simpler version of Bousebata et al. (2023, Proposition 1) but it extends to the functional case.

Proof of Proposition 1. We start by rewriting the conditional covariance. Let $w \in H$ such that $\|w\| = 1$ and let $y \in \mathbb{R}$ such that $\bar{F}(y) > 0$. By Cauchy-Schwarz's inequality in H, $|\langle w, X\rangle Y| \leq \|X\|\,|Y|$ and $|\langle w, X\rangle| \leq \|X\|$, so that the moment assumption ensures that all the expectations at hand below are finite; in particular, $m_{XY}(y)$ and $m_X(y)$ are well-defined Bochner integrals. Then,

$$\mathrm{Cov}(\langle w, X\rangle, Y | Y \geq y) = \frac{\mathbb{E}(\langle w, X\rangle Y 1_{\{Y\geq y\}})}{\bar{F}(y)} - \frac{\mathbb{E}(\langle w, X\rangle 1_{\{Y\geq y\}})\,\mathbb{E}(Y 1_{\{Y\geq y\}})}{\bar{F}(y)^2}.$$

Next, since Bochner integrals commute with bounded linear forms such as $\langle w, \cdot\rangle$, see Hytonen et al. (2016, Eq (1.2)), one has $\mathbb{E}(\langle w, X\rangle Y 1_{\{Y\geq y\}}) = \langle w, m_{XY}(y)\rangle$ and $\mathbb{E}(\langle w, X\rangle 1_{\{Y\geq y\}}) = \langle w, m_X(y)\rangle$, whence

$$\bar{F}^2(y)\ \mathrm{Cov}(\langle w, X\rangle, Y | Y \geq y) = \bar{F}(y)\langle w, m_{XY}(y)\rangle - \langle w, m_Y(y) m_X(y)\rangle = \langle w, v(y)\rangle.$$

The optimization problem now becomes $\arg\max_{\|w\|=1}\langle w, v(y)\rangle$. By Cauchy-Schwarz's inequality, $\langle w, v(y)\rangle \leq \|v(y)\|$ for any $w \in H$ such that $\|w\| = 1$, with equality if and only if $w = v(y)/\|v(y)\|$ since $v(y) \neq 0$. This proves both optimality and uniqueness, and ends the proof. □

The next lemma describes the exact conditional distribution of the whole sample given the intermediate order statistic $Y_{n-k+1,n}$. It is the key tool to handle tail-moments at the random threshold: it extends Bousebata et al. (2023, Lemma 2) to this framework, see Lemma 3 below, and provides the second-moment identity of Lemma 4, both of which are extensively used in the sequel. Recall that, given $\{Y_{n-k+1,n} = y\}$, the conditional expectation is denoted by $\mathbb{E}_{k,y}(\cdot) := \mathbb{E}(\cdot \mid Y_{n-k+1,n} = y)$ and the conditional tail-moment of a generic random object W_i, $1 \le i \le n$, by $m_{k,y}(W_i) := \mathbb{E}_{k,y}(W_i 1_{\{Y_i \ge Y_{n-k+1,n}\}})$. In the sequel, $\mathcal{L}(\cdot)$ and $\mathcal{L}(\cdot \mid \cdot)$ stand for the distribution and the conditional distribution of the random object at hand.

Lemma 2. *Let $(X_i, Y_i)_{1\le i\le n}$ be independent copies of $(X, Y) \in H \times \mathbb{R}$ and assume that Y admits a density f. Let k be an integer such that $2 \le k < n$ and let $y \in \mathbb{R}$ such that $f(y) > 0$ and $0 < F(y) < 1$. Then, conditionally on $\{Y_{n-k+1,n} = y\}$, the sample $(X_i, Y_i)_{1\le i\le n}$ is distributed as follows: a pair (i_0, S) is drawn uniformly at random among the $n\binom{n-1}{k-1}$ pairs made of an index $i_0 \in \{1, \dots, n\}$ and a subset $S \subset \{1, \dots, n\} \setminus \{i_0\}$ with $|S| = k-1$; given (i_0, S), the pairs (X_i, Y_i), $1 \le i \le n$, are independent, with*

$$(X_j, Y_j) \sim \mathcal{L}\big((X, Y) \mid Y > y\big) \text{ for } j \in S,$$
$$(X_j, Y_j) \sim \mathcal{L}\big((X, Y) \mid Y < y\big) \text{ for } j \notin S \cup \{i_0\},$$

while $Y_{i_0} = y$ and $X_{i_0} \sim \mathcal{L}(X \mid Y = y)$. In words, i_0 is the index of the observation achieving the order statistic, the indices of S are those of the $k-1$ observations falling above the threshold y, and the $n-k$ remaining observations fall below. In particular, writing $e_>(\Psi) := \mathbb{E}(\Psi(X, Y) \mid Y > y)$ and $e_=(\Psi) := \mathbb{E}(\Psi(X, Y) \mid Y = y)$ for any measurable function $\Psi : H \times \mathbb{R} \to [0, +\infty]$, one has, for any such Ψ, Ψ' and any $1 \le i \ne \ell \le n$:

$$\mathbb{E}_{k,y}\left(\Psi(X_i, Y_i) 1_{\{Y_i \ge y\}}\right) = \frac{k-1}{n}\, e_>(\Psi) + \frac{1}{n}\, e_=(\Psi), \tag{23}$$

$$\begin{aligned}\mathbb{E}_{k,y}\left(\Psi(X_i, Y_i)\Psi'(X_\ell, Y_\ell) 1_{\{Y_i \ge y\}} 1_{\{Y_\ell \ge y\}}\right) &= \frac{(k-1)(k-2)}{n(n-1)}\, e_>(\Psi)\, e_>(\Psi') \\ &\quad + \frac{k-1}{n(n-1)}\Big(e_>(\Psi)\, e_=(\Psi') + e_=(\Psi)\, e_>(\Psi')\Big).\end{aligned} \tag{24}$$

By decomposing into positive and negative parts, (23) extends to any measurable $\Psi : H \times \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}_{k,y}\left(|\Psi(X_i, Y_i)| 1_{\{Y_i \ge y\}}\right) < +\infty$.

Proof. We start by proving the representation for the sub-sample $(Y_1, \dots, Y_n)$. Since F is continuous, the values $Y_1, \dots, Y_n$ are a.s. pairwise distinct. For (i_0, S) as in the statement, define the event

$$E(i_0, S) := \big\{Y_j > Y_{i_0} \text{ for any } j \in S\big\} \cap \big\{Y_j < Y_{i_0} \text{ for any } j \notin S \cup \{i_0\}\big\}.$$

On $E(i_0, S)$, exactly $k-1$ values exceed Y_{i_0}, so that $Y_{n-k+1,n} = Y_{i_0}$. Up to a $\mathbb{P}$-null set, the event $\{Y_{n-k+1,n} \le y\}$ is thus the disjoint union over the $n\binom{n-1}{k-1}$ pairs (i_0, S) of the events $E(i_0, S) \cap \{Y_{i_0} \le y\}$. For any Borel sets $B_1, \dots, B_n$ of

$\mathbb{R}$, conditioning on Y_{i_0} and using independence yield

$$\mathbb{P}\Big(\bigcap_{i=1}^{n}\{Y_i \in B_i\} \cap E(i_0, S) \cap \{Y_{i_0} \le y\}\Big) = \int_{-\infty}^{y} f(u)\, 1_{B_{i_0}}(u) \prod_{j\in S} \mathbb{P}\big(Y \in B_j,\, Y > u\big) \prod_{j \notin S\cup\{i_0\}} \mathbb{P}\big(Y \in B_j,\, Y < u\big)\, \mathrm{d}u.$$

Choosing $B_1 = \cdots = B_n = \mathbb{R}$ and summing over the pairs (i_0, S) recovers the classical density of the order statistic, $f_{Y_{n-k+1,n}}(u) = n\binom{n-1}{k-1} f(u)\bar{F}^{k-1}(u)F^{n-k}(u)$. The conditional distribution given $\{Y_{n-k+1,n} = y\}$ is, for almost every such y, the ratio of the integrand at $u = y$ to this density:

$$\mathbb{P}\Big(\bigcap_{i=1}^{n}\{Y_i \in B_i\} \cap E(i_0, S) \,\Big|\, Y_{n-k+1,n} = y\Big) = \frac{1}{n\binom{n-1}{k-1}}\, 1_{B_{i_0}}(y) \prod_{j\in S} \mathbb{P}\big(Y \in B_j \mid Y > y\big) \prod_{j \notin S\cup\{i_0\}} \mathbb{P}\big(Y \in B_j \mid Y < y\big).$$

In other words, each pair (i_0, S) has probability $\big(n\binom{n-1}{k-1}\big)^{-1}$ and, given (i_0, S), the random variables Y_j, $j \neq i_0$, are independent with $Y_j \sim \mathcal{L}(Y \mid Y > y)$ for $j \in S$ and $Y_j \sim \mathcal{L}(Y \mid Y < y)$ otherwise, while $Y_{i_0} = y$.

Let us now incorporate the X_i's. Conditionally on $(Y_1, \ldots, Y_n)$, the random variables $X_1, \ldots, X_n$ are independent with $X_i \sim \mathcal{L}(X \mid Y = Y_i)$. Both (i_0, S) and the event $\{Y_{n-k+1,n} = y\}$ only depend on $(Y_1, \ldots, Y_n)$, so this remains true under the above conditioning. Hence, given (i_0, S), the pairs (X_i, Y_i), $1 \le i \le n$, are independent, $Y_{i_0} = y$ with $X_{i_0} \sim \mathcal{L}(X \mid Y = y)$, and, for $j \in S$, the pair (X_j, Y_j) is obtained by drawing $Y_j \sim \mathcal{L}(Y \mid Y > y)$ and then $X_j \sim \mathcal{L}(X \mid Y = Y_j)$. This two-step draw has distribution $\mathcal{L}((X, Y) \mid Y > y)$ since, for any measurable sets $A \subset H$ and $B \subset \mathbb{R}$, the tower property gives

$$\mathbb{P}(X_j \in A,\, Y_j \in B) = \mathbb{E}\left(1_{\{Y\in B\}}\, \mathbb{P}(X \in A \mid Y)\,\middle|\, Y > y\right) = \mathbb{P}\big(X \in A,\, Y \in B \mid Y > y\big).$$

Likewise, $(X_j, Y_j) \sim \mathcal{L}((X, Y) \mid Y < y)$ for $j \notin S \cup \{i_0\}$, which proves the representation.

