Functional Extreme-PLS
Stéphane Girard, Cambyse Pakzad

To cite this version:
Stéphane Girard, Cambyse Pakzad. Functional Extreme-PLS. 2024. hal-04488561

HAL Id: hal-04488561
https://inria.hal.science/hal-04488561
Preprint submitted on 4 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
FUNCTIONAL EXTREME-PLS

STÉPHANE GIRARD AND CAMBYSE PAKZAD

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

Abstract. We propose an extreme dimension reduction method extending the Extreme-PLS approach to the case where the covariate lies in a possibly infinite-dimensional Hilbert space. The ideas are partly borrowed from both Partial Least-Squares and Sliced Inverse Regression techniques. As such, the method relies on the projection of the covariate onto a subspace and maximizes the covariance between its projection and the response conditionally to an extreme event to capture most of the tail information. Moreover, we link the covariate and the heavy-tailed response 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. Finally, we illustrate our results on a finite-sample study with synthetic functional data.

1. Introduction

Over the last decades, the amount of available data has massively increased, urging the statistical community to shift focus toward the high-dimensional regime where the dimension of the observations is large compared to the sample size, or even infinite for the functional case. This situation turns out to be challenging for many reasons, see the monographs Bosq, 2000; Ramsay and Silverman, 2005; Ferraty and Vieu, 2006. First, from the computational point of view, the time and energy costs may exceed the feasible limits. Second, a deeper phenomena called the curse of dimensionality occurs, whose one facet is sparsity among the observations. A prejudice may be overfitting in the regression context where standard estimation methods exhibit high variance and hence may be untrustworthy.

Regression analysis aims at extracting a relationship between a real response variable $Y$ and a high-dimensional predictor/covariate $X$, based on their observations. Doing so, it has inherent connections with sufficient dimension reduction (SDR), a generalization of the single-index model (see e.g., Cook, 1998), whose generic purpose is to alleviate the curse of dimensionality in a regression setting. In essence, the idea of dimension reduction is to substitute $X$ by its orthogonal projection $P_S X$ onto a low-dimensional subspace $S$ that carries the same amount of information on $Y|X$. Specifically, one seeks a subspace $S$ such that $Y$ conditionally to $X$ has same distribution as $Y$ conditionally to $P_S X$, or equivalently $Y$ and $X$ are independent conditionally to $P_S X$ (Cook, 2007). As an example, a powerful, flexible and abundantly studied tool is assuming a single-index model which itself extends standard linear models. Such a setting prescribes $\dim(S) = 1$ and a relationship between $X$ and $Y$ through an unknown link function in such a way that one may explain the values of $Y$...
using a linear combination of the $X$ coordinates. The literature is rich and we only mention Horowitz (2009); Chen et al. (2011) on this topic.

Among the popular methods for dimension reduction, the Partial Least Squares (PLS) method, initiated by Wold (1975), 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 $X$ having high variance and correlation with $Y$ (namely high covariance) meaning that the information in the covariate $X$ that impacts the response $Y$ is encompassed in $P_S X$ while being the most exhaustive.

The literature on high-dimensional PLS regression is scarce. A breakthrough is Chun and Keleş (2010) which show that, under some model, the PLS estimator in linear regression is not consistent unless the dimension of $X$ grows to infinity much slower than the sample size. Interestingly and in the opposite direction, Cook and Forzani (2018, 2019) retrieve consistency with standard rates in the multi-index model and without any regime on the dimension. When the predictor/covariate $X$ is functional, i.e., $X$ lies in an infinite-dimensional space, Preda and Saporta (2005) introduce the first PLS regression method adapted to the particular $L^2([0,1])$-case. Most of developments focus on estimating the slope in functional linear models and we refer to Reiss and Ogden (2007); Delaigle and Hall (2012); Aguilera et al. (2016).

Another standard line of research is Sliced Inverse Regression (SIR), first introduced by Li (1991) in the context of SDR, which targets the central subspace $S_{Y|X}$ defined as the minimal subspace $S$ such that $Y$ and $X$ are independent conditionally to $P_S X$. It hinges on the inverse 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 contradicts the philosophy of Fisher, as noted and encouraged by Cook (2007), where the inference is rather conditioned on the predictors/covariates $X$. The literature on SIR is vast and has seen many developments; we refer to Li et al. (2007) for an alternative method, Lin et al. (2018, 2021) for the fine properties (optimality and minimax rates) of SIR in high-dimension, Tan et al. (2020) in the sparse case and more recently Huang et al. (2023) when $\dim(S_{Y|X})$ may not be bounded.

Concerning the functional framework (FSIR), the seminal papers introducing the topic are Dauxois et al. (2001); Ferré and Yao (2003) which prove the consistency (without rate) provided that the true dimension of the central subspace is known. In a complementary way, Li and Hsing (2010) propose a test on the value of the reduced dimension. The question of consistency is complemented by Forzani and Cook (2007); Cook et al. (2010). Herein, it is proven that the central subspace may only be consistently estimated with rate $\sqrt{n}$ by FSIR under strict restrictions on the covariance operator of $X$. Next, Lian (2015) proves the consistency of FSIR and show that the convergence rate coincide with the minimax rate for the functional linear regression. This connection is explored by Wang and Lian (2020) in the framework of reproducing kernel Hilbert spaces (RKHS) which has already found success in SDR (see e.g., Fukumizu et al. (2009); Lee and Li (2022)). Let us mention two generalizations of FSIR using RKHS to nonlinear spaces: non-Hilbertian sample paths for $X$ in Hsing and Ren (2009), and nonlinear spaces in Li and Song (2017) (see also Li and Song (2022)). For the more realistic situation where the functional data are longitudinal, meaning they are sparsely observed, we may mention Jiang et al. (2014) or very recently Chatla and Liu (2024). About the latest development of FSIR, we refer to Chen et al. (2023) on optimality and minimax rates, and Tian et al. (2023) for a slice-free version.

Now, one may wonder what happens at 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 the vivid research branch of the association of dimension reduction and extreme-value statistics. In this case, the sparsity is doubled 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)). Thereby, one main aspect in this branch builds up on the existence of a projection $P_S X$ that captures all the information on the extreme values of $Y$. Instead of using functional estimation methods, one would then expect that replacing $X$ by $P_S X$ as the covariate, and hence opting for a lower-dimensional method, yields comparable or event better performances. Generally speaking, one may distinguish two classes of method, the unsupervised case where the variables at hand are somehow exchangeable in their roles, and the supervised case where different statuses exist (such as response or covariate). Concerning the former setting, many developments are available in the literature and we refer to Aghbalou et al. (2024) for the latest bibliography material.

Adversely, the supervised situation, especially from a regression perspective, is mostly untouched. This branch builds up on the existence of a projection of $X$ that captures all the information on the extreme values of $Y$. The pioneer work may be identified as Gardes (2018), later followed up by Gardes (2020), which proposes a new notion of tail conditional independence, the latter property underlying SDR. 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 properties are lacking and the computational cost is too high. In this direction, Xu et al. (2020) develop a three-step estimation procedure for such large conditional quantiles and show the $\sqrt{n}$-consistency under a new tail single-index model when the indexes are unknown; yielding better computational performances. Alternatively and having in sight the prediction of tail events, Aghbalou et al. (2024) introduce a new mathematical framework, including a new definition for the tail conditional independence, for estimating with theoretical guarantees, in the extreme regime and by means of inverse regression, the sufficient dimension reduction subspace, under the standard assumptions for SIR. Intersecting dimension reduction and conditional extreme value theory, a novel method based on the PLS in the extreme regime was recently introduced by Bousebata et al. (2023).

Herein, the authors introduce an inverse single-index model. This means that one expresses the covariate 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. The nature of their assumptions is quite different from the standard literature of dimension reduction as no conditions on the inverse conditional mean $E(X|Y)$, such as linearity, nor noise independence or centering are required.

