



Functional estimation of extreme conditional expectiles

Stéphane Girard, Gilles Stupfler, Antoine Usseglio-Carleve

► To cite this version:

Stéphane Girard, Gilles Stupfler, Antoine Usseglio-Carleve. Functional estimation of extreme conditional expectiles. *Econometrics and Statistics* , 2022, 21, pp.131-158. 10.1016/j.ecosta.2021.05.006 . hal-03117547v2

HAL Id: hal-03117547

<https://inria.hal.science/hal-03117547v2>

Submitted on 1 Mar 2021

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.

Functional estimation of extreme conditional expectiles

Stéphane Girard⁽¹⁾, Gilles Stupfler⁽²⁾ & Antoine Usseglio-Carleve⁽¹⁾

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

⁽²⁾ Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

Abstract. Quantiles and expectiles can be interpreted as solutions of convex minimization problems. Unlike quantiles, expectiles are determined by tail expectations rather than tail probabilities, and define a coherent risk measure. For these reasons, among others, they have recently been the subject of renewed attention in actuarial and financial risk management. Here, we focus on the challenging problem of estimating extreme expectiles, whose order converges to one as the sample size increases, given a functional covariate. We construct a functional kernel estimator of extreme conditional expectiles by writing expectiles as quantiles of a different distribution. The asymptotic properties of the estimators are studied in the context of conditional heavy-tailed distributions. We also provide and analyse different ways of estimating the functional tail index, as a way to extrapolate our estimates to the very far conditional tails. A numerical illustration of the finite-sample performance of our estimators is provided on simulated and real datasets.

Keywords. Conditional tail index, expectiles, extrapolation, extremes, functional kernel estimator, heavy tails, nonparametric estimation.

1 Introduction

The quantile, or Value-at-Risk in the actuarial and financial literature, is arguably the most widespread tool in risk management. This is largely due to the simplicity and interpretability of the quantile: if Y is a random variable with cumulative distribution function F , the quantile of level $\alpha \in (0, 1)$ of Y is given by the generalized inverse $q_\alpha(Y) = \inf\{y \in \mathbb{R} | F(y) \geq \alpha\}$, and when F is continuous and (strictly) increasing, $q_\alpha(Y)$ is the unique solution q of the equation $F(q) = \mathbb{P}(Y \leq q) = \alpha$. An estimation of high quantiles, for α close to 1, is then a reasonable way to get an understanding of extreme risk, routinely used when considering large claim amounts in insurance, financial losses, or extreme wave heights during storms in environmental science, for example. However, in the context of financial and actuarial risk management, it has been repeatedly emphasized that the quantile has important shortcomings. One of them is that the quantile does not induce a coherent risk measure in the sense of Artzner et al. (1999), in particular because it does not fulfill the subadditivity property (see Acerbi (2002)). Another major drawback is the reliance of quantiles on the frequency of tail events and not on their actual magnitudes; this is of course an issue in risk management, where it is important to find not only what constitutes an extreme level of loss but also what a typical extreme loss will be.

These pitfalls motivated part of the risk management community to propose alternative risk measures. One of them is induced by expectiles, introduced in Newey and Powell (1987) and defined by the following minimization problem:

$$e_\alpha(Y) = \arg \min_{t \in \mathbb{R}} \mathbb{E} [\eta_\alpha(Y - t) - \eta_\alpha(Y)],$$

where $\eta_\alpha(y) = |\alpha - \mathbb{1}_{\{y \leq 0\}}|y^2$ is the so-called expectile loss function and $\mathbb{1}_{\{\cdot\}}$ is the indicator function. Expectiles can be seen as L^2 -analogues of quantiles, in the sense that (see Koenker and Bassett (1978)):

$$q_\alpha(Y) \in \arg \min_{t \in \mathbb{R}} \mathbb{E} [\rho_\alpha(Y - t) - \rho_\alpha(Y)],$$

where $\rho_\alpha(y) = |\alpha - \mathbb{1}_{\{y \leq 0\}}||y|$ is called quantile loss function. Both quantiles and expectiles are part of the wider family of L^p -quantiles (introduced in Chen (1996)), a subfamily of M -quantiles (introduced in Breckling and Chambers (1988)). Unlike quantiles, expectiles are determined by tail expectations rather than tail probabilities, and their calculation is thus driven both by the probability of tail values and their actual realizations (see Kuan et al., 2009). Moreover, as shown in Bellini et al. (2014), expectiles with $\alpha \geq 1/2$ are the only M -quantiles to fulfill the coherence property. In fact, such expectiles define the only coherent law-invariant risk measure that is elicitable (see Ziegel, 2016), meaning that they benefit from the existence of a natural backtesting methodology (see Gneiting, 2011). The interest for expectiles is therefore growing in the financial and actuarial risk management literature (see, for instance, Taylor (2008) and Cai and Weng (2016)). In particular, the problem of extreme expectile estimation for heavy-tailed random variables has recently been considered in Daouia et al. (2018) and Daouia et al. (2020). A substantial difficulty when working with expectiles is that they do not benefit from closed form expressions, making their analysis significantly harder than that of quantiles.