It remains to derive (23) and (24). Under the representation, $\{Y_i \ge y\} = \{i \in S \cup \{i_0\}\}$ a.s., so that, conditionally on (i_0, S),

$$\mathbb{E}_{k,y}\left(\Psi(X_i, Y_i) 1_{\{Y_i \ge y\}} \mid (i_0, S)\right) = e_{>}(\Psi)\, 1_{\{i\in S\}} + e_{=}(\Psi)\, 1_{\{i_0 = i\}},$$

and, for $i \neq \ell$, the conditional independence of (X_i, Y_i) and (X_ℓ, Y_ℓ) given (i_0, S) entails

$$\mathbb{E}_{k,y}\left(\Psi(X_i, Y_i)\Psi'(X_\ell, Y_\ell) 1_{\{Y_i \ge y\}} 1_{\{Y_\ell \ge y\}} \mid (i_0, S)\right) = \Big(e_{>}(\Psi)\, 1_{\{i\in S\}} + e_{=}(\Psi)\, 1_{\{i_0 = i\}}\Big)\Big(e_{>}(\Psi')\, 1_{\{\ell\in S\}} + e_{=}(\Psi')\, 1_{\{i_0 = \ell\}}\Big).$$

By exchangeability of (i_0, S),

$$\mathbb{P}(i_0 = i) = \frac{1}{n}, \quad \mathbb{P}(i \in S) = \frac{k-1}{n},$$
$$\mathbb{P}(i \in S,\, \ell = i_0) = \frac{k-1}{n(n-1)}, \quad \mathbb{P}(i \in S,\, \ell \in S) = \frac{(k-1)(k-2)}{n(n-1)},$$

and $\{i_0 = i\} \cap \{i_0 = \ell\} = \emptyset$. Taking the expectation with respect to (i_0, S) in the two previous displays yields (23) and (24), which ends the proof. □

The first consequence extends Bousebata et al. (2023, Lemma 2) to the random threshold framework; it corrects the deterministic-level statement by the diagonal contribution of the observation achieving the order statistic.

Lemma 3. *Under the assumptions of Lemma 2, let $h : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $\mathbb{E}(|h(Y)|1_{\{Y>y\}}) < +\infty$. Then, for any $1 \le i \le n$,*

$$m_{k,y}(h(Y_i)) = \frac{k-1}{n}\,\frac{1}{\bar{F}(y)}\int_y^{+\infty} h(t)f(t)\,\mathrm{d}t + \frac{h(y)}{n}; \tag{25}$$

in particular, $m_{k,y}(1) = k/n$. If moreover $h \in \mathcal{RV}_\rho(+\infty)$ with $\rho\gamma < 1$ and (8) holds, then, as $y \to +\infty$,

$$m_{k,y}(h(Y_i)) = \frac{k-1}{n}\,\frac{h(y)}{1-\rho\gamma}\big(1 + o(1)\big) + \frac{h(y)}{n}, \tag{26}$$

where the $o(1)$ term depends neither on k nor on n.

Proof. Identity (25) is (23) applied to $\Psi(x,t) = h(t)$, since

$$e_>(\Psi) = \bar{F}(y)^{-1}\int_y^{+\infty} h(t)f(t)\,\mathrm{d}t \quad \text{and} \quad e_=(\Psi) = h(y).$$

The asymptotic equivalent $\bar{F}(y)^{-1}\int_y^{+\infty} h(t)f(t)\mathrm{d}t \to h(y)/(1-\rho\gamma)$ as $y \to +\infty$, which involves neither k nor n, is provided by Bousebata et al. (2023, Lemma 2), which ends the proof. □

Remark 9. The diagonal term $h(y)/n$ in (25) is negligible in all our uses, where $k = k_n \to +\infty$, but it must be kept for the identity to be exact, as the check $m_{k,y}(1) = k/n$ shows. In the sequel, Bousebata et al. (2023, Lemma 2) remains the reference for tail-moments at a deterministic level, and Lemma 3 whenever the threshold is the random order statistic $Y_{n-k+1,n}$.

The second consequence is the exact second conditional moment of empirical tail-moments. Note that the inflation factors of the two conditional variances below are $(k-1)/n^2$ and $1/n^2$, and not $1/n$ times the corresponding probabilities: the indicator functions $1_{\{Y_i \ge Y_{n-k+1,n}\}}$, $1 \le i \le n$, are not conditionally independent given $Y_{n-k+1,n}$, and the exact combinatorial structure of Lemma 2 is what makes the computation below valid.

Lemma 4. *Under the assumptions of Lemma 2, let $Z_i := \Psi(X_i, Y_i)$, $1 \le i \le n$, for some measurable function $\Psi : H \times \mathbb{R} \to H$ such that $e_>(\|\Psi\|^2) < +\infty$ and $e_=(\|\Psi\|^2) < +\infty$, and set*

$$\hat{m}_Z(Y_{n-k+1,n}) := \frac{1}{n}\sum_{i=1}^{n} Z_i 1_{\{Y_i \ge Y_{n-k+1,n}\}},$$

$$\mu_> := \mathbb{E}_{\mathcal{B}}(Z_1 \mid Y_1 > y) \quad \textit{and} \quad \mu_= := \mathbb{E}_{\mathcal{B}}(Z_1 \mid Y_1 = y).$$

Then $m_{k,y}(Z_1) = \frac{k-1}{n}\mu_> + \frac{1}{n}\mu_=$ and

$$\begin{aligned}\mathbb{E}_{k,y}\left(\|\hat{m}_Z(Y_{n-k+1,n})\|^2\right) &= \|m_{k,y}(Z_1)\|^2 + \frac{k-1}{n^2}\Big(e_>(\|\Psi\|^2) - \|\mu_>\|^2\Big) + \frac{1}{n^2}\Big(e_=(\|\Psi\|^2) - \|\mu_=\|^2\Big)\\ &\le \|m_{k,y}(Z_1)\|^2 + \frac{1}{n}\,m_{k,y}\big(\|Z_1\|^2\big).\end{aligned} \tag{27}$$

The same statements hold for real-valued Z_i, replacing inner products and norms accordingly.

Proof. The formula for $m_{k,y}(Z_1)$ is (23); it extends to H-valued integrands by testing against a countable orthonormal basis, since bounded linear forms commute with Bochner conditional expectations. Next, expanding the squared norm,

$$\mathbb{E}_{k,y}\left(\|\hat{m}_Z(Y_{n-k+1,n})\|^2\right) = \frac{1}{n^2}\sum_{i\neq\ell}\mathbb{E}_{k,y}\left(\langle Z_i 1_{\{Y_i\geq y\}}, Z_\ell 1_{\{Y_\ell\geq y\}}\rangle\right) + \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{k,y}\left(\|Z_i\|^2 1_{\{Y_i\geq y\}}\right).$$

Conditionally on (i_0, S) from the representation of Lemma 2, the random variables $Z_i 1_{\{Y_i\geq y\}}$, $1\leq i\leq n$, are mutually independent, so that, for any $i\neq\ell$, the unconditional version of Lemma 1 applied under $\mathbb{P}(\cdot \mid (i_0,S),\, Y_{n-k+1,n}=y)$ factorizes each off-diagonal term into the inner product of the two corresponding conditional means. Averaging over (i_0,S) with the probabilities collected in the proof of Lemma 2 gives, exactly as in (24),

$$\sum_{i\neq\ell}\mathbb{E}_{k,y}\left(\langle Z_i 1_{\{Y_i\geq y\}}, Z_\ell 1_{\{Y_\ell\geq y\}}\rangle\right) = (k-1)(k-2)\,\|\mu_>\|^2 + 2(k-1)\,\langle\mu_>,\mu_=\rangle.$$

Besides, (23) applied to $\|\Psi\|^2$ yields $\sum_{i=1}^n \mathbb{E}_{k,y}(\|Z_i\|^2 1_{\{Y_i\geq y\}}) = (k-1)\,e_>(\|\Psi\|^2) + e_=(\|\Psi\|^2)$, while

$$\|m_{k,y}(Z_1)\|^2 = \Big(\frac{k-1}{n}\Big)^2\|\mu_>\|^2 + \frac{2(k-1)}{n^2}\langle\mu_>,\mu_=\rangle + \frac{1}{n^2}\|\mu_=\|^2.$$

Collecting the last three displays, the cross terms $\langle\mu_>,\mu_=\rangle$ cancel out and the stated identity follows from $(k-1)(k-2)-(k-1)^2 = -(k-1)$. At last, the bound (27) is obtained by discarding the nonpositive terms $-\|\mu_>\|^2$ and $-\|\mu_=\|^2$ and recognizing $\frac{1}{n}m_{k,y}(\|Z_1\|^2) = \frac{k-1}{n^2}\,e_>(\|\Psi\|^2) + \frac{1}{n^2}\,e_=(\|\Psi\|^2)$ from (23), which ends the proof. □

Remark 10. If, in addition, $\mathbb{E}_{\mathcal{B}}(Z_1 \mid Y_1 = t) = 0$ for almost every $t\geq y$, then $\mu_> = \mu_= = 0$, $m_{k,y}(Z_1) = 0$, and (27) is an equality:

$$\mathbb{E}_{k,y}\left(\|\hat{m}_Z(Y_{n-k+1,n})\|^2\right) = \frac{1}{n}\,m_{k,y}(\|Z_1\|^2).$$

This covers in particular the conditionally centred residual ξ of Section 3 through $Z_i = \xi_i$, and makes rigorous the corresponding step in the proof of Proposition 2.

A.2. **Empirical tail-moments.** The next lemma establishes a bound on the (random) tail moments of $\varphi(Y)\varepsilon$.

Lemma 5. *Assume that* (8), (4), (7), (9), (11) *and* (12) *hold. Let $k := k_n \to +\infty$ be an integer deterministic sequence such that $k/n\to 0$ and $y_{n,k}\sim U(n/k)$ as $n\to+\infty$. Let $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$. Then,*

$$\frac{\|\hat{m}_{\varphi(Y)\varepsilon}(Y_{n-k+1,n})\|}{m_{\varphi g(Y)}(y_{n,k})} = O_{\mathbb{P}}\left(\delta_{n,k}\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Proof. For clarity, denote $Z_1 := \varphi(Y_1)\varepsilon_1$ and $Y_{n,k} := Y_{n-k+1,n}$. Let $y>0$. Highlighting the dependence on y, we denote the conditional expectation given $\{Y_{n,k}=y\}$ by $\mathbb{E}_{k,y}(\cdot) := \mathbb{E}(\cdot|Y_{n,k}=y)$ and, similarly, let us denote the conditional tail-moments of any generic random variable W_i, $1\leq i\leq n$, by

$$m_{k,y}(W_i) := \mathbb{E}_{k,y}(W_i 1_{\{Y_i\geq Y_{n,k}\}}).$$

We start by computing the expectation of $\|\hat{m}_Z(Y_{n,k})\|^2$ which is by construction,

$$\|\hat{m}_Z(Y_{n,k})\|^2 = \frac{1}{n^2}\sum_{i_1,i_2=1}^{n}\langle Z_{i_1}, Z_{i_2}\rangle 1_{\{Y_{i_1}\geq Y_{n,k}\}}1_{\{Y_{i_2}\geq Y_{n,k}\}}.$$