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 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. As such, we consider covariates that lie in an infinite-dimensional Hilbert space $H$ and the projection is expressed in terms of the associated inner product. Our results equally apply when $H = \mathbb{R}^d$ with $d$ large but finite, thereby positioning Bousebata et al. (2023) as a particular sub-case. Following the philosophy of PLS, we maximize the covariance between the projected predictor/covariate and the response conditionally to 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 to 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 $H$. We 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 allows more freedom to the statistician. Our theoretical guarantees take form of results on the asymptotic consistency with rates under similar conditions as in Bousebata et al. (2023). In particular, the assumptions are expressed in the language of regular variation and we require minimal conditions about the noise, unlike the standard literature on SIR and forward single-index model. Finally, the final-sample properties of the method are studied by numerical simulations on synthetic functional data.

The remaining of the paper is organized as follows. First, we introduce in Section 2 the mathematical framework for the Functional Extreme Partial Least-Squares (FEPLS in short) and, more specifically, the inverse single-index model. The estimators are presented in Section 3 with the associated consistency results. Next, the FEPLS finite-sample properties are illustrated in Section 4. The respective code in Python is publicly available on the website of the authors\(^1\). The mathematical details needed for the derivation of our main results are gathered in Section 5.

### 2. Theoretical framework

Let \((H, \langle \cdot, \cdot \rangle)\) be a separable Hilbert space with its Borel \(\sigma\)-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 (Hytönen et al., 2016, Definition 1.1.3).

We define the FEPLS direction as
\[
  w(y) := \arg \max_{\|w\|=1} \text{Cov} \left( \langle w, X \rangle, Y \mid Y \geq y \right), \quad y \in \mathbb{R},
\]
where \(\|\cdot\|\) is the norm induced by the inner product \(\langle \cdot, \cdot \rangle\) and \(y \in \mathbb{R}^+\) is a large deterministic threshold. The expectation underlying the covariance in (2.1) is understood in the sense of Lebesgue. The interpretation of (2.1) 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 (and not the correlation) conditionally to the tail event of \(Y\), shared between \(Y\) and the projection of \(X\) on a single direction.

In the sequel, we denote 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 \(\mathbb{E}_B\). Note that, since we consider a separable space, we could equivalently use the Pettis integral instead. We refer to Hytönen et al. (2016) for more details about the notion.

A key fact is that the solution of the optimization problem (2.1) is explicit as we have:

**Proposition 2.1.** Suppose that \(Y\) is integrable and \(X\) is Bochner-integrable. Then, the unique solution of the optimization problem (2.1) is given by, for any \(y \in \mathbb{R}\),
\[
  w(y) = v(y)/\|v(y)\| \text{ where } v(y) = F(y)m_{XY}(y) - m_X(y)m_Y(y).
\]
This result allows us to undertake an inference procedure based on the empirical counterparts of tail-moments in (2.2). Before that, let us introduce the characteristics of our model. To make sense of the conditional tail event in (2.1), we consider an extreme framework where the response variable \(Y\) is heavy-tailed. Equivalently, its associated survival function is such that
\[
  y \mapsto F(y) := \mathbb{P}(Y \geq y) \in \mathcal{RV}_{-1/\gamma}(+\infty), \quad \gamma \in (0, 1),
\]
where \(\mathcal{RV}_a(+\infty), \ a \in \mathbb{R}\), is the set of real-valued measurable functions \(f\) defined on a neighbourhood \([x_0, +\infty)\) of infinity and such that, for any \(x > x_0 \geq 0\),
\[
  \lim_{t \to +\infty} \frac{f(tx)}{f(t)} = x^a.
\]

\(^1\)https://cpakzad.github.io/download/FEPLS.zip
In such a case, \( f \) is said to be regularly-varying at infinity with index \( a \) and we refer to Bingham et al. (1989) for further details.

Next, in order to provide theoretical guarantees on the inference method described in the next section and based on (2.1)–(2.2), we consider the following (non-linear) inverse single-index functional model,

\[
X = g(Y)\beta + \varepsilon, \quad \beta \in H, \quad \|\beta\| = 1, \tag{2.4}
\]

with \( \varepsilon : \Omega \to H \) a random variable being the noise and an (unknown) deterministic link function \( g : \mathbb{R} \to \mathbb{R} \) such that \( g \in \mathcal{R}V_\kappa(+\infty) \) with \( \kappa > 0 \).

The appeal of the model (2.4) stems from the fact that, for instance when \( \varepsilon \) is independent of \( X \) and centered, the FEPLS direction (2.1) coincides with the true index: \( w(y) = \beta \) for any \( y \geq 0 \). Heuristically, one expects, assuming more generally that \( Y \) and \( \varepsilon \) are dependent but \( \varepsilon \) has a small contribution in the tail regime of \( Y \), that

\[
\arg \max_{\|w\|=1} \text{Cov} \left( \langle w, X \rangle, Y \mid Y \geq y \right) \to \beta, \quad \text{as } y \to +\infty.
\]

To deal with the conditional expectation involving the tail event \( \{ Y > y \} \) for large \( y \), we need some regularity on the distribution of \( Y \). Hence, let us assume that the density of \( Y \) exists and denote it by \( f := (1-F)' \). Concerning the noise, we require that there exists \( q > 2 \) such that

\[
\mathbb{E}(\|\varepsilon\|^q) < +\infty. \tag{2.5}
\]

In view of the norms injection, the noise \( \varepsilon \) also satisfies \( \mathbb{E}(\|\varepsilon\|) < +\infty \). By (Hytönen et al., 2016, Proposition 1.2.2), since \( \varepsilon \) is strongly \( \mathbb{P} \)-measurable as a random variable, it is also Bochner-integrable.

### 3. Inference and consistency

Let \((X_i, Y_i)_{1 \leq i \leq n}\) be independent copies of \((X, Y)\) with \( X_i \in H \) and \( Y_i \in \mathbb{R} \). Let also \( y_n \to +\infty \) be a deterministic sequence such that

\[
n \tilde{F}(y_n) \xrightarrow[n \to +\infty]{} +\infty. \tag{3.1}
\]

The empirical counterpart of the tail-moment \( m_W(y) \) is defined by

\[
\hat{m}_W(y_n) := \frac{1}{n} \sum_{i=1}^{n} W_i 1\{Y_i \geq y_n\}
\]

basing on the \( n \)-sample. It is then possible to estimate \( w(y_n) \) by \( \hat{w}(y_n) \), obtained by replacing all four tail moments in (2.2) by their empirical counterparts. Note that \( \hat{m}_Y(y_n) \) is sometimes referred to as the empirical Conditional Tail Expectation (see El-Methni et al., 2014). It readily appears from (2.2) that \( w(y_n) \) is a linear combination of \( m_{XY}(y_n) \in H \) and \( m_X(y_n) \in H \). Since both of them point in the same direction in \( H \) which is given by \( X \), it seems reasonable to investigate the asymptotic behaviour of more general empirical tail moments defined as

\[
\hat{v}_\varphi(y_n) = \hat{m}_{\varphi(Y)X}(y_n) = \frac{1}{n} \sum_{i=1}^{n} \varphi(Y_i)X_i 1\{Y_i \geq y_n\}, \quad \hat{\beta}_\varphi(y_n) := \frac{\hat{v}_\varphi(y_n)}{\|\hat{v}_\varphi(y_n)\|}.
\]

where \( \varphi : \mathbb{R} \to \mathbb{R} \) is some test function such that \( \varphi \in \mathcal{R}V_\tau(+\infty) \) with \( \tau \in \mathbb{R} \).

We may now state our consistency result under the inverse single-index model. From now on, we adopt the notation \( a_n \ll b_n \) for two real sequences if and only if \( a_n/b_n \to 0 \) as \( n \to +\infty \) and \( b_n \neq 0 \) eventually.
Proposition 3.1. Suppose (2.3), (2.4) and (2.5) hold. Let \((y_n)\) be a deterministic sequence such that \(y_n \to +\infty\) and (3.1) holds. Assume moreover that
\[
2(\kappa + \tau)\gamma < 1, \tag{3.2}
\]
\[
q\kappa\gamma > 1. \tag{3.3}
\]
If \(\tau > 0\), assume in addition that
\[
q(1 - 2\tau\gamma) > 2. \tag{3.4}
\]
Let \(\delta_n := (g(y_n)F^{1/q}(y_n))^{-1} \ll 1\). Then,
\[
\|\hat{\beta}_\varphi(y_n) - \beta\| = O_\mathbb{P}(\delta_n).
\]

In the next result, we show that the same amount of the tail information is asymptotically preserved when projecting on \(\text{Span}(\hat{\beta}_\varphi(y_n))\) instead of \(\text{Span}(\beta)\).