Taking the conditional expectation and splitting the sum, it follows,

$$\begin{aligned}\mathbb{E}_{k,y}(\|\hat{m}_Z(Y_{n,k})\|^2) = \frac{1}{n^2}\sum_{i_1\neq i_2}^{n}\mathbb{E}_{k,y}\left(\langle Z_{i_1}1_{\{Y_{i_1}\geq Y_{n,k}\}}, Z_{i_2}1_{\{Y_{i_2}\geq Y_{n,k}\}}\rangle\right)\\ + \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{k,y}\left(\|Z_i\|^2 1_{\{Y_i\geq Y_{n,k}\}}\right).\end{aligned}$$

Lemma 4 and equidistribution entail

$$\mathbb{E}_{k,y}(\|\hat{m}_Z(Y_{n,k})\|^2) \leq \|m_{k,y}(Z_1)\|^2 + \frac{1}{n}\, m_{k,y}(\|Z_1\|^2).$$

The main tool of the proof is the conditional Markov's inequality which gives, for any $\eta' > 0$,

$$\mathbb{P}_{k,y}\left(\frac{\|\hat{m}_Z(Y_{n,k})\|}{m_{\varphi g(Y)}(y_{n,k})} > \eta'\right) \leq \frac{1}{\eta'^2}\,\mathbb{E}_{k,y}\left(\|\hat{m}_Z(y)\|^2\right) m^{-2}_{\varphi g(Y)}(y_{n,k}). \tag{28}$$

Here $y_{n,k}$ is deterministic, so $m_{\varphi g(Y)}(y_{n,k})$ is a tail-moment at a deterministic level. According to Bousebata et al. (2023, Lemma 2) under (11), there exists some constant $c \in (0, +\infty)$ independent of n such that, for n large enough, Markov's inequality (28) becomes

$$\mathbb{P}_{k,y}\left(\frac{\|\hat{m}_Z(Y_{n,k})\|}{m_{\varphi g(Y)}(y_{n,k})} > \eta'\right) \leq \frac{c}{\eta'^2}(\varphi g \bar{F})^{-2}(y_{n,k})\left(\|m_{k,y}(Z_1)\|^2 \vee \frac{1}{n}m_{k,y}(\|Z_1\|^2)\right).$$

The next step is to bound the term involving the conditional tail-moments. To this end, observe that the conditional expectation is a contraction for the norm of H, by the conditional Jensen inequality for Bochner integrals, so that $\|m_{k,y}(Z_1)\| \leq m_{k,y}(\|Z_1\|) = m_{k,y}(\varphi(Y_1)\|\varepsilon_1\|)$, where the last equality uses $\varphi \geq 0$. By conditional Cauchy-Schwarz's inequality in $L^2(\Omega, \mathcal{A}, \mathbb{P})$, we may then bound

$$\|m_{k,y}(Z_1)\|^2 \leq m^2_{k,y}(\varphi(Y_1)\|\varepsilon_1\|) \leq m_{k,y}(\|Z_1\|^2)\, m_{k,y}(1), \tag{29}$$

and therefore

$$\|m_{k,y}(Z_1)\|^2 \vee \frac{1}{n}m_{k,y}(\|Z_1\|^2) \leq m_{k,y}(\|Z_1\|^2)\left(\frac{1}{n} \vee m_{k,y}(1)\right).$$

Besides, Lemma 3 yields $m_{k,y}(1) = k/n$, so that

$$\|m_{k,y}(Z_1)\|^2 \vee \frac{1}{n}m_{k,y}(\|Z_1\|^2) \leq \frac{k}{n}\, m_{k,y}(\|Z_1\|^2). \tag{30}$$

The focus now turns to the tail-moment of $\|Z_1\|^2$ in the previous inequality. The q-moment of $\|\varepsilon_1\|$ is finite under (9) so that Hölder's inequality yields

$$m_{k,y}(\|Z_1\|^2) \leq \mathbb{E}_{k,y}\left(\|\varepsilon_1\|^q\right)^{2/q} m^{1-2/q}_{k,y}(\varphi(Y_1)^{2q/(q-2)}). \tag{31}$$

Note that the latter tail moment is finite if $2\tau q/(q-2) < 1/\gamma$ which is satisfied for any $q > 2$ and $\gamma \in (0, 1)$ whenever $\tau \leq 0$. This existence condition is also satisfied when $\tau > 0$ under (11)-(12). Indeed, the condition at hand is equivalent to $\tau < (1 - 2/q)/(2\gamma)$ while (11) is the same as $\tau < (1 - 2\kappa\gamma)/(2\gamma)$ and (12)

writes as $2/q < 2\kappa\gamma$. Moreover, $\varphi^{2q/(q-2)} \in \mathcal{RV}_{2\tau q/(q-2)}(+\infty)$ so that Lemma 3 yields, uniformly in $y \geq y_{n,k}/2$, as $n \to +\infty$:

$$m_{k,y}^{1-2/q}(\varphi(Y_1)^{2q/(q-2)}) \sim \left(1 - \frac{2\tau q\gamma}{q-2}\right)^{(2/q)-1} \varphi^2(y) \left(\frac{k-1}{n}\right)^{1-2/q}. \tag{32}$$

It follows from (31) and (32) that there exists a constant $c \in (0, +\infty)$ independent of n such that for n large enough, uniformly in $y \geq y_{n,k}/2$,

$$m_{k,y}(\|Z_1\|^2) \leq c\, \mathbb{E}_{k,y}\left(\|\varepsilon_1\|^q\right)^{2/q} \varphi^2(y) \left(\frac{k-1}{n}\right)^{1-2/q}. \tag{33}$$

Combining (30) and (33), one has for some constant $c > 0$ independent of n and for n large enough,

$$\|m_{k,y}(Z_1)\|^2 \vee \frac{1}{n} m_{k,y}(\|Z_1\|^2) \leq c\, \mathbb{E}_{k,y}\left(\|\varepsilon_1\|^q\right)^{2/q} \varphi^2(y) \left(\frac{k-1}{n}\right)^{2(1-1/q)}.$$

Going back to Markov's bound (28), for some constant $c \in (0, +\infty)$ independent of n,

$$\mathbb{P}_{k,y}\left(\frac{\|\hat{m}_Z(Y_{n,k})\|}{m_{\varphi g(Y)}(y_{n,k})} > \eta'\right) \leq \frac{c}{\eta'^2}\, \mathbb{E}_{k,y}\left(\|\varepsilon_1\|^q\right)^{2/q} \Delta_{n,k}(y),$$

where

$$\Delta_{n,k}(y) := (\varphi g \bar{F})^{-2}(y_{n,k}) \left(\frac{k-1}{n}\right)^{2(1-1/q)} \varphi^2(y). \tag{34}$$

Define two sequences $y_{n,k}^+ := 3y_{n,k}/2$ and $y_{n,k}^- := y_{n,k}/2$. Since $\varphi \in \mathcal{RV}_\tau(+\infty)$, the uniform convergence theorem for regularly varying functions yields

$$\sup_{y \in [y_{n,k}^-, y_{n,k}^+]} \frac{\varphi(y)}{\varphi(y_{n,k})} = \sup_{\lambda \in [1/2, 3/2]} \frac{\varphi(\lambda y_{n,k})}{\varphi(y_{n,k})} \longrightarrow \left(\frac{1}{2}\right)^\tau \vee \left(\frac{3}{2}\right)^\tau \quad \text{as } n \to +\infty.$$

Hence, when $y \in [y_{n,k}^-, y_{n,k}^+]$ and n is large enough, $\varphi^2(y) \leq c\varphi^2(y_{n,k})$ for some constant $c > 0$ independent of n. At last, $y_{n,k} \sim U(n/k)$ implies $\bar{F}(y_{n,k}) \sim \bar{F}(U(n/k)) = k/n$ as $n \to +\infty$ since asymptotic equivalences are stable under regularly-varying composition, and therefore, for some other constant $c > 0$,

$$\Delta_{n,k}(y) \leq c g^{-2}(y_{n,k}) \left(\frac{k}{n}\right)^{-2/q}.$$

Taking account of (9), it follows that there exists a constant $c > 0$ independent of n such that for n large enough,

$$\sup_{y \in [y_{n,k}^-, y_{n,k}^+]} \left\{ \mathbb{E}_{k,y}\left(\|\varepsilon_1\|^q\right)^{2/q} \Delta_{n,k}(y) \right\} \leq c\delta_{n,k}^2.$$

Now, let any $\eta > 0$. Since η' is arbitrary, one may pick η' of the form $\delta_{n,k}(2c/\eta)^{1/2}$ and combining with what precedes, we have shown that (28) becomes, when $y_{n,k}^- \leq y \leq y_{n,k}^+$,

$$\mathbb{P}_{k,y}\left(\delta_{n,k}^{-1} \frac{\|\hat{m}_Z(Y_{n,k})\|}{m_{\varphi g(Y)}(y_{n,k})} > (2c/\eta)^{1/2}\right) \leq \frac{\eta}{2}. \tag{35}$$

In order to derive the result, we need in fact to integrate the conditional probability (35) with respect to y on the whole real line, in view of Bayes' formula. After splitting the integral domain, the proof is complete if the next quantity is negligible when n is large,

$$P_{n,k,\eta}(-\infty,+\infty) = P_{n,k,\eta}(-\infty, y_{n,k}^-) + P_{n,k,\eta}(y_{n,k}^-, y_{n,k}^+) + P_{n,k,\eta}(y_{n,k}^+, +\infty),$$

where

$$P_{n,k,\eta}(a,b) := \int_a^b \mathbb{P}_{k,y}\left(\delta_{n,k}^{-1}\frac{\|\hat{m}_Z(Y_{n,k})\|}{m_{\varphi g(Y)}(y_{n,k})} > (2c/\eta)^{1/2}\right) f_{Y_{n,k}}(y)\mathrm{d}y,$$

for any $-\infty \le a \le b \le +\infty$. First, note that (35) readily provides that $P_{n,k,\eta}(y_{n,k}^-, y_{n,k}^+) \le \eta/2$ for any $\eta > 0$. Concerning the remaining two terms, we bound the conditional probability by one, use the complementary event and invoke the fact that $Y_{n,k}/U(n/k) \xrightarrow{\mathbb{P}} 1$ as $n \to +\infty$ to get, for n large enough:

$$P_{n,k,\eta}(-\infty, y_{n,k}^-) + P_{n,k,\eta}(y_{n,k}^+, +\infty) \le 1 - \mathbb{P}\left(y_{n,k}^- < Y_{n,k} < y_{n,k}^+\right) \le \frac{\eta}{2}.$$

By Bayes' formula, the conclusion follows as we have shown that, for any $\eta > 0$ and n large enough,

$$\mathbb{P}\left(\delta_{n,k}^{-1}\frac{\|\hat{m}_Z(Y_{n,k})\|}{m_{\varphi g(Y)}(y_{n,k})} > (2c/\eta)^{1/2}\right) \le \frac{\eta}{2} + \frac{\eta}{2} = \eta,$$

which is exactly the claimed $O_{\mathbb{P}}(\delta_{n,k})$ bound.