Proposition 3.2. Let \((X_0, Y_0)\) with same distribution as \((X, Y)\) and independent of \((X_1, Y_1)\).
Under the conditions of Proposition 3.1, it holds that, as \(n \to +\infty\),
\[
\frac{\text{Cov}(Y_0, \langle \hat{\beta}_\varphi(y_n), X_0 \rangle \mid Y_0 \geq y_n)}{\text{Cov}(Y_0, \langle \beta, X_0 \rangle \mid Y_0 \geq y_n)} - 1 = O(\delta_n),
\]
where it is recalled that \(\delta_n = (g(y_n)F^{1/q}(y_n))^{-1} \ll 1\).

Remark 3.1. We state some comments about the set of hypotheses and their implications.

- The assumptions may be divided into two classes. First, there are the model hypotheses which are constituted by (2.3), (2.4), (2.5) and (3.3). Second, the hypotheses related to the estimation are (3.1), (3.2), and (3.4).
- The regime (3.1) entails that the average number of tail observations \(Y_i > y_n\) increases as the sample size grows. This roughly ensures that the threshold \(y_n\) does not grow too fast so we have enough data points in our inference scheme. In particular, rephrasing (3.1) and using (2.3) in terms of regular variations yield \(y_n \ll n^{\gamma - \eta}\) for any small \(\eta > 0\).
- The set of \((\gamma, \tau, \kappa, q)\) with \(\gamma \in (0, 1), \kappa > 0, 0 < \tau < \frac{1}{\sqrt{2}}\) and \(q > 2\) such that (3.2), (3.3) and (3.4) hold is non-empty.
- The convergence rate \(\delta_n\) only depends on the tail distribution of \(Y\), the integrability order of the noise \(\varepsilon\) and the asymptotic behaviour of the deterministic link function \(g\).

4. Illustration on simulated data

In this section, we assess the validity of our method by a Monte Carlo simulation experiment with \(N = 500\) independent repetitions. The ambient space is \(H = L^2([0, 1])\) endowed with its usual inner product.

Throughout this numerical study, all functions are discretized on a regular grid \(0 = x_1 < \ldots < x_d = 1\) with \(d = 101\) and the \(L^2([0, 1])\)-inner product of two functions \(W_1, W_2\) is approximated by
\[
\langle W_1, W_2 \rangle_d := \frac{1}{d} \sum_{k=1}^{d} W_1(x_k)W_2(x_k).
\]

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., \(\tilde{F}(y) = (1 + y^{-\rho})^{-\frac{1}{\rho}}\) with \(y \geq 0, \) and \(\gamma \in (0, 1)\) being the tail-index and \(\rho < 0\) being the second-order parameter.
- \(\varepsilon \mid Y = y\) 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 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 $B^H_y$ is centered Gaussian with conditional covariance function

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

Note that $H(y) = 1/2$ gives a Brownian motion $B^H_y$ with mean $\mu(y)$ and variance $\sigma^2(y)$.

Once both the response variable $Y$ and the noise $\varepsilon$ are simulated, and the deterministic functions $\beta$ and $g$ are chosen, the covariate sampling $(X_i)_{1 \leq i \leq n}$ is readily derived from (2.4). The choice of $\sigma$ is guided by the following reasonable requirement: One should ensure that the contribution in (2.4) of $\varepsilon$ 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 in our study case $(H, \mu) \equiv (1/3, 200)$ regardless of $Y = y$.

As usual in extreme value theory, we parametrize the diverging threshold by the quantiles of $(Y_i)_{1 \leq i \leq n}$. Hence, we set up $y_n$ of the form $Y_{n-k+1:n}$ where $\min(Y_i) = Y_{1:n} \leq \ldots \leq Y_{n:n} = \max(Y_i)$ are the order statistics of $(Y_i)$. We consider the following data-driven selection procedure on the order $k$, based on the maximization of the correlation between $Y$ and $\langle \hat{\beta}_\phi(y_n), X \rangle$ when $Y$ is large.

For any $1 \leq k \leq n$, we introduce $\text{PCC}(k, n)$ the Pearson correlation coefficient between $(Y_{n-i+1:n})_{1 \leq i \leq k}$ and $(\langle \hat{\beta}_\phi(Y_{n-k+1:n}), X_{\{n-i+1,n\}} \rangle)_{1 \leq i \leq k}$ where, for any $1 \leq i \leq k$, $X_{\{n-i+1,n\}}$ denotes the concomitant of $Y_{n-i+1:n}$, i.e., the value of $X_s$ with $s \in \{1, \ldots, n\}$ being the unique index such that $Y_s = Y_{n-i+1:n}$ (since $Y$ is continuous, there is no tie). The measure of association $\text{PCC}(j, n)$ is explicitly written as

$$\frac{1}{k} \sum_{i=1}^{k} Y_{n-i+1:n} \langle \hat{\beta}_\phi(Y_{n-k+1:n}), X_{\{n-i+1,n\}} \rangle - \frac{1}{k} \sum_{i=1}^{k} Y_{n-i+1:n} \frac{1}{k} \sum_{i=1}^{k} \langle \hat{\beta}_\phi(Y_{n-k+1:n}), X_{\{n-i+1,n\}} \rangle.$$

We select the order $k$ defined as

$$k := \arg \max_{5 \leq j \leq n} \text{PCC}(j, n).$$

Under the described simulation model, we compute the values of $\text{PCC}(k, n)$ for $1 \leq k \leq 100$ and $n = 500$ for different admissible $\tau$. The results are represented in Figure 1. As one may see, the estimator should be the most performing at the highest admissible value of $\tau$ where it always capture the most of the correlation in the extremes between $Y$ and $\langle \hat{\beta}_\phi, X \rangle$.

Using the selection method for $k$, we may now set the threshold $y_n$ according to the data and infer the single-index parameter $\beta$ thanks to $\hat{\beta}_\phi$. The results are shown in Figure 2 for multiple tail indexes. One may observe that the estimation is more accurate when $\gamma$ or (mostly) $\tau$ increases.
Figure 1. Finite sample behaviour comparison of $\hat{\beta}_p(Y_{n-k-1:n})$ for different admissible $\tau$ under the 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 over $N = 500$ Monte Carlo repetitions of the correlation between $(Y_{n-i+1:n})_{1 \leq i \leq k}$ et $(\langle \hat{\beta}_p(Y_{n-k+1:n}), X_{(n-i+1:n)} \rangle)_{1 \leq i \leq k}$. Here, the tail indexes $\kappa$ and $\gamma$ are varying and $\rho = -1/2$. 

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

[Graphs and Figures]
Figure 2. Simulation results on the inverse model with Burr response and conditional fBm noise. The orange curve is the graph of $\beta$ on $[0, 1]$ while the blue one is the averaged value of $\hat{\beta}_{\phi}(Y_{n-k+1:n})$ over $N = 500$ Monte Carlo repetitions, where $k = \arg \max_{5 \leq j \leq n} \text{PCC}(j, n)$. The light blue area corresponds to the confidence region comprising the top 5–95% values of the Monte Carlo repetitions. Here, the tail indexes $\kappa$ and $\gamma$ are varying while we fix $\rho = -1/2$ and $\tau = -2$. 
5. Technical results

We gather in this section all the mathematical details behind Section 2 and Section 3. We start with Section 5.1 and an useful technical tool which, among other things, enables the proof of Proposition 2.1. We also give a result on the asymptotic equivalent of theoretical tail-moments, which will be extensively used throughout the paper. Next, we derive a central limit theorem on the univariate empirical tail-moments in Section 5.2. The relevant asymptotics of empirical tail-moments involving the noise and its inner product with \( \beta \) are respectively investigated in Section 5.3 and Section 5.4. Finally, we establish our main consistency results in Section 5.5.