□

Finally, Lemma 6 below is dedicated to the control of the (random) tail moments of $\langle \beta, \varphi(Y)\varepsilon\rangle$.

Lemma 6. *Let any $\beta \in H$ with $\|\beta\| = 1$ and suppose (8) holds. Assume that (4), (7), (9), (11) and (12) hold. Let $k := k_n \to +\infty$ be an integer deterministic sequence such that $k/n \to 0$ and $y_{n,k} \sim U(n/k)$ as $n \to +\infty$. Let $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$. Then,*

$$\frac{\hat{m}_{\langle \beta, \varphi(Y)\varepsilon\rangle}(Y_{n-k+1,n})}{m_{\varphi g(Y)}(y_{n,k})} = O_{\mathbb{P}}\left(\delta_{n,k}\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Proof. Let $y > 0$. Again, we denote $Z_1 := \varphi(Y_1)\varepsilon_1$, $Y_{n,k} := Y_{n-k+1,n}$, the conditional expectation given $\{Y_{n,k} = y\}$ by $\mathbb{E}_{k,y}(\cdot) := \mathbb{E}(\cdot | Y_{n,k} = y)$ and $m_{k,y}(W_i) := \mathbb{E}_{k,y}(W_i 1_{\{Y_i \ge Y_{n,k}\}})$. The proof relies on the conditional Markov's inequality which states that

$$\mathbb{P}_{k,y}\left(\left|\frac{\hat{m}_{\langle \beta, \varphi(Y)\varepsilon\rangle}(Y_{n,k})}{m_{\varphi g(Y)}(y_{n,k})}\right| > \eta'\right) \le \frac{1}{\eta'^2}\frac{\mathbb{E}_{k,y}\left(\hat{m}^2_{\langle \beta, \varphi(Y)\varepsilon\rangle}(Y_{n,k})\right)}{m^2_{\varphi g(Y)}(y_{n,k})}, \quad \forall \eta' > 0. \quad (36)$$

The goal is to show that the RHS in (36) converges to zero for large n. Let us start with a preliminary computation of the expectation. One has

$$\hat{m}^2_{\langle \beta, \varphi(Y)\varepsilon\rangle}(Y_{n,k}) = \frac{1}{n^2}\sum_{i,j=1}^n \langle \beta, \varphi(Y_i)\varepsilon_i\rangle\langle \beta, \varphi(Y_j)\varepsilon_j\rangle 1_{\{Y_i \ge Y_{n,k}\}} 1_{\{Y_j \ge Y_{n,k}\}}.$$

Taking the conditional expectation and applying the real-valued case of Lemma 4 to $z_i := \langle \beta, \varphi(Y_i)\varepsilon_i \rangle$, $1 \leq i \leq n$, together with equidistribution, yield

$$\mathbb{E}_{k,y}\left(\hat{m}^2_{\langle \beta, \varphi(Y)\varepsilon \rangle}(Y_{n,k})\right) \leq \mathbb{E}_{k,y}\left(\varphi(Y_1)\langle \beta, \varepsilon_1 \rangle 1_{\{Y_1 \geq Y_{n,k}\}}\right)^2 + \frac{1}{n}\,\mathbb{E}_{k,y}\left(\varphi^2(Y_1)\langle \beta, \varepsilon_1 \rangle^2 1_{\{Y_1 \geq Y_{n,k}\}}\right). \tag{37}$$

On the one hand, Cauchy-Schwarz's inequality in H with $\|\beta\| = 1$ as well as the bound (33) in the proof of Lemma 5, which is allowed since (9), (11) and (12) hold, entail that there exists a constant $c \in (0, +\infty)$ independent of n such that for n large enough, uniformly in $y \geq y_{n,k}/2$,

$$\mathbb{E}_{k,y}\left(\varphi^2(Y_1)\langle \beta, \varepsilon_1 \rangle^2 1_{\{Y_1 \geq Y_{n,k}\}}\right) \leq m_{k,y}(\varphi^2(Y)\|\varepsilon\|^2) \leq c\,\mathbb{E}_{k,y}\left(\|\varepsilon\|^q\right)^{2/q} \varphi^2(y)\left(\frac{k-1}{n}\right)^{1-2/q}. \tag{38}$$

On the other hand, recalling that $Z_1 := \varphi(Y_1)\varepsilon_1$, one may apply Cauchy-Schwarz's inequality in H to obtain

$$\mathbb{E}_{k,y}\left(\varphi(Y_1)\langle \beta, \varepsilon_1 \rangle 1_{\{Y_1 \geq Y_{n,k}\}}\right)^2 \leq \mathbb{E}_{k,y}\left(\varphi(Y_1)\|\varepsilon_1\| 1_{\{Y_1 \geq Y_{n,k}\}}\right)^2 = m^2_{k,y}(\|Z_1\|). \tag{39}$$

The focus now turns to the conditional tail-moment of $\|Z_1\|$ in the previous inequality. The q-moment of $\|\varepsilon\|$ is finite under (9) so that Hölder's inequality allows us to write,

$$m^2_{k,y}(\|Z_1\|) \leq \mathbb{E}_{k,y}\left(\|\varepsilon_1\|^q\right)^{2/q} m^{2(1-1/q)}_{k,y}(\varphi^{q/(q-1)}(Y_1)). \tag{40}$$

Lemma 3 applies since $\tau q/(q-1) < 1/\gamma$ whenever $\tau \leq 0$. This very condition is also satisfied when $\tau > 0$ under (11) and (12). Indeed, the condition at hand is equivalent to $\tau < (1-1/q)/\gamma$ while (11) gives $\tau < (1/2 - \kappa\gamma)/\gamma < (1-\kappa\gamma)/\gamma$ and (12) writes as $1/q < \kappa\gamma$ which concludes. Finally, it yields, for some constant $c > 0$ independent of n, for n large enough, uniformly in $y \geq y_{n,k}/2$,

$$m^{2(1-1/q)}_{k,y}(\varphi^{q/(q-1)}(Y_1)) \leq c\varphi^2(y)\left(\frac{k-1}{n}\right)^{2(1-1/q)}. \tag{41}$$

Collecting (39), (40) and (41) entails

$$\mathbb{E}_{k,y}\left(\varphi(Y_1)\langle \beta, \varepsilon_1 \rangle 1_{\{Y_1 \geq Y_{n,k}\}}\right)^2 \leq c\,\mathbb{E}_{k,y}\left(\|\varepsilon_1\|^q\right)^{2/q} \varphi^2(y)\left(\frac{k-1}{n}\right)^{2(1-1/q)}, \tag{42}$$

for n large enough, where c is another constant independent of n. It readily follows from (37), (38) and (42) that the numerator of the RHS in (36) can be upper bounded as

$$\mathbb{E}_{k,y}\left(\hat{m}^2_{\langle \beta, \varphi(Y)\varepsilon \rangle}(Y_{n,k})\right) \leq c\,\mathbb{E}_{k,y}\left(\|\varepsilon_1\|^q\right)^{2/q} \varphi^2(y)\left(\frac{k-1}{n}\right)^{2(1-1/q)}. \tag{43}$$

Besides, the denominator of the RHS in (36) is controlled through Bousebata et al. (2023, Lemma 2), for some constant $c > 0$ independent of n and for $n \to +\infty$,

$$m^2_{\varphi g(Y)}(y_{n,k}) \sim c\varphi^2(y_{n,k}) g^2(y_{n,k}) \bar{F}^2(y_{n,k}). \tag{44}$$

Thus, (43) altogether with (44) give, for some constant $c > 0$ independent of n, for n large enough,

$$\frac{\mathbb{E}_{k,y}\left(\hat{m}^2_{\langle\beta,\varphi(Y)\varepsilon\rangle}(Y_{n,k})\right)}{m^2_{\varphi g(Y)}(y_{n,k})} \leq c(\varphi g\bar{F})^{-2}(y_{n,k})\,\mathbb{E}_{k,y}\left(\|\varepsilon\|^q\right)^{2/q}\varphi^2(y)\left(\frac{k-1}{n}\right)^{2(1-1/q)}.$$

Overall, the Markov's inequality (36) becomes, for some constant $c \in (0,+\infty)$ independent of n, for any n large enough and any $\eta' > 0$,

$$\mathbb{P}_{k,y}\left(\left|\frac{\hat{m}_{\langle\beta,\varphi(Y)\varepsilon\rangle}(Y_{n,k})}{m_{\varphi g(Y)}(y_{n,k})}\right| > \eta'\right) \leq \frac{c}{\eta'^2}\,\mathbb{E}_{k,y}\left(\|\varepsilon\|^q\right)^{2/q}\Delta_{n,k}(y),$$

where $\Delta_{n,k}(y)$ is defined by (34) in the proof of Lemma 5.

The remaining part of the proof is a mere repetition of the same steps as in the proof of Lemma 5. □

A.3. **Consistency.** Theorem 1 follows from the upcoming Proposition 3, which is proved based on Lemma 5 and Lemma 6, combined with Proposition 4: the triangle inequality entails

$$\begin{aligned}\|\tilde{\beta}_\varphi - \beta\| &\leq \|\tilde{\beta}_\varphi - \hat{\beta}_\varphi\| + \|\hat{\beta}_\varphi - \beta\| \\ &= O_{\mathbb{P}}\big(N^{-\alpha_b} \vee \{N^{-\alpha_e}\delta_{n,k}\}\big) + O_{\mathbb{P}}(\delta_{n,k}) = O_{\mathbb{P}}\big(\delta_{n,k} \vee N^{-\alpha_b}\big),\end{aligned}$$

all quantities being evaluated at $Y_{n-k+1,n}$, where the last equality uses $N^{-\alpha_e}\delta_{n,k} \leq \delta_{n,k}$.