5.1. Preliminaries. A standard result in the literature of Bochner spaces, e.g., (Hytönen 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.

Let \( (\mathcal{B}, \langle \cdot, \cdot \rangle) \) be a separable Hilbert space, \( \beta \) a separable Hilbert space, \( \| \cdot \|_\beta \) a norm on \( \beta \), \( (e_n)_{n \geq 1} \) an orthonormal basis of \( \beta \), and \( (\|W_i\|)_{i \in \mathbb{N}} \) a sequence of nonnegative real numbers. Then, for any \( i \in \{1, 2\} \),

\[
\langle \mathbb{E}(W_i), e_n \rangle = \sum_{n \geq 1} \langle \mathbb{E}(W_i), e_n \rangle e_n.
\]

Now, \( w \mapsto \langle w, z \rangle \) is a bounded linear operator for any fixed \( z \in \mathcal{B} \) and hence Hille’s theorem (Hytönen et al., 2016, Theorem 1.2.4) (which is more than what we actually need) yields

\[
\langle \mathbb{E}(W_i), e_n \rangle = \langle \mathbb{E}(W_i), e_n \rangle.
\]

Overall,

\[
\langle \mathbb{E}(W_1), \mathbb{E}(W_2) \rangle = \sum_{n \geq 1} \langle \mathbb{E}(W_1), e_n \rangle \langle \mathbb{E}(W_2), e_n \rangle
\]

holds by continuity of the inner product on \( \mathcal{H} \). Commuting again the Bochner expectation with the inner product and independence, it follows that

\[
\langle \mathbb{E}(W_1), \mathbb{E}(W_2) \rangle = \sum_{n \geq 1} \langle \mathbb{E}(W_1), e_n \rangle \langle \mathbb{E}(W_2), e_n \rangle.
\]

The last step is to exchange the standard Lebesgue expectation and the infinite sum in the RHS of the last equality. This would conclude the proof as it implies that

\[
\langle \mathbb{E}(W_1), \mathbb{E}(W_2) \rangle = \mathbb{E} \left( \sum_{n \geq 1} \langle W_1, e_n \rangle \langle W_2, e_n \rangle \right) = \mathbb{E} \left( \langle W_1, W_2 \rangle \right).
\]
We start by rewriting the conditional covariance. For any $y \in \mathbb{R}$,

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

By (Hytönen et al., 2016, Proposition 1.2.2), we have $\mathbb{E}(\|X\|) < +\infty$. Since $\|w\| = 1$, a straightforward application of Cauchy-Schwarz’s inequality and monotony of expectations imply that the expectations in the last display are finite. Next, by Lemma 5.1, one has

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

The optimization problem now becomes $\arg\max_{\|w\|=1} \langle w, v(y) \rangle$ for which the solution is known to be $v(y)/\|v(y)\|$. Indeed, for any $w \in H$ such that $\|w\| = 1$, Cauchy-Schwarz’s inequality implies that $\langle w, v(y) \rangle \leq \|v(y)\|$, while taking $w = v(y)/\|v(y)\|$ realizes the latter upper-bound.

The next result is a slight extension to (Bousebata et al., 2023, Lemma 2) that we quote here for the sake of self-containedness.

Lemma 5.2. Let any $h \in \mathcal{RV}_\rho(+\infty)$ with $\rho \in \mathbb{R}$ and any real random variable $Y$ with density $f \in \mathcal{RV}_{-1/\gamma-1}(+\infty)$ with $\gamma > 0$. Assume that $\rho \gamma < 1$. Then, as $y \to +\infty$,

$$
m_h(y)(y) := \mathbb{E}(h(Y) 1_{\{Y \geq y\}}) \sim \frac{1}{1 - \rho \gamma} h(y) \bar{F}(y),
$$

where $\bar{F}(y) := 1 - \int_y^\infty f(t)dt \in \mathcal{RV}_{-1/\gamma}(+\infty)$ is the survival distribution function of $Y$.

Proof. By transfer formula, one may write

$$
\mathbb{E}(h(Y) 1_{\{Y \geq y\}}) = \int_y^{+\infty} h(t) f(t) dt.
$$

Since $h \cdot f \in \mathcal{RV}_{\rho - 1 - 1/\gamma}(+\infty)$, we may write $h(x) f(x) = x^{\rho - 1 - 1/\gamma} \ell(x)$ for any $x$ where $\ell$ is slowly-varying. Now, since $\rho \gamma < 1$ ensures that $\rho - 1 - 1/\gamma < -1$, one may apply (Bingham et al., 1989, Proposition 1.5.10) to get

$$
\int_y^{+\infty} h(t) f(t) dt \sim \frac{1}{-\rho + \frac{1}{\gamma}} y^{\rho - 1 - 1/\gamma} \ell(y) = \frac{1}{-\rho + \frac{1}{\gamma}} y h(y) f(y), \quad y \to +\infty.
$$

Since $\gamma > 0$ so that $-1 - 1/\gamma < -1$ and since $f \in \mathcal{RV}_{-1/\gamma-1}(+\infty)$, we may apply once more (Bingham et al., 1989, Proposition 1.5.10) to get $y f(y) \sim \bar{F}(y)/\gamma$ as $y \to +\infty$. The conclusion follows by combining the previous asymptotics and transitivity of equivalents.
5.2. Empirical tail-moment of \( g(Y) \).

**Lemma 5.3.** Let \( \varphi \in \mathcal{RV}_\tau(+\infty) \) with \( \tau \in \mathbb{R} \) and \( g \in \mathcal{RV}_\kappa(+\infty) \) with \( \kappa > 0 \). Let \( Y \) a real random variable with survival function such that (2.3) holds. Let \((y_n)\) be a deterministic sequence such that (3.1) holds. Assume that (3.2) holds for \( \tau \in \mathbb{R} \). Then,

\[
\sqrt{nF(y_n)} \left( \frac{\bar{m}_{\varphi,g}(y_n)}{m_{\varphi,g}(y_n)} - 1 \right) \xrightarrow{d} n \to +\infty N \left( 0, \frac{(1 - (\kappa + \tau)\gamma)^2}{1 - 2(\kappa + \tau)\gamma} \right), \tag{5.1}
\]

**Proof.** Let \( D_{ni} := \frac{\varphi(Y_i)g(Y_i)}{m_{\varphi,g}(y_n)} - 1 \) and \( s_n^2 := \sum_{i=1}^n \text{Var}(D_{ni}) = n \left( \frac{m_{\varphi^2,g}(y_n)}{m_{\varphi,g}(y_n)} - 1 \right) \).