Proposition 3. *Assume* (6), (4), (5), (7), (8), (9), (11) *and* (12) *hold, with* $\sqrt{k}A(n/k) = O(1)$. *Assume* φ *and* g *are continuously differentiable in a neighbourhood of infinity such that* $t(\varphi g)'(t)/(\varphi g)(t) \to \tau + \kappa$ *as* $t \to +\infty$. *Let* $k := k_n \to +\infty$ *be an intermediate sequence and let* $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$ *where* $y_{n,k} \sim U(n/k)$ *as* $n \to +\infty$.

$$\|\hat{\beta}_\varphi(Y_{n-k+1,n}) - \beta\| = O_{\mathbb{P}}(\delta_{n,k}) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Proof of Proposition 3. Denote $Y_{n,k} := Y_{n-k+1,n}$. Under (6), the estimator $\hat{m}_{X\varphi(Y)}$ may be expressed as

$$\hat{m}_{X\varphi(Y)}(Y_{n,k}) = \hat{m}_{\varphi g(Y_1)}(Y_{n,k})\beta + \hat{m}_{\varphi(Y_1)\varepsilon_1}(Y_{n,k}).$$

So that, since $\|\beta\| = 1$, one may write the inner product between β and $\hat{m}_{X\varphi(Y)}(Y_{n,k})$ as

$$\langle\hat{m}_{X\varphi(Y)}(Y_{n,k}), \beta\rangle = \hat{m}_{\varphi g(Y_1)}(Y_{n,k}) + \hat{m}_{\langle\varphi(Y_1)\varepsilon_1,\beta\rangle}(Y_{n,k}),$$

and

$$\begin{aligned}\|\hat{m}_{X\varphi(Y)}(Y_{n,k})\|^2 = \hat{m}^2_{\varphi g(Y_1)}(Y_{n,k}) &+ \|\hat{m}_{\varphi(Y_1)\varepsilon_1}(Y_{n,k})\|^2 \\ &+ 2\hat{m}_{\varphi g(Y_1)}(Y_{n,k})\hat{m}_{\langle\varphi(Y_1)\varepsilon_1,\beta\rangle}(Y_{n,k}).\end{aligned}$$

It follows that

$$1 - \frac{\langle\hat{m}_{X\varphi(Y)}(Y_{n,k}), \beta\rangle^2}{\|\hat{m}_{X\varphi(Y)}(Y_{n,k})\|^2} = \frac{\|\hat{m}_{\varphi(Y_1)\varepsilon_1}(Y_{n,k})\|^2 - \hat{m}^2_{\langle\varphi(Y_1)\varepsilon_1,\beta\rangle}(Y_{n,k})}{\hat{m}^2_{\varphi g(Y_1)}(Y_{n,k}) + 2\hat{m}_{\varphi g(Y_1)}(Y_{n,k})\hat{m}_{\langle\varphi(Y_1)\varepsilon_1,\beta\rangle}(Y_{n,k}) + \|\hat{m}_{\varphi(Y_1)\varepsilon_1}(Y_{n,k})\|^2},$$

which equals

$$\frac{\big(\|\hat{m}_{\varphi(Y_1)\varepsilon_1}(Y_{n,k})\|/\hat{m}_{\varphi g(Y_1)}(Y_{n,k})\big)^2 - \big(\hat{m}_{\langle\varphi(Y_1)\varepsilon_1,\beta\rangle}(Y_{n,k})/\hat{m}_{\varphi g(Y_1)}(Y_{n,k})\big)^2}{1+2\big(\hat{m}_{\langle\varphi(Y_1)\varepsilon_1,\beta\rangle}(Y_{n,k})/\hat{m}_{\varphi g(Y_1)}(Y_{n,k})\big)+\big(\|\hat{m}_{\varphi(Y_1)\varepsilon_1}(Y_{n,k})\|/\hat{m}_{\varphi g(Y_1)}(Y_{n,k})\big)^2}.$$

We have everything needed to apply Stupfler (2019, Theorem 2) to the function φg. Indeed, $t(\varphi g)'(t)/(\varphi g)(t) \to a := \tau + \kappa$ as $t \to +\infty$ with $a > 0$ in view of (11), so that φg is ultimately increasing and $(\varphi g)' \in \mathcal{RV}_{a-1}(+\infty)$; the second-order condition on U is (5); the moment condition $0 < 2a\gamma < 1$ is exactly (11); and the bias condition $\sqrt{k}A(n/k) = O(1)$ is assumed. The required conditions for Lemma 5 and Lemma 6 are also fulfilled. Therefore, their combination together with Bousebata et al. (2023, Lemma 2) and the well-known fact that $\bar{F}(y_{n,k}) \sim \bar{F}(U(n/k)) = k/n$ since $y_{n,k} \sim U(n/k)$, allows us to chain asymptotic equivalences in probability:

$$\|\hat{m}_{X\varphi(Y)}(Y_{n,k})\|^2 = \hat{m}^2_{\varphi g(Y_1)}(Y_{n,k})(1+o_{\mathbb{P}}(1)) = m^2_{\varphi g(Y)}(y_{n,k})(1+o_{\mathbb{P}}(1)). \quad (45)$$

Whence after a first order Taylor expansion, $1-\langle\hat{m}_{X\varphi(Y)}(Y_{n,k}),\beta\rangle^2/\|\hat{m}_{X\varphi(Y)}(Y_{n,k})\|^2$ is asymptotically equivalent, with respect to the convergence in probability, to

$$\frac{\big(\|\hat{m}_{\varphi(Y_1)\varepsilon_1}(Y_{n,k})\|/m_{\varphi g(Y)}(y_{n,k})\big)^2 - \big(\hat{m}_{\langle\varphi(Y_1)\varepsilon_1,\beta\rangle}(Y_{n,k})/m_{\varphi g(Y)}(y_{n,k})\big)^2}{1+2\big(\hat{m}_{\langle\varphi(Y_1)\varepsilon_1,\beta\rangle}(Y_{n,k})/m_{\varphi g(Y)}(y_{n,k})\big)+\big(\|\hat{m}_{\varphi(Y_1)\varepsilon_1}(Y_{n,k})\|/m_{\varphi g(Y)}(y_{n,k})\big)^2}.$$

Again, Lemma 5 and Lemma 6, together with a Taylor expansion, yield

$$1 - \frac{\langle\hat{m}_{X\varphi(Y)}(Y_{n,k}),\beta\rangle^2}{\|\hat{m}_{X\varphi(Y)}(Y_{n,k})\|^2} = O_{\mathbb{P}}(\delta^2_{n,k}) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Finally, note that $\|\hat{\beta}_\varphi(Y_{n,k}) - \beta\|^2 = 2(1 - \langle\hat{m}_{X\varphi(Y)}(Y_{n,k}),\beta\rangle/\|\hat{m}_{X\varphi(Y)}(Y_{n,k})\|)$ to conclude the proof with the classic identity $1-\lambda^2 = (1+\lambda)(1-\lambda)$, where $\lambda := \langle\hat{m}_{X\varphi(Y)}(Y_{n,k}),\beta\rangle/\|\hat{m}_{X\varphi(Y)}(Y_{n,k})\|$. Indeed, $\langle\hat{m}_{X\varphi(Y)}(Y_{n,k}),\beta\rangle = \hat{m}_{\varphi g(Y_1)}(Y_{n,k})(1+o_{\mathbb{P}}(1))$ with $\hat{m}_{\varphi g(Y_1)} \geq 0$, so that $1+\lambda \geq 1$ with probability tending to one and $0 \leq 1-\lambda \leq 1-\lambda^2 = O_{\mathbb{P}}(\delta^2_{n,k})$. □

The next proposition controls the effect of the grid-based reconstruction on the FEPLS estimator; its first assertion replaces the former intermediate result on $\|\hat{m}_{\tilde{X}\varphi(Y)}\|$.

Proposition 4. *Assume* (6) *holds with* $\beta \in \mathcal{C}^{0,\alpha_b}([0,1],\mathbb{R})$ *for some* $\alpha_b \in (0,1]$, *together with* (8), (4), (7), (9), (11) *and* (12). *Suppose the grid satisfies* (10) *and that* φ *and* g *are continuously differentiable in a neighbourhood of infinity such that* $t(\varphi g)'(t)/(\varphi g)(t) \to \tau+\kappa$ *as* $t \to +\infty$. *Let* $k := k_n \to +\infty$ *be an intermediate sequence such that* $\sqrt{k}A(n/k) = O(1)$ *and let* $y_{n,k} \sim U(n/k)$ *as* $n \to +\infty$. *Then, for any deterministic sequence* $N := N_n \to +\infty$, *it holds that:*

$$\big\|\hat{m}_{\tilde{X}\varphi(Y)}(Y_{n-k+1,n})\big\| = m_{\varphi g(Y)}(y_{n,k})\,\big(1+o_{\mathbb{P}}(1)\big), \quad (46)$$

$$\big\|\tilde{\beta}_\varphi(Y_{n-k+1,n}) - \hat{\beta}_\varphi(Y_{n-k+1,n})\big\| = O_{\mathbb{P}}\Big(N^{-\alpha_b} \vee \big\{N^{-\alpha_e}\,\delta_{n,k}\big\}\Big), \quad (47)$$

where $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$.

Proof. Denote $Y_{n,k} := Y_{n-k+1,n}$, $\hat{m} := \hat{m}_{X\varphi(Y)}(Y_{n,k})$, $\tilde{m} := \hat{m}_{\tilde{X}\varphi(Y)}(Y_{n,k})$ and $m := m_{\varphi g(Y)}(y_{n,k})$, and recall from the proof of Lemma 5 the two sequences $y^+_{n,k} := 3y_{n,k}/2$ and $y^-_{n,k} := y_{n,k}/2$.

We start by controlling the pathwise discretisation error. Let $\alpha \in (0,1]$ and $h \in \mathcal{C}^{0,\alpha}([0,1],\mathbb{R})$, and recall from Section 3 that $\|h\|_{0,\alpha}$ denotes the Hölder norm of h on $[0,1]$; in particular $|h(s)-h(t)| \le \|h\|_{0,\alpha}|s-t|^\alpha$ for any $s,t \in [0,1]$. Since $\max_{1\le j\le N}|I_j| \le c_0/N$ by (10) and $t_j \in I_j$, one has

$$\Big\|h - \sum_{j=1}^{N} h(t_j)1_{I_j}\Big\|^2 = \sum_{j=1}^{N}\int_{I_j}\big(h(t)-h(t_j)\big)^2\mathrm{d}t \le \|h\|_{0,\alpha}^2\left(\frac{c_0}{N}\right)^{2\alpha}\sum_{j=1}^{N}|I_j| = \|h\|_{0,\alpha}^2\left(\frac{c_0}{N}\right)^{2\alpha}. \tag{48}$$

Under (6), $\tilde{X}_i - X_i = g(Y_i)\big(\sum_{j=1}^N \beta(t_j)1_{I_j} - \beta\big) + \big(\sum_{j=1}^N \varepsilon_i(t_j)1_{I_j} - \varepsilon_i\big)$, so that (48), applied to β and to ε_i, yields almost surely, for any $1 \le i \le n$,

$$\|\tilde{X}_i - X_i\| \le c_0^{\alpha_b}\,\|\beta\|_{0,\alpha_b}\,g(Y_i)\,N^{-\alpha_b} + c_0^{\alpha_e}\,\|\varepsilon_i\|_{0,\alpha_e}\,N^{-\alpha_e}. \tag{49}$$

Since $\varphi \ge 0$ and $g \ge 0$, the triangle inequality and (49) entail

$$\begin{aligned}\|\tilde{m} - \hat{m}\| &= \Big\|\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i - X_i)\,\varphi(Y_i)1_{\{Y_i \ge Y_{n,k}\}}\Big\| \le \hat{m}_{\|\tilde{X}-X\|\varphi(Y)}(Y_{n,k}) \\ &\le c\,N^{-\alpha_b}\,\hat{m}_{\varphi g(Y)}(Y_{n,k}) + c\,N^{-\alpha_e}\,\hat{m}_{\varphi(Y)\|\varepsilon\|_{0,\alpha_e}}(Y_{n,k}),\end{aligned} \tag{50}$$

for some constant $c > 0$ depending only on c_0, α_b, α_e and $\|\beta\|_{0,\alpha_b}$.

Concerning the link-function term, all the assumptions of Proposition 3 are in force, so that (45) in its proof gives both $\hat{m}_{\varphi g(Y)}(Y_{n,k}) = m\,(1+o_{\mathbb{P}}(1))$ and $\|\hat{m}\| = m\,(1+o_{\mathbb{P}}(1))$.

Let us now turn to the noise term $S := \hat{m}_{\varphi(Y)\|\varepsilon\|_{0,\alpha_e}}(Y_{n,k})$. By equidistribution and the conditional Hölder inequality with conjugate exponents $q/(q-1)$ and q,

$$\mathbb{E}_{k,y}(S) = m_{k,y}\big(\varphi(Y_1)\|\varepsilon_1\|_{0,\alpha_e}\big) \le m_{k,y}\big(\varphi^{q/(q-1)}(Y_1)\big)^{(q-1)/q}\;\mathbb{E}_{k,y}\left(\|\varepsilon_1\|_{0,\alpha_e}^q\right)^{1/q}.$$

Note that $\tau\gamma\,q/(q-1) < 1$: this is immediate when $\tau \le 0$ while, when $\tau > 0$, (11) and (12) give $\tau\gamma < 1/2 - \kappa\gamma < 1/2 - 1/q < (q-1)/q$. Since $\varphi^{q/(q-1)} \in \mathcal{RV}_{\tau q/(q-1)}(+\infty)$, Lemma 3 thus provides a constant $c' > 0$ independent of n such that $m_{k,y}(\varphi^{q/(q-1)}(Y_1)) \le c'\,\varphi^{q/(q-1)}(y)\,k/n$ for all y large enough, uniformly in n and k. Besides, (9) provides a constant $C > 0$ independent of n such that, for n large enough, $\sup_{y\ge 0}\mathbb{E}_{k,y}(\|\varepsilon_1\|_{0,\alpha_e}^q) \le C$. Collecting these bounds, for n large enough,

$$\sup_{y\in[y_{n,k}^-,y_{n,k}^+]}\frac{\mathbb{E}_{k,y}(S)}{\varphi(y)} \le c''\,(k/n)^{1-1/q},$$

for some constant $c'' > 0$ independent of n. In addition, Bousebata et al. (2023, Lemma 2) yields $m \ge c'''\,(\varphi g\bar{F})(y_{n,k})$ for n large enough, since $(\tau+\kappa)\gamma < 1$ by (11). Now, let $M > 0$. Splitting according to the localisation of $Y_{n,k}$ in $[y_{n,k}^-, y_{n,k}^+]$ and applying the conditional Markov inequality, exactly as in the proof of Lemma 5, entail

$$\mathbb{P}\big(S > M\,\delta_{n,k}\,m\big) \le \mathbb{P}\big(Y_{n,k} \notin [y_{n,k}^-, y_{n,k}^+]\big) + \sup_{y\in[y_{n,k}^-,y_{n,k}^+]}\frac{\mathbb{E}_{k,y}(S)}{M\,\delta_{n,k}\,m}.$$

For n large enough, $\varphi \in \mathcal{RV}_\tau(+\infty)$ gives $\sup_{y\in[y^-_{n,k},y^+_{n,k}]} \varphi(y) \le c'''' \, \varphi(y_{n,k})$, while $\bar F(y_{n,k}) \sim k/n$. Hence, for n large enough,

$$\sup_{y\in[y^-_{n,k},y^+_{n,k}]} \frac{\mathbb{E}_{k,y}(S)}{M\,\delta_{n,k}\,m} \le \frac{c''c''''\,\varphi(y_{n,k})\,(k/n)^{1-1/q}}{M\,\delta_{n,k}\,c'''\,(\varphi g\bar F)(y_{n,k})} \le \frac{c''c''''}{c'''}\cdot\frac{2}{M\,\delta_{n,k}\,g(y_{n,k})\,(k/n)^{1/q}} = \frac{2c''c''''}{c'''\,M},$$

since $\delta_{n,k}\,g(y_{n,k})\,(k/n)^{1/q} = 1$ by definition of $\delta_{n,k}$. As $\mathbb{P}(Y_{n,k} \notin [y^-_{n,k}, y^+_{n,k}]) \to 0$ when $n \to +\infty$, letting $n \to +\infty$ and then $M \to +\infty$ shows that

$$S/m = O_{\mathbb{P}}(\delta_{n,k}). \tag{51}$$

It remains to conclude. Collecting (50), the link-function term and (51),

$$\frac{\|\tilde m - \hat m\|}{m} = O_{\mathbb{P}}\big(N^{-\alpha_b}\big) + N^{-\alpha_e}\,O_{\mathbb{P}}\big(\delta_{n,k}\big) = o_{\mathbb{P}}(1), \tag{52}$$

since $\delta_{n,k} \to 0$, see Proposition 3. The triangle inequality $\big|\,\|\tilde m\| - \|\hat m\|\,\big| \le \|\tilde m - \hat m\| = o_{\mathbb{P}}(m)$ combined with $\|\hat m\| = m(1 + o_{\mathbb{P}}(1))$ proves (46). Next, on the event $\{\|\tilde m - \hat m\| \le \|\hat m\|/2\}$, whose probability tends to one by (52), one has $\tilde m \neq 0$ and, for any $u, w \in H \setminus \{0\}$,

$$\Big\|\frac{u}{\|u\|} - \frac{w}{\|w\|}\Big\| \le \frac{\|u-w\|}{\|u\|} + \|w\|\,\Big|\frac{1}{\|u\|} - \frac{1}{\|w\|}\Big| \le \frac{2\,\|u-w\|}{\|u\|},$$

where the last inequality stems from $|\,\|u\| - \|w\|\,| \le \|u-w\|$. Applying this bound with $u = \hat m$ and $w = \tilde m$ yields

$$\big\|\tilde\beta_\varphi(Y_{n,k}) - \hat\beta_\varphi(Y_{n,k})\big\| \le \frac{2\,\|\tilde m - \hat m\|}{\|\hat m\|} = O_{\mathbb{P}}\Big(N^{-\alpha_b} \vee \big\{N^{-\alpha_e}\,\delta_{n,k}\big\}\Big),$$

which proves (47) and ends the proof. □

Remark 11. In (47), the noise smoothness α_e only appears multiplied by $\delta_{n,k}$: the corresponding term is thus always asymptotically negligible with respect to $\delta_{n,k} \vee N^{-\alpha_b}$, the rate appearing in Theorem 1 below. In particular, no condition relating the grid size N to n or k is required.

Proof of Proposition 2. For clarity, denote $Y_{n,k} := Y_{n-k+1,n}$, $y_{n,k} := U(n/k)$ and $m_{s^2(Y)}(y) := \mathbb{E}(s^2(Y)1_{\{Y\ge y\}})$. Let us bound everything relatively to $m_{s(Y)}(y_{n,k})$ and consider the empirical threshold at the end. Since $\mathbb{E}_{\mathcal{F},\mathcal{B}}(\xi) = 0$, one has $m_W = m_{\mu(Y)}$ and the following expansion holds:

$$\hat m_W(Y_{n,k}) - m_W(y_{n,k}) = A_{1,n} + A_{2,n} + A_{3,n},$$

where $A_{1,n} := \hat m_{\mu(Y)}(Y_{n,k}) - \hat m_{\mu(Y)}(y_{n,k})$, $A_{2,n} := \hat m_{\mu(Y)}(y_{n,k}) - m_W(y_{n,k})$ and $A_{3,n} := \hat m_\xi(Y_{n,k})$, with $\hat m_{\mu(Y)}(Y_{n,k}) = \frac{1}{n}\sum_{i=1}^n \mu(Y_i)1_{\{Y_i\ge Y_{n,k}\}}$.

We start by computing the deterministic tail-moments. Since $s \in \mathcal{RV}_\lambda(+\infty)$ with $\lambda\gamma < 1$, which follows from $2\lambda\gamma < 1$, the integrand $sf \in \mathcal{RV}_{\lambda-1/\gamma-1}(+\infty)$ has a regular variation index below -1, so that Karamata's theorem together with $yf(y) \sim \gamma^{-1}\bar F(y)$ yield $m_{s(Y)}(y) \sim s(y)\bar F(y)/(1-\lambda\gamma)$. Likewise, $m_{s^2(Y)}(y) \sim$

$s^2(y)\bar{F}(y)/(1-2\lambda\gamma)$ since $2\lambda\gamma<1$. Using $\bar{F}(y_{n,k})\sim k/n$, it follows that

$$\frac{m_{s^2(Y)}(y_{n,k})}{n\,m^2_{s(Y)}(y_{n,k})}\sim\frac{(1-\lambda\gamma)^2}{1-2\lambda\gamma}\cdot\frac{1}{k}. \tag{53}$$

Besides, the triangle inequality for the Bochner integral gives $\|m_W(y)\|\le m_{s(Y)}(y)$ for every y, hence the announced bound at $y=Y_{n,k}$.

The focus now turns to the term $A_{2,n}$, the fluctuation at the deterministic threshold. The variables $Z_i:=\mu(Y_i)1_{\{Y_i\ge y_{n,k}\}}$ are i.i.d. in H with $\mathbb{E}_{\mathcal{B}}(Z_1)=m_W(y_{n,k})$ and $\|Z_1\|=s(Y_1)1_{\{Y_1\ge y_{n,k}\}}$, so that $A_{2,n}=\frac{1}{n}\sum_{i=1}^n(Z_i-\mathbb{E}_{\mathcal{B}}(Z_i))$. By the orthogonality of centred i.i.d. summands in H, which follows from Lemma 1, and by (53),

$$\mathbb{E}(\|A_{2,n}\|^2)=\frac{1}{n}\,\mathbb{E}(\|Z_1-\mathbb{E}_{\mathcal{B}}(Z_1)\|^2)\le\frac{1}{n}\,\mathbb{E}(\|Z_1\|^2)=\frac{1}{n}m_{s^2(Y)}(y_{n,k}).$$