First, observe that Lemma 5.2 yields, under condition (3.2), the following two first-order approximations:

\[
m_{\varphi,g}(y_n) = F(y_n) \mathbb{E}(\varphi(Y)g(Y) \mid Y \geq y_n) \sim \frac{1}{1 - (\kappa + \tau)\gamma} \varphi(y_n)g(y_n)F(y_n),
\]

\[
m_{\varphi^2,g}(y_n) = F(y_n) \mathbb{E}(\varphi^2(Y)g^2(Y) \mid Y \geq y_n) \sim \frac{1}{1 - 2(\kappa + \tau)\gamma} \varphi^2(y_n)g^2(y_n)F(y_n),
\]

as \( n \to +\infty \). In particular, one has

\[
\frac{m_{\varphi^2,g}(y_n)}{m_{\varphi,g}(y_n)} \sim \frac{(1 - (\kappa + \tau)\gamma)^2}{1 - 2(\kappa + \tau)\gamma} \frac{1}{F(y_n)} \xrightarrow{n \to +\infty} +\infty,
\]

and thus,

\[
s_n^2 \sim \frac{(1 - (\kappa + \tau)\gamma)^2}{1 - 2(\kappa + \tau)\gamma} \frac{n}{F(y_n)}, \quad n \to +\infty.
\]

Since \((D_{ni})_{1 \leq i \leq n}\) is a centered independent triangular array, and according to (Billingsley, 1995, Theorem 27.3), it is enough to show that

\[
\exists \delta > 0 \text{ such that } \sum_{i=1}^n \frac{1}{s_n^{2+\delta}} \mathbb{E}(|D_{ni}|^{2+\delta}) \xrightarrow{n \to +\infty} 0. \tag{5.2}
\]

This would provide that \( \sum_{i=1}^n D_{ni}/s_n \xrightarrow{d} n \to +\infty N(0, 1) \) which is equivalent to (5.1). Now, let \( \delta \in (0, \frac{1}{\tau(\kappa + \tau)} - 2) \) which is possible by hypothesis (3.2). We write, for \( i \in \{1, \ldots, n\} \):

\[
D_{ni} = \frac{1}{m_{\varphi,g}(y_n)} \left( \varphi(Y_i)g(Y_i)1_{\{Y_i \geq y_n\}} - m_{\varphi,g}(y_n) \right).
\]

Using the triangle inequality, Jensen inequality, the fact that the \( L^p \)-norms are ordered and that regularly-varying functions are eventually positive,

\[
\mathbb{E}(|D_{ni}|^{2+\delta}) \xrightarrow{1+\delta} m_{\varphi,g}(y_n)^{-1} \mathbb{E} \left( (\varphi(Y_i)g(Y_i))^{2+\delta}1_{\{Y_i \geq y_n\}} \right)^{\frac{1}{1+\delta}} \leq m_{\varphi,g}(y_n)^{-1} \mathbb{E} \left( (\varphi(Y_i)g(Y_i))^{2+\delta}1_{\{Y_i \geq y_n\}} \right)^{\frac{1}{1+\delta}} \leq 2m_{\varphi,g}(y_n)^{-1} \mathbb{E} \left( (\varphi(Y_i)g(Y_i))^{2+\delta}1_{\{Y_i \geq y_n\}} \right)^{\frac{1}{1+\delta}}.
\]

It follows that

\[
0 \leq \sum_{i=1}^n \frac{1}{s_n^{2+\delta}} \mathbb{E}(|D_{ni}|^{2+\delta}) \leq \sum_{i=1}^n \frac{2^{2+\delta}}{s_n^{2+\delta}m_{\varphi,g}(y_n)} \mathbb{E} \left( (\varphi(Y_i)g(Y_i))^{2+\delta}1_{\{Y_i \geq y_n\}} \right).
\]

Using equidistribution, Lemma 5.2 again with the fact that \( (\varphi \cdot g)^{2+\delta} \in \mathcal{RV}_{(2+\delta)(\kappa + \tau)}(+\infty) \), there exists a constant \( c \in (0, +\infty) \) depending only on \( \kappa, \tau, \gamma, \delta \) such that, as \( n \to +\infty \),

\[
\sum_{i=1}^n \frac{1}{s_n^{2+\delta}m_{\varphi,g}(y_n)} \mathbb{E}(\varphi(Y_i)g(Y_i)1_{\{Y_i \geq y_n\}})^{2+\delta} = \frac{n}{s_n^{2+\delta}} \frac{m_{\varphi,g}(y_n)^{2+\delta}(y_n)}{m_{\varphi,g}(y_n)} \sim c \left( nF(y_n) \right)^{-\delta/2},
\]
and the conclusion follows from (3.1).

5.3. **Empirical tail-moments for ε.**

**Lemma 5.4.** Assume that (2.3), (2.5), (3.2) and (3.4) hold. Let \((y_n)\) be a deterministic sequence such that (3.1) holds and let \(\delta_n := (g(y_n)F^{1/q}(y_n))^{-1}\). Then,

\[
\left\| \frac{\hat{m}_\varphi(Y)\varepsilon(y_n)}{m_{\varphi Y}(y_n)} \right\| = O_P(\delta_n), \quad n \to +\infty.
\]

**Remark 5.5.** Observe that \(\delta_n \ll 1\) under assumption (3.3).

**Proof.** Denoting \(Z := \varphi(Y)\varepsilon\) for clarity, we start by computing the expectation of \(\|\hat{m}_Z(y_n)\|^{2}\) which is by construction,

\[
\|\hat{m}_Z(y_n)\|^{2} = \frac{1}{n^2} \sum_{i_1 \neq i_2}^{n} \langle Z_{i_1}, Z_{i_2} \rangle 1\{Y_{i_1} \geq y_n\} 1\{Y_{i_2} \geq y_n\}.
\]

Taking the expectation and splitting the sum, we may write

\[
\mathbb{E}(\|\hat{m}_Z(y_n)\|^{2}) = \frac{1}{n^2} \sum_{i_1 \neq i_2}^{n} \mathbb{E} \left( \langle Z_{i_1} 1\{Y_{i_1} \geq y_n\}, Z_{i_2} 1\{Y_{i_2} \geq y_n\} \rangle \right) + \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}(\|Z_i\|^{2} 1\{Y_i \geq y_n\}).
\]

Lemma 5.1 and equidistribution entail

\[
\mathbb{E}(\|\hat{m}_Z(y_n)\|^{2}) = \|m_Z(y_n)\|^{2} + \frac{1}{n} \left( m_{\|Z\|^2}(y_n) - \|m_Z(y_n)\|^{2} \right).
\]

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

\[
\mathbb{P} \left( \frac{\|\hat{m}_Z(y_n)\|}{m_{\varphi Y}(y_n)} > \eta' \right) \leq \frac{1}{\eta'^2} \mathbb{E}(\|\hat{m}_Z(y_n)\|^{2}) m_{\varphi Y}(y_n). \tag{5.3}
\]

According to Lemma 5.2 and under (3.2), there exists some constant \(c \in (0, +\infty)\) independent of \(n\) such that, for \(n\) large enough, Markov’s inequality (5.3) becomes

\[
\mathbb{P} \left( \frac{\|\hat{m}_Z(y_n)\|}{m_{\varphi Y}(y_n)} > \eta' \right) \leq \frac{c}{\eta'^2} (\varphi \cdot g \cdot \tilde{F})^{-2}(y_n) \left( \|m_Z(y_n)\|^{2} \vee \frac{1}{n} m_{\|Z\|^2}(y_n) \right) .
\]

The next step is to bound the term involving the tail-moments. To this end, observe that we may write by Lemma 5.1,

\[
\|m_Z(y_n)\|^{2} = \langle \mathbb{E}(Z_1 1\{Y_1 \geq y_n\}), \mathbb{E}(Z_2 1\{Y_2 \geq y_n\}) \rangle = \mathbb{E}(\langle Z_1, Z_2 \rangle 1\{Y_1, Y_2 \geq y_n\}).
\]

By independence and Cauchy-Schwarz’s inequality in \(H\) and then by Cauchy-Schwarz’s inequality in \(L^2(\Omega, \mathcal{A}, \mathbb{P})\), we may bound

\[
\|m_Z(y_n)\|^{2} \leq m_{\varphi Y}(y_n) \|\varepsilon\|^{2} \leq m_{\|Z\|^2}(y_n) \tilde{F}(y_n). \tag{5.4}
\]

The regime (3.1) implies that there exists a constant \(c \in (0, +\infty)\) independent of \(n\) such that for \(n\) large enough,

\[
\frac{1}{n} m_{\|Z\|^2}(y_n) \leq c \tilde{F}(y_n) m_{\|Z\|^2}(y_n).
\]

It follows from (5.4) that

\[
\|m_Z(y_n)\|^{2} \vee \frac{1}{n} m_{\|Z\|^2}(y_n) \leq m_{\|Z\|^2}(y_n) \tilde{F}(y_n). \tag{5.5}
\]

The focus now turns to the tail-moment of \(\|Z\|^{2}\) in the previous inequality. The \(q\)-moment of \(\|\varepsilon\|\) is finite under (2.5) so that \(c\’s\) inequality allows us to write,

\[
m_{\|Z\|^2}(y_n) = \mathbb{E} \left( \varphi^2(Y) \|\varepsilon\|^{2} 1\{Y \geq y_n\} \right) \leq \mathbb{E} (\|\varepsilon\|^{q/2} m_{\varphi Y}(y_n)^{2/q})^{2/q} (\varphi(Y))^{2/q} (y_n).
\]
Note that the tail moment \(m_{\varphi(Y)^{2q}}(y_n)\) is finite if \(\frac{2q}{q-2} < \frac{1}{\gamma}\) 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 (3.4). Moreover, \(\varphi^{2q} \in \mathcal{R}V_{\frac{2q}{q-2}}(0, \infty)\) so that Lemma 5.2 and (3.4) yield

\[
m_{\varphi(Y)^{2q}}(y_n) \sim \frac{1}{(1-\frac{2\varphi^2}{q-2})^{1-2/q}} \varphi^2(y_n) F^{1-2/q}(y_n).
\]

It follows that there exists a constant \(c \in (0, +\infty)\) independent of \(n\) such that for \(n\) large enough,

\[
m_{\|Z\|^2}(y_n) \leq \mathbb{E}(\|\varepsilon\|^{2q} m_{\varphi(Y)^{2q}}(y_n)) \leq c \varphi^2(y_n) F^{1-2/q}(y_n).
\] (5.6)

Combining (5.5) and (5.6), we have for some constant \(c \in (0, +\infty)\) independent of \(n\),

\[
\|m_Z(y_n)\|^2 \vee \frac{1}{n} m_{\|Z\|^2}(y_n) \leq c \varphi^2(y_n) F^{2(1-1/q)}(y_n).
\] (5.7)

Going back to Markov’s bound (5.3), for some constant \(c \in (0, +\infty)\) independent of \(n\),

\[
\mathbb{P}\left(\frac{\|\hat{m}_Z(y_n)\|}{m_{\varphi(Y)\varphi(Y)}(y_n)} > \eta'\right) \leq \frac{c}{\eta'^2} g^{-2}(y_n) F^{-2/q}(y_n) = \frac{c\delta_n^2}{\eta'^2}.
\]

Now, let any \(\eta > 0\). Since \(\eta'\) is arbitrary, we may pick \(\eta'\) of the form \(\delta_n (c/\eta)^{-1/2}\) so that it concludes the proof as we have shown that

\[
\mathbb{P}\left(\delta_n^{-1} \left|\frac{\hat{m}_{\beta, \varphi(Y)\varphi(Y)}(y_n)}{m_{\varphi(Y)\varphi(Y)}(y_n)}\right| > \eta^{-1/2}\right) < \eta,
\]

for all \(\eta > 0\).

\[\square\]

5.4. Empirical tail-moment of the inner product.

**Lemma 5.6.** Let any \(\beta \in H\) with \(\|\beta\| = 1\). Assume that (2.3), (2.5), (3.2) and (3.4) hold. Let \((y_n)\) be a deterministic sequence such that (3.1) holds and let \(\delta_n := (g(y_n) F^{1/q}(y_n))^{-1}\). Then,

\[
\frac{\hat{m}_{\beta, \varphi(Y)\varphi(Y)}(y_n)}{m_{\varphi(Y)\varphi(Y)}(y_n)} = O_p(\delta_n), \quad n \to +\infty.
\]

**Proof.** We start with a preliminary computation of the expectation. One has

\[
\hat{m}_{\beta, \varphi(Y)\varphi(Y)}(y_n) = \frac{1}{n} \sum_{i=1}^n \langle \beta, \varphi(Y_i) \varepsilon_i \rangle 1_{\{Y_i \geq y_n\}}.
\]

Since \(\langle \beta, \cdot \rangle : H \to \mathbb{R}\) is a continuous linear mapping,

\[
\mathbb{E}(\hat{m}_{\beta, \varphi(Y)\varphi(Y)}(y_n)) = \frac{1}{n} \sum_{i=1}^n \langle \beta, m_{\varphi(Y)\varphi(Y)}(y_n) \rangle = \langle \beta, m_{\varphi(Y)\varphi(Y)}(y_n) \rangle.
\]

We may now invoke the main tool for the proof, namely the Chebyshev’s inequality which states that

\[
\mathbb{P}\left(\left|\frac{\hat{m}_{\beta, \varphi(Y)\varphi(Y)}(y_n)}{m_{\varphi(Y)\varphi(Y)}(y_n)} - \frac{\langle \beta, m_{\varphi(Y)\varphi(Y)}(y_n) \rangle}{m_{\varphi(Y)\varphi(Y)}(y_n)}\right| > \eta'\right) \leq \frac{\text{Var}(\hat{m}_{\beta, \varphi(Y)\varphi(Y)}(y_n))}{m_{\varphi(Y)\varphi(Y)}(y_n)^2}, \quad \forall \eta' > 0.
\] (5.8)

The first step is to show that the RHS in (5.8) converges to zero for large \(n\). Focusing on the variance, it is straightforward to see that, by Lemma 5.1, independence and equidistribution, \(\delta_n = O(1)\),

\[
\text{Var}(\hat{m}_{\beta, \varphi(Y)\varphi(Y)}(y_n)) = \frac{1}{n} \mathbb{V}(\langle \beta, \varphi(Y) \varepsilon 1_{\{Y \geq y_n\}} \rangle).
\]

By positivity, one has

\[
\text{Var}(\langle \beta, \varphi(Y) \varepsilon 1_{\{Y \geq y_n\}} \rangle) \leq \mathbb{E}(\langle \beta, \varphi(Y) \varepsilon 1_{\{Y \geq y_n\}} \rangle)^2 = \mathbb{E}(\varphi^2(Y) 1_{\{Y \geq y_n\}} \langle \beta, \varepsilon \rangle^2).
\]
Now, Cauchy-Schwarz’s inequality in \( H \) entails
\[
\text{Var} \left( \hat{m}_{(\beta,\varepsilon)}(y_n) \right) \leq \frac{1}{n} \mathbb{E} \left( \varphi^2(Y)1_{\{Y \geq y_n\}} \| \varepsilon \|^2 \right) = \frac{1}{n} m_{\varphi^2(Y)}(y_n).\]
Next, using the bound (5.6) in the proof of Lemma 5.4, which is allowed since (2.5) and (3.4) hold, there exists a constant \( c \in (0, +\infty) \) independent of \( n \) such that for \( n \) large enough,
\[
\text{Var} \left( \hat{m}_{(\beta,\varepsilon)}(y_n) \right) \leq cn^{-1}\varphi^2(y_n)F^{1/2}(y_n).
\]
The denominator of the RHS in (5.8) is controlled with Lemma 5.2:
\[
m_{\varphi^2(Y)}(y_n) \sim \frac{1}{(1 - (\kappa + \tau)^2)}\varphi^2(y_n)g^2(y_n)F^2(y_n) \quad \text{as} \quad n \to +\infty.
\]
Overall, the Chebychev’s inequality (5.8) becomes, for some constant \( c \in (0, +\infty) \) independent of \( n \), for any \( n \) large enough and any \( \eta' > 0 \),
\[
\mathbb{P} \left( \left| \frac{\hat{m}_{(\beta,\varepsilon)}(y_n)}{m_{\varphi^2(Y)}(y_n)} - \frac{\langle \beta, m_{\varphi(Y)}(y_n) \rangle}{m_{\varphi^2(Y)}(y_n)} \right| > \eta' \right) \leq \frac{c}{\eta'^2} (nF(y_n)\delta_n)^{-1}.
\]
Now, since \( \eta' \) is arbitrary, we may pick \( \eta' \) of the form \( \delta_n(nF(y_n)\eta/c)^{-1/2} \) for any \( \eta > 0 \) leading to
\[
\frac{\hat{m}_{(\beta,\varepsilon)}(y_n)}{m_{\varphi^2(Y)}(y_n)} = \frac{\langle \beta, m_{\varphi(Y)}(y_n) \rangle}{m_{\varphi^2(Y)}(y_n)} + O_P \left( \delta_n(nF(y_n))^{-1/2} \right).
\]
The next step is to evaluate the first term of the RHS in the previous asymptotic expansion. While the denominator is controlled by Lemma 5.2, the numerator is bounded using Cauchy-Schwarz’s inequality in \( H \) and then (5.7) from the proof of Lemma 5.4. This chain of arguments gives:
\[
\left| \frac{\langle \beta, m_{\varphi(Y)}(y_n) \rangle}{m_{\varphi^2(Y)}(y_n)} \right| \leq \frac{m_{^1 - 1/q, Y}(y_n)}{m_{\varphi^2(Y)}(y_n)} \leq c \left( g(y_n)F^{1/q}(y_n) \right)^{-1} = c\delta_n.
\]
Finally, combining the previous bound with the regime (3.1) and the fact that we derived
\[
\frac{\hat{m}_{(\beta,\varepsilon)}(y_n)}{m_{\varphi^2(Y)}(y_n)} = O_P \left( \delta_n + \delta_n(nF(y_n))^{-1/2} \right) = O_P \left( \delta_n \right),
\]
the conclusion follows. \( \square \)