In regime (i), Markov's inequality and the deterministic denominator give

$$\frac{\|A_{2,n}\|}{m_{s(Y)}(y_{n,k})}=O_{\mathbb{P}}(k^{-1/2}). \tag{54}$$

In regime (ii), we turn this into an almost-sure statement. The main tool is de Acosta's inequality (Acosta, 1981, Theorem 2.1, case $p>2$). Since $2\lambda\gamma<1$, pick $p\in(2,1/(\lambda\gamma))$; then $s^pf\in\mathcal{RV}_{p\lambda-1/\gamma-1}(+\infty)$ has index below -1, so that $m_{s^p(Y)}(y)\sim C_p\,s^p(y)\bar{F}(y)$ is finite for n large enough. The summands $Z_i-\mathbb{E}_{\mathcal{B}}(Z_i)$ are independent in H with

$$\mathbb{E}(\|Z_i-\mathbb{E}_{\mathcal{B}}(Z_i)\|^p)\le 2^p\,\mathbb{E}(\|Z_1\|^p)=2^p m_{s^p(Y)}(y_{n,k}),$$

so de Acosta's inequality applied to $nA_{2,n}=\sum_{i=1}^n(Z_i-\mathbb{E}_{\mathcal{B}}(Z_i))$ gives

$$\mathbb{E}\left(\big|\,\|nA_{2,n}\|-\mathbb{E}(\|nA_{2,n}\|)\,\big|^p\right)\le C_p\Big[\big(n\,m_{s^2(Y)}(y_{n,k})\big)^{p/2}+n\,m_{s^p(Y)}(y_{n,k})\Big],$$

the first bracket dominating since $p/2\ge1$. Dividing by n^p, it follows that

$$\mathbb{E}\left(\big|\,\|A_{2,n}\|-\mathbb{E}(\|A_{2,n}\|)\,\big|^p\right)\le C_p\,m^p_{s(Y)}(y_{n,k})\,k^{-p/2}(1+o(1)).$$

Besides,

$$\mathbb{E}(\|A_{2,n}\|)\le\mathbb{E}(\|A_{2,n}\|^2)^{1/2}=O(k^{-1/2})\,m_{s(Y)}(y_{n,k}),$$

in virtue of (53), so that, for any $\eta'>0$ and n large enough, $\{\|A_{2,n}\|>\eta' m_{s(Y)}(y_{n,k})\}\subseteq\{\,|\,\|A_{2,n}\|-\mathbb{E}(\|A_{2,n}\|)\,|\,>(\eta'/2)\,m_{s(Y)}(y_{n,k})\,\}$. Markov's inequality then gives

$$\mathbb{P}(\|A_{2,n}\|/m_{s(Y)}(y_{n,k})>\eta')\le C_p\,(2/\eta')^p\,k_n^{-p/2}(1+o(1)).$$

Picking p close enough to $1/(\lambda\gamma)$ that $2/p<\theta$, this is summable since $k_n\ge n^\theta$, whence Borel–Cantelli Lemma yields $\|A_{2,n}\|/m_{s(Y)}(y_{n,k})\to0$ almost surely.

The next step is to bound the term $A_{1,n}$. The indicators $1_{\{Y_i\ge Y_{n,k}\}}$ and $1_{\{Y_i\ge y_{n,k}\}}$ are nested, so that their difference has the same sign for every i. Since $\|\mu(Y_i)\|=s(Y_i)\ge0$, the triangle inequality in H gives

$$\|A_{1,n}\|\le\frac{1}{n}\sum_{i=1}^n s(Y_i)\big|1_{\{Y_i\ge Y_{n,k}\}}-1_{\{Y_i\ge y_{n,k}\}}\big|=\big|\hat{m}_{s(Y)}(Y_{n,k})-\hat{m}_{s(Y)}(y_{n,k})\big|.$$

In regime (i), s is continuously differentiable near infinity with $s'\in\mathcal{RV}_{\lambda-1}(+\infty)$, and $2\lambda\gamma<1$ is exactly the index condition of Stupfler (2019, Theorem 2)

with $a = \lambda$, as in the proof of Proposition 3. Thus, it yields $\hat{m}_{s(Y)}(Y_{n,k}) = m_{s(Y)}(y_{n,k})(1+O_{\mathbb{P}}(k^{-1/2}))$. Besides, the scalar version of (54), with s in place of μ, gives $\hat{m}_{s(Y)}(y_{n,k}) = m_{s(Y)}(y_{n,k})(1 + O_{\mathbb{P}}(k^{-1/2}))$. Subtracting, it follows that $\|A_{1,n}\|/m_{s(Y)}(y_{n,k}) = O_{\mathbb{P}}(k^{-1/2})$.

In regime (ii), we localise the random threshold. Fix $\vartheta \in (0,1)$ and set $y^{\pm}_{n,k} := (1 \pm \vartheta)y_{n,k}$. For $y > 0$, let $N_n(y) := \#\{i : Y_i \geq y\}$, which is binomial with parameters n and $\bar{F}(y)$. Since $\bar{F} \in \mathcal{RV}_{-1/\gamma}(+\infty)$ and $\bar{F}(y_{n,k}) \sim k/n$, one has $\mathbb{E}(N_n(y^{\pm}_{n,k})) = n\bar{F}(y^{\pm}_{n,k}) = \mu^{\pm}_n$, with $\mu^{\pm}_n \sim \rho_{\pm}k$ and $\rho_{\pm} := (1 \pm \vartheta)^{-1/\gamma}$, so that $\rho_+ < 1 < \rho_-$. Set $\delta_+ := \rho_+^{-1} - 1 > 0$ and $\delta_- := 1 - \rho_-^{-1} \in (0,1)$. Since $\mu^+_n \sim \rho_+ k$ with $\rho_+ < 1$, one has $k \geq (1 + \delta_+/2)\mu^+_n$ for n large enough; likewise $\mu^-_n \sim \rho_- k$ with $\rho_- > 1$ gives $k \leq (1 - \delta_-/2)\mu^-_n$ for n large enough. Together with the standard identity $\{Y_{n,k} \geq y\} = \{N_n(y) \geq k\}$, the Chernoff bound for the binomial distribution gives, for n large enough,

$$\mathbb{P}(Y_{n,k} \geq y^+_{n,k}) = \mathbb{P}\big(N_n(y^+_{n,k}) \geq k\big) \leq \mathbb{P}\big(N_n(y^+_{n,k}) \geq (1 + \tfrac{\delta_+}{2})\mu^+_n\big) \leq \exp(-c_+\mu^+_n),$$

$$\mathbb{P}(Y_{n,k} < y^-_{n,k}) = \mathbb{P}\big(N_n(y^-_{n,k}) < k\big) \leq \mathbb{P}\big(N_n(y^-_{n,k}) \leq (1 - \tfrac{\delta_-}{2})\mu^-_n\big) \leq \exp(-c_-\mu^-_n),$$

where $c_+, c_- > 0$ depend only on ϑ and γ. Since $\mu^{\pm}_n \sim \rho_{\pm}k$, both bounds are of the form $\exp(-c'_{\pm}k)$ with $c'_{\pm} > 0$, and as $k_n \geq n^{\theta}$ the series $\sum_n \exp(-c'_{\pm}k_n)$ converge, whence Borel-Cantelli Lemma yields $y^-_{n,k} \leq Y_{n,k} < y^+_{n,k}$ for n large enough almost surely. On this event, $s \geq 0$ and the monotonicity of $y \mapsto \hat{m}_{s(Y)}(y)$ give

$$\big|\hat{m}_{s(Y)}(Y_{n,k}) - \hat{m}_{s(Y)}(y_{n,k})\big| \leq \hat{m}_{s(Y)}(y^-_{n,k}) - \hat{m}_{s(Y)}(y^+_{n,k}).$$

Now $y^{\pm}_{n,k} \to +\infty$ are deterministic with $\bar{F}(y^{\pm}_{n,k}) \sim \rho_{\pm}k/n$, so the scalar version of the regime (ii) bound for $A_{2,n}$, with s in place of μ, applies at $y^-_{n,k}$ and $y^+_{n,k}$ and yields $\hat{m}_{s(Y)}(y^{\pm}_{n,k}) = m_{s(Y)}(y^{\pm}_{n,k})(1 + o(1))$ almost surely. Besides, $m_{s(Y)} \in \mathcal{RV}_{\lambda-1/\gamma}(+\infty)$ together with $y^{\pm}_{n,k} = (1 \pm \vartheta)y_{n,k}$ gives $m_{s(Y)}(y^{\pm}_{n,k})/m_{s(Y)}(y_{n,k}) \to (1 \pm \vartheta)^{\lambda-1/\gamma}$. It follows that

$$\limsup_{n\to+\infty} \frac{\|A_{1,n}\|}{m_{s(Y)}(y_{n,k})} \leq (1-\vartheta)^{\lambda-1/\gamma} - (1+\vartheta)^{\lambda-1/\gamma} \quad \text{almost surely.}$$

The bound holds for every $\vartheta \in (0,1)$, and its right-hand side is positive, since $\lambda - 1/\gamma < 0$, and tends to 0 as $\vartheta \downarrow 0$; letting $\vartheta \downarrow 0$ along a sequence, we conclude that $\|A_{1,n}\|/m_{s(Y)}(y_{n,k}) \to 0$ almost surely.

It remains to bound the residual $A_{3,n}$. As in the proof of Lemma 5, let $\mathbb{E}_{k,y}(\cdot) := \mathbb{E}(\cdot \mid Y_{n,k} = y)$ and $m_{k,y}(W_i) := \mathbb{E}_{k,y}(W_i 1_{\{Y_i \geq Y_{n,k}\}})$. The conditional centring $\mathbb{E}_{\mathcal{F},\mathcal{B}}(\xi) = 0$ reads $\mathbb{E}_{\mathcal{B}}(\xi_1 \mid Y_1 = t) = 0$ for almost every t, while the moment conditions of Lemma 4 hold in view of (16) and Jensen's inequality, so that Remark 10, a consequence of Lemma 2 and Lemma 4, applies to $Z_i = \xi_i$ and yields $m_{k,y}(\xi_1) = 0$ together with

$$\mathbb{E}_{k,y}(\|A_{3,n}\|^2) = \frac{1}{n}\, m_{k,y}(\|\xi_1\|^2).$$

Contrary to the proof of Lemma 5, where the noise ε in (6) is not assumed conditionally centred and the tail-moment $m_{k,y}(Z_1)$ does not vanish, only the diagonal term remains here. By Hölder's inequality with exponents $(q/2, q/(q-2))$, (16) and $m_{k,y}(1) = k/n$ from Lemma 3, there is a constant $c > 0$ independent

of (n, y) such that, for n large enough and uniformly in $y \geq 0$, $m_{k,y}(\|\xi\|^2) \leq c(k/n)^{1-2/q}$. In regime (i), the tower property and $m^2_{s(Y)}(y_{n,k}) \sim cs^2(y_{n,k})(k/n)^2$ then give, with $\delta^{\mu}_{n,k} := (s(y_{n,k})(k/n)^{1/q})^{-1}$, $\|A_{3,n}\|/m_{s(Y)}(y_{n,k}) = O_{\mathbb{P}}(\delta^{\mu}_{n,k}/\sqrt{k})$. Besides, (17) ensures that $1/\delta^{\mu}_{n,k}$ is regularly varying of positive index $\lambda\gamma - 1/q$ in n/k, hence $\delta^{\mu}_{n,k} \to 0$, so this term is $o_{\mathbb{P}}(k^{-1/2})$. In regime (ii), we argue conditionally and integrate over the threshold, as in the proof of Lemma 5. Conditionally on $\{Y_{n,k} = y\}$ and on the pair (i_0, S) of Lemma 2, $nA_{3,n}$ is a sum of k independent centred residuals valued in H, whose conditional moments are bounded by $c_r := \sup_t \mathbb{E}(\|\xi\|^r \mid Y = t) < +\infty$ for $r \in \{2, q\}$, in view of (16) and Jensen's inequality. The same argument as for $A_{2,n}$ then applies conditionally on (i_0, S): writing $\mathbb{E}_{i_0,S,y}(\cdot) := \mathbb{E}(\cdot \mid (i_0, S), Y_{n,k} = y)$ for brevity, de Acosta's inequality (Acosta, 1981, Theorem 2.1, case $p > 2$) with exponent q gives