5.5. Consistency of \( \hat{\beta}_\varphi \).

**Proof of Proposition 3.1.** Under (2.4), the estimator \( \hat{\varphi} \) may be expressed as,
\[
\hat{\varphi}(y_n) = \frac{1}{n} \sum_{i=1}^{n} \varphi(Y_i)g(Y_i)1_{\{Y_i \geq y_n\}} + \frac{1}{n} \sum_{i=1}^{n} \varphi(Y_i)\varepsilon_i1_{\{Y_i \geq y_n\}} = \hat{m}_{\varphi g}(y_n)\beta + \hat{m}_{\varphi(Y)}(y_n)\varepsilon(y_n).
\]
So that, since \( \| \beta \| = 1 \), one may write the inner product between \( \beta \) and \( \hat{\varphi}(y_n) \) as
\[
\langle \hat{\varphi}(y_n), \beta \rangle = \hat{m}_{\varphi g}(y_n) + \hat{m}_{\varphi(Y)}(y_n)\varepsilon(y_n),
\]
and \( \| \hat{\varphi}(y_n) \|^2 = \hat{m}_{\varphi^2(Y)}(y_n) + \| \hat{m}_{\varphi(Y)}(y_n) \|^2 + 2\hat{m}_{\varphi g}(y_n)\hat{m}_{\varphi(Y)}(y_n)\varepsilon(y_n) + 2\hat{m}_{\varphi(Y)}(y_n)\hat{m}_{\varphi(Y)}(y_n)\varepsilon^2(y_n) \). It follows that
\[
1 - \frac{\langle \hat{\varphi}(y_n), \beta \rangle^2}{\| \hat{\varphi}(y_n) \|^2} = \frac{\| \hat{m}_{\varphi(Y)}(y_n) \|^2 - \hat{m}_{\varphi(Y)}^2(\varepsilon(y_n)\varepsilon(y_n))}{\hat{m}_{\varphi^2(Y)}(y_n) + 2\hat{m}_{\varphi g}(y_n)\hat{m}_{\varphi(Y)}(y_n)\varepsilon(y_n) + \| \hat{m}_{\varphi(Y)}(y_n) \|^2} = \frac{\left( \| \hat{m}_{\varphi(Y)}(y_n) \|^2/\hat{m}_{\varphi g}(y_n) \right) - \left( \hat{m}_{\varphi(Y)}(\varepsilon(y_n))/\hat{m}_{\varphi g}(y_n) \right)^2}{1 + 2 \left( \hat{m}_{\varphi(Y)}(\varepsilon(y_n))/\hat{m}_{\varphi g}(y_n) \right) + \left( \| \hat{m}_{\varphi(Y)}(y_n) \|^2/\hat{m}_{\varphi^2(Y)}(y_n) \right)^2}.
\]
In view of Lemma 5.3, we may write after a first order Taylor expansion, with respect to the convergence in probability,

\[ 1 - \frac{(\langle \tilde{v}_\varphi(y_n), \beta \rangle)^2}{\| \tilde{v}_\varphi(y_n) \|^2} \overset{P}{\sim} 1 + 2 \langle \tilde{m}_\varphi(y_n), \beta \rangle \frac{\| \tilde{m}_\varphi(y_n) \|^2}{\| \tilde{v}_\varphi(y_n) \|^2} + \left( \frac{\| \tilde{m}_\varphi(y_n) \|^2}{\| \tilde{v}_\varphi(y_n) \|^2} \right)^2. \]

The conditions to apply Lemma 5.4 and Lemma 5.6 are fulfilled so that a Taylor expansion yields

\[ 1 - \frac{(\langle \tilde{v}_\varphi(y_n), \beta \rangle)^2}{\| \tilde{v}_\varphi(y_n) \|^2} = O_P(\sigma_n^2) \xrightarrow{n \to \infty} 0. \]

Finally, note that $\| \hat{\beta}_\varphi(y_n) - \beta \|^2 = 2(1 - (\langle \tilde{v}_\varphi(y_n), \beta \rangle)^2/\| \tilde{v}_\varphi(y_n) \|^2)$ to conclude the proof.

**Proof of Proposition 3.2.** To ease the notations, let us introduce

\[ \hat{C}_n := \text{Cov}(Y_0, \langle \hat{\beta}_\varphi(y_n), X_0 \rangle | Y_0 \geq y_n), \]

\[ C_n := \text{Cov}(Y_0, \langle \beta, X_0 \rangle | Y_0 \geq y_n). \]

Under the model (2.4), one may decompose

\[ \hat{C}_n = \text{Cov}(Y_0, g(Y_0) \langle \hat{\phi}_\varphi(y_n), \beta \rangle | Y_0 \geq y_n) + \text{Cov}(Y_0, \langle \hat{\beta}_\varphi(y_n), \varepsilon \rangle | Y_0 \geq y_n) \]

\[ = \mathbb{E}(Y_0g(Y_0) \langle \hat{\beta}_\varphi(y_n), \beta \rangle | Y_0 \geq y_n) - \mathbb{E}(Y_0 \text{Cov}(\hat{\phi}_\varphi(y_n), \beta) | Y_0 \geq y_n) \]

\[ + \mathbb{E}(Y_0 \langle \hat{\phi}_\varphi(y_n), \varepsilon_0 \rangle | Y_0 \geq y_n) - \mathbb{E}(Y_0 \text{Cov}(\varepsilon_0, \beta) | Y_0 \geq y_n). \]

Taking the difference yields $\hat{C}_n - C_n = \sum_{\ell=1}^4 A_{\ell,n}$ where,

\[ A_{1,n} := \mathbb{E}(Y_0g(Y_0)(\langle \hat{\beta}_\varphi(y_n), \beta \rangle - 1) | Y_0 \geq y_n), \]

\[ A_{2,n} := \mathbb{E}(Y_0 \langle \beta, \varepsilon_0 \rangle | Y_0 \geq y_n) \mathbb{E}(g(Y_0)(1 - \langle \hat{\beta}_\varphi(y_n), \beta \rangle) | Y_0 \geq y_n), \]

\[ A_{3,n} := \mathbb{E}(Y_0 \langle \beta, \varepsilon_0 \rangle | Y_0 \geq y_n), \]

\[ A_{4,n} := \mathbb{E}(Y_0 \langle \beta, \varepsilon_0 \rangle | Y_0 \geq y_n) \mathbb{E}(\langle \varepsilon_0, \beta - \hat{\beta}_\varphi(y_n) \rangle | Y_0 \geq y_n). \]

The first step for deriving the statement is to control $\hat{C}_n - C_n$. To this end, we apply the triangular inequality which gives $|\hat{C}_n - C_n| \leq \sum_{\ell=1}^4 |A_{\ell,n}|$ and we subsequently bound each term $|A_{\ell,n}|, \ell = 1, \ldots, 4$. Starting with $A_{1,n}$, one observes that $\hat{\beta}_\varphi(y_n)$ only depends on $X_1, \ldots, X_n$ so that, by independence, equidistribution and Jensen’s inequality,

\[ |A_{1,n}| \leq \frac{1}{F(y_n)} \mathbb{E}(Y_0g(Y_0)1_{\{Y_0 \geq y_n\}}) \mathbb{E}(\langle \hat{\beta}_\varphi(y_n), \beta \rangle - 1). \]

Now, from Proposition 3.1 and a Taylor expansion of the square root, we have that $\langle \hat{\beta}_\varphi(y_n), \beta \rangle = 1 + O(\delta_n)$. Since, moreover, $|\langle \hat{\beta}_\varphi(y_n), \beta \rangle - 1| \leq 2$ holds almost surely by Cauchy-Schwarz’s inequality in $H$, we may apply the dominated convergence theorem and Lemma 5.2, leading to $|A_{1,n}| = O(y_n g(y_n) \delta_n)$. Indeed, a general fact is that, if a generic sequence $(W_n)$ of real random variables bounded almost surely by $M > 0$ converges to zero in probability with some rate, then it also converges in expectation to zero with the same rate. This may be derived by noticing that, for any $\eta > 0$,

\[ \mathbb{E}(|W_n|) = \int_0^{+\infty} \mathbb{P}(|W_n| \geq x) \, dx = \int_0^{\eta} \mathbb{P}(|W_n| \geq x) \, dx + \int_{\eta}^M \mathbb{P}(|W_n| \geq x) \, dx \leq \eta + M \mathbb{P}(|W_n| > \eta). \]
Similarly, $|A_{2,n}| = O(y_n g(y_n) \delta_n)$. Focussing on the third term, we give a first bound using independence between $(Y_0, \epsilon_0)$ and $(X_i)_{1 \leq i \leq n}$, Lemma 5.1, Cauchy-Schwarz’s inequality in $H$ and equidistribution:

$$|A_{3,n}| = \frac{1}{F(y_n)} \left| \mathbb{E} \left( (\hat{\beta}_n(y_n) - \beta, Y_0 \epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}}) \right) \right| = \frac{1}{F(y_n)} \left| \mathbb{E}_{\mathcal{B}}(\hat{\beta}_n(y_n) - \beta, \mathbb{E}_B(Y_0 \epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}})) \right|$$

$$\leq \frac{1}{F(y_n)} \mathbb{E}_B(Y_0 \epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}}) \mathbb{E}_B(\hat{\beta}_n(y_n) - \beta).$$

Next, an application of Hölder’s inequality and Lemma 5.2 yields, for some $c \in (0, +\infty)$,

$$\left\| \mathbb{E}_B(Y_0 \epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}}) \right\| \leq \mathbb{E}(\|\epsilon_0\| q)^{1/q} m_n^{1 - 1/q} \sqrt[n]{y_0^2} \leq c y_n F^{-1/q}(y_n).$$

(5.9)

Besides, since $\hat{\beta}_n(y_n)$ and $\beta$ are unit vectors of $H$, using Jensen’s inequality and the same argument based on the dominated convergence theorem altogether with Proposition 3.1, we may write

$$\left\| \mathbb{E}_B(\hat{\beta}_n(y_n) - \beta)^2 \leq \mathbb{E} \left\| \hat{\beta}_n(y_n) - \beta \right\|^2 = \mathbb{E} \left\| \hat{\beta}_n(y_n) \right\|^2 + \mathbb{E} \|\beta\|^2 - 2 \mathbb{E}(\hat{\beta}_n(y_n), \beta)$$

$$\leq 2(1 - \mathbb{E}(\hat{\beta}_n(y_n), \beta)) = O(\delta_n).$$

Combining the last three inequalities, we have shown that $|A_{3,n}| = O(y_n F^{-1/q}(y_n) \delta_n^{1/2})$. Lastly, for some constant $c \in (0, +\infty)$ independent of $n$ and for $n$ large enough,

$$|A_{4,n}| = \frac{m_Y(y_n)}{F^2(y_n)} \left| \mathbb{E} \left( (\hat{\beta}_n(y_n) - \beta, \epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}}) \right) \right| = \frac{m_Y(y_n)}{F^2(y_n)} \left| \mathbb{E}_{\mathcal{B}}(\hat{\beta}_n(y_n) - \beta, \mathbb{E}_B(\epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}})) \right|$$

$$\leq \frac{m_Y(y_n)}{F^2(y_n)} \left\| \mathbb{E}_B(\epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}}) \right\| \left\| \mathbb{E}_B(\hat{\beta}_n(y_n) - \beta) \right\| \leq c \frac{m_Y(y_n)}{F^2(y_n)} \left\| \mathbb{E}_B(\epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}}) \right\| \delta_n.$$ Applying Hölder’s inequality once more so that, for some constant $c \in (0, +\infty)$ only depending on $\gamma$ and $q$, we may write

$$\left\| \mathbb{E}_B(\epsilon_0\mathbb{1}_{\{Y_0 \geq y_n\}}) \right\| \leq \mathbb{E}(\|\epsilon_0\| q)^{1/q} F^{-1/q}(y_n) \leq c F^{-1/q}(y_n),$$

(5.10) It follows that $|A_{4,n}| = O(y_n F^{-1/q}(y_n) \delta_n^{1/2}) = o(y_n g(y_n) \delta_n)$ under assumption (3.3). Overall, this yields

$$C_n - C_n = O\left( \max_{\epsilon \in \{1, 2, 3, 4\}} |A_{\epsilon,n}| \right) = O\left( y_n g(y_n) \delta_n \right) = O\left( y_n F^{-1/q}(y_n) \right).$$

(5.11) In a second step, we give an asymptotic equivalent of $C_n$ and start by noticing that, by equidistribution, it may be written as,

$$C_n = \frac{m_Y g(Y)(y_n)}{F(y_n)} - \frac{m_Y(y_n)m_Y(Y)(y_n)}{F^2(y_n)} + \frac{\mathbb{E}(Y(\beta, \epsilon) \mathbb{1}_{\{Y \geq y_n\}})}{F(y_n)} + \frac{\mathbb{E}(\langle \beta, \epsilon \rangle \mathbb{1}_{\{Y \geq y_n\}})}{F^2(y_n)}.$$

Lemma 5.2 yields the asymptotic expansion

$$C_n = c_{\gamma,n} y_n g(y_n) (1 + o(1)) + \frac{\mathbb{E}(Y(\beta, \epsilon) \mathbb{1}_{\{Y \geq y_n\}})}{F(y_n)} + \frac{1}{1 - \gamma} \frac{\mathbb{E}(\langle \beta, \epsilon \rangle \mathbb{1}_{\{Y \geq y_n\}})}{F(y_n)} + \frac{1}{(1 - \gamma)c_{\gamma,n}} \frac{\mathbb{E}(\langle \beta, \epsilon \rangle \mathbb{1}_{\{Y \geq y_n\}})}{y_n g(y_n) F(y_n)} (1 + o(1)) + o(1)$$

where $c_{\gamma,n} := \frac{1}{1 - (\epsilon + 1)\gamma} - \frac{1}{(1 - \gamma)(1 - c\gamma)}$. Next, combining (5.9)–(5.10) with the RHS of the previous asymptotic expansion readily implies that, as $n \to +\infty$,

$$C_n = c_{\gamma,n} y_n g(y_n) \left[ 1 + O(\delta_n) + O(\delta_n/y_n) + o(1) \right] \sim c_{\gamma,n} y_n g(y_n).$$

(5.12)
The final step consists in assembling the previous ingredients (5.11) and (5.12):

\[
\frac{\hat{C}_n}{C_n} - 1 = \frac{\hat{C}_n - C_n}{|C_n|} = O\left(\frac{y_nF^{-1/q}(y_n)}{c_{\gamma,\kappa}y_n g(y_n)(1+o(1))}\right) = O(\delta_n) \ll 1,
\]

in view of (3.3).

\[
\square
\]

References