$$\mathbb{E}_{i_0,S,y}\left(\left|\, \|nA_{3,n}\| - \mathbb{E}_{i_0,S,y}(\|nA_{3,n}\|) \,\right|^q\right) \leq C_q\big[(k\, c_2)^{q/2} + k\, c_q\big] \leq C\, k^{q/2},$$

while the recentring is controlled by $\mathbb{E}_{i_0,S,y}(\|nA_{3,n}\|) \leq (k\, c_2)^{1/2}$. Since $n\, m_{s(Y)}(y_{n,k}) \sim s(y_{n,k})\, k/(1-\lambda\gamma)$ with $s(y_{n,k}) \to +\infty$, the recentring is negligible with respect to the threshold $\eta'\, n\, m_{s(Y)}(y_{n,k})$ and Markov's inequality provides a constant $c > 0$ independent of (n, y) and of (i_0, S) such that, uniformly in $y \in [y^-_{n,k}, y^+_{n,k}]$ and for n large enough,

$$\mathbb{P}_{i_0,S,y}\big(\|A_{3,n}\|/m_{s(Y)}(y_{n,k}) > \eta'\big) \leq c\, \eta'^{-q}\, k_n^{-q/2}\, s^{-q}(y_{n,k}) \leq c\, \eta'^{-q}\, k_n^{-q/2}$$

for any $\eta' > 0$ and n large enough. Averaging over (i_0, S) yields the same bound for $\mathbb{P}_{k,y}(\|A_{3,n}\|/m_{s(Y)}(y_{n,k}) > \eta')$. Integrating this conditional bound over y by Bayes' formula and bounding the contribution of $\{Y_{n,k} \notin [y^-_{n,k}, y^+_{n,k}]\}$ by the Chernoff estimate established for $A_{1,n}$, one obtains $\mathbb{P}(\|A_{3,n}\|/m_{s(Y)}(y_{n,k}) > \eta') \leq c\, \eta'^{-q}\, k_n^{-q/2} + \exp(-c'k)$ for n large enough. Since $\theta > 2\lambda\gamma$ and $q\lambda\gamma > 1$ from (17) entail $\theta q/2 > q\lambda\gamma > 1$: this is summable, whence Borel–Cantelli Lemma yields $\|A_{3,n}\|/m_{s(Y)}(y_{n,k}) \to 0$ almost surely.

We are now able to conclude. Dividing the splitting by $m_{s(Y)}(y_{n,k})$, the three terms are $O_{\mathbb{P}}(k^{-1/2})$ in regime (i) and converge to 0 almost surely in regime (ii). It remains to replace $y_{n,k}$ by $Y_{n,k}$ inside m_W and $m_{s(Y)}$. To this end, in regime (i), (8) ensures $\sqrt{k}\,(Y_{n,k}/y_{n,k} - 1) = O_{\mathbb{P}}(1)$, see de Haan and Ferreira (2006, Theorem 2.2.1); since $m_{s(Y)}$ is continuously differentiable with $-y\, m'_{s(Y)}(y)/m_{s(Y)}(y) = y\, s(y) f(y)/m_{s(Y)}(y) \to (1-\lambda\gamma)/\gamma$, the mean value theorem combined with the uniform convergence theorem for regularly varying functions yields

$$m_{s(Y)}(Y_{n,k})/m_{s(Y)}(y_{n,k}) = 1 + O_{\mathbb{P}}(k^{-1/2}).$$

In regime (ii), $Y_{n,k}/y_{n,k} \to 1$ almost surely, as established for $A_{1,n}$, and $m_{s(Y)} \in \mathcal{RV}_{\lambda-1/\gamma}(+\infty)$ gives $m_{s(Y)}(Y_{n,k})/m_{s(Y)}(y_{n,k}) \to 1$ almost surely. In both regimes, the Bochner triangle inequality $\|m_W(Y_{n,k}) - m_W(y_{n,k})\| \leq |m_{s(Y)}(Y_{n,k}) - m_{s(Y)}(y_{n,k})|$, as for $A_{1,n}$, then lets one replace the deterministic threshold by the empirical one without affecting the rate. This proves the first claim in each regime.

Finally, assume (18), that is $\|m_W(y_{n,k})\| \geq c_0\, m_{s(Y)}(y_{n,k})$ for some $c_0 > 0$ and n large enough, so that $m_W(y_{n,k}) \neq 0$ and $\bar{a}(y_{n,k})$ is well defined. Denote $v_n := \hat{m}_W(Y_{n,k})/m_{s(Y)}(y_{n,k})$ and $\bar{v}_n := m_W(y_{n,k})/m_{s(Y)}(y_{n,k})$, so that

$\|\bar{v}_n\| \in [c_0, 1]$, $\bar{v}_n/\|\bar{v}_n\| = \bar{a}(y_{n,k})$ and $v_n - \bar{v}_n \to 0$ in the mode at hand by the first claim. The reverse triangle inequality provides $\big|\|v_n\| - \|\bar{v}_n\|\big| \to 0$, that is $\|\hat{m}_W(Y_{n,k})\|/\|m_W(y_{n,k})\| \to 1$. Combined with $\|m_W(Y_{n,k})\|/\|m_W(y_{n,k})\| \to 1$, which follows from Bochner triangle inequality and the threshold swap above, this implies the norm statement. Finally, $\|u/\|u\| - w/\|w\|\| \leq 2\|u - w\|/\|u\|$ for nonzero u, w, applied on the event $\{\|v_n\| \geq c_0/2\}$, of probability tending to one in regime (i) and holding for n large enough almost surely in regime (ii), with $u = v_n$ and $w = \bar{v}_n$, gives $\hat{m}_W(Y_{n,k})/\|\hat{m}_W(Y_{n,k})\| - \bar{a}(y_{n,k}) \to 0$ in the mode at hand. Besides, the threshold swap above gives $\|m_W(Y_{n,k}) - m_W(y_{n,k})\| \leq |m_{s(Y)}(Y_{n,k}) - m_{s(Y)}(y_{n,k})| = o\big(m_{s(Y)}(y_{n,k})\big)$ in the mode at hand, while $\|m_W(y_{n,k})\| \geq c_0\, m_{s(Y)}(y_{n,k})$: whence $m_W(Y_{n,k}) \neq 0$ for n large enough, in the mode at hand, and the same triangle bound, now applied with $u = m_W(Y_{n,k})$ and $w = m_W(y_{n,k})$, gives $\bar{a}(Y_{n,k}) - \bar{a}(y_{n,k}) \to 0$. The direction statement at the empirical threshold follows, which ends the proof. □

References

Acosta, A. D. (1981). Inequalities for B-Valued Random Vectors with Applications to the Strong Law of Large Numbers. *The Annals of Probability 9*(1), 157 – 161.

Aghbalou, A., F. Portier, A. Sabourin, and C. Zhou (2024). Tail inverse regression: Dimension reduction for prediction of extremes. *Bernoulli 30*(1), 503–533.

Arbel, J., S. Girard, and H. Lorenzo (2024). Shrinkage for Extreme Partial Least Squares. *Statistics and Computing 34*, 181.

Beirlant, J., Y. Goegebeur, J. Segers, and J. L. Teugels (2004). *Statistics of extremes: theory and applications*, Volume 558. John Wiley & Sons, New-York.

Bingham, N., C. Goldie, and J. Teugels (1987). *Regular Variation*, Volume 27 of *Encyclopedia of Mathematics and its application*. Cambridge University Press.

Bingham, N., C. Goldie, and J. Teugels (1989). *Regular Variation*. Number n° 1 in Encyclopedia of Mathematics and its Applications. Cambridge University Press.

Bousebata, M., G. Enjolras, and S. Girard (2023). Extreme Partial Least-Squares. *Journal of Multivariate Analysis 194*, 105101.

Cai, J.-J., J. H. J. Einmahl, L. de Haan, and C. Zhou (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 77*(2), 417–442.

Cardot, H., F. Ferraty, and P. Sarda (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 571–591.

Cook, R. D. (2007). Fisher lecture: Dimension reduction in regression. *Statistical Science 22*(1), 1–26.

Crambes, C., A. Kneip, and P. Sarda (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics 37*(1), 35 – 72.

Dauxois, J., L. Ferré, and A.-F. Yao (2001). Un modèle semi-paramétrique pour variables aléatoires hilbertiennes. *Comptes Rendus de l'Académie des Sciences - Series I - Mathématiques 333*(10), 947–952.

de Haan, L. and A. Ferreira (2006). *Extreme value theory: an introduction*. Springer Science & Business Media, New-York.

Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics)*. Berlin, Heidelberg: Springer-Verlag.

Gardes, L. (2018). Tail dimension reduction for extreme quantile estimation. *Extremes 21*(1), 57–95.

Gardes, L. (2020). Nonparametric confidence intervals for conditional quantiles with large-dimensional covariates. *Electronic Journal of Statistics 14*(1), 661–701.

Gardes, L. and S. Girard (2012). Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics 6*, 1715–1744.

Girard, S., G. Stupfler, and A. Usseglio-Carleve (2022). Functional estimation of extreme conditional expectiles. *Econometrics and Statistics 21*, 131–158.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*(5), 1163–1174.

Hytonen, T., J. van Neerven, M. Veraar, and L. Weis (2016). *Analysis in Banach Spaces. Volume I: Martingales and Littlewood-Paley Theory*, Volume 63 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics*. Cham: Springer International Publishing.

Kallenberg, O. (2002). *Foundations of Modern Probability* (2nd ed.). Probability and its Applications. New York: Springer-Verlag.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc. 86*(414), 316–327.

Martens, H. and T. Naes (1992). *Multivariate Calibration*. New York: John Wiley & Sons.

Portier, F. (2016). An empirical process view of inverse regression. *Scandinavian Journal of Statistics 43*(3), 827–844.

Preda, C. and G. Saporta (2005). PLS regression on a stochastic process. *Computational Statistics & Data Analysis 48*(1), 149–158.

Stupfler, G. (2019). On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. *Extremes 22*, 749–769.

Wold, H. (1975). Soft modelling by latent variables: The non-linear iterative partial least squares (nipals) approach. *Journal of Applied Probability 12*(S1), 117–142.

Xu, W., D. Li, and H. Wang (2020). Extreme quantile estimation based on the tail single-index model. *Statistica Sinica 32*, 1–22.