In many applications, the random variable Y is recorded along with auxiliary information represented by a random covariate \mathbf{X} . The difficulty of working with extreme expectiles is then compounded by the necessity of integrating this auxiliary information into the estimation. When $\mathbf{X} \in \mathbb{R}^p$ is a vector, the problem of estimating such extreme conditional expectiles has been considered in Girard et al. (2021) from a nonparametric perspective, by writing expectiles as quantiles of a different distribution and using kernel estimation of distribution functions. The basic idea behind this procedure dates back to Jones (1994). This followed a series of earlier papers by Daouia et al. (2018, 2019, 2020) on the estimation of extreme unconditional expectiles using their interpretation as convex minimizers. To the best of our knowledge, when \mathbf{X} belongs to a more general, potentially infinite-dimensional, functional space, the problem of estimating extreme conditional expectiles has not yet been considered. This functional setup is precisely the focus of this paper, in which we deal with the case when Y given \mathbf{X} has a heavy right tail. The heavy-tailed assumption, which is ubiquitous in actuarial and financial risk management (see e.g. the discussions on p.9 of Embrechts et al. (1997) and p.1 of Resnick (2007)), makes it possible to estimate extreme conditional quantiles and expectiles using a well-known extrapolation argument of Weissman (1978), via the estimation of the so-called tail index: conditional expectiles are first estimated at an intermediate level (that is, increasing with sample size, but

not too high), and these estimates are then extrapolated using the shape of the conditional heavy-tailed distribution to obtain estimators of properly extreme conditional expectiles.

Our paper is organized as follows. Section 2 introduces our notation, assumptions, and the general idea behind our approach, in which we adapt the two-step, extreme conditional quantile estimation methodology of Gardes and Girard (2012) to the estimation of extreme functional expectiles. Section 3 gives the asymptotic properties of our estimator of the so-called intermediate functional expectiles, which are the basic building blocks of our estimators of properly extreme functional expectiles considered in Section 4. In the latter section, we will explain that the estimation of the functional tail index of Y given \mathbf{X} is crucial, and in Section 5 we suggest and study a handful of estimators of this quantity. Section 6 proposes a simulation study to give an overview of the performance of our procedure. Section 7 concludes by showcasing our technique on a real financial data example. All proofs are relegated to the Appendix.

2 Notation and assumptions

Let (\mathbf{X}_i, Y_i) , $i = 1, \dots, n$ be independent copies of a random pair (\mathbf{X}, Y) in $E \times \mathbb{R}$, where E is a functional space endowed with a semi-metric d . We assume that the probability distribution of \mathbf{X} is non-atomic, and that the topology of E makes it possible to define the family of conditional distributions of Y given $\mathbf{X} = \mathbf{x}$. For all $\mathbf{x} \in E$, the conditional survival function (c.s.f.) of Y given $\mathbf{X} = \mathbf{x}$ is denoted by $\bar{F}(y|\mathbf{x}) = \mathbb{P}(Y > y | \mathbf{X} = \mathbf{x})$. Discussing the existence of regular versions of $\bar{F}(\cdot|\cdot)$ is beyond the scope of this paper; it is guaranteed when (E, d) is a Polish space, see Theorem 1 and its discussion in Chang and Pollard (1997). When d is a semi-metric defined exclusively on low-frequency components of a truncated basis expansion (for example when $E = L^2[0, 1]$), the relevant conditional distributions will be defined with respect to the quotient space of E by the equivalence relation considering that two elements are different if and only if one pair of their low-frequency components are different. This quotient space is then essentially a finite-dimensional linear space on which conditional distributions indeed exist, the price to pay being of course that the conditional distributions only identify lower-frequency components.

In this context, Gardes and Girard (2012) address the estimation of the conditional quantiles of Y given $\mathbf{X} = \mathbf{x}$, defined by $q(\alpha|\mathbf{x}) = \inf \{y \in \mathbb{R} | \bar{F}(y|\mathbf{x}) \leq 1 - \alpha\}$, when the quantile level α is extreme, i.e. $\alpha = \alpha_n \rightarrow 1$ as $n \rightarrow \infty$. We briefly describe here their two-step approach. First, the c.s.f. is estimated thanks to a functional kernel estimator, also considered for instance in Ferraty et al. (2006) and Ferraty et al. (2007):

$$\widehat{\bar{F}}_n(y|\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n K \left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n} \right) \mathbb{1}_{\{Y_i > y\}} / \widehat{\mu}_K^{(1)}(\mathbf{x}, h_n), \quad (2.1)$$

where (h_n) is a positive bandwidth sequence such that $h_n \rightarrow 0$ as $n \rightarrow \infty$, $K : [0, \infty) \rightarrow [0, \infty)$ is a kernel function (namely, positive and measurable), and for all $b \geq 0$,

$$\widehat{\mu}_K^{(b)}(\mathbf{x}, h_n) = \frac{1}{n} \sum_{i=1}^n K^b \left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n} \right)$$

is the empirical counterpart of the quantity $\mu_K^{(b)}(\mathbf{x}, h_n) := \mathbb{E} [K^b(d(\mathbf{x}, \mathbf{X})/h_n)]$. We highlight that the particular case of the uniform kernel $K(t) = \mathbb{1}_{\{0 \leq t \leq 1\}}$ leads to $\mu_K^{(b)}(\mathbf{x}, h_n) = \mathbb{P}(\mathbf{X} \in B(\mathbf{x}, h_n)) =: \pi(\mathbf{x}, h_n)$, the so-called small ball probability of \mathbf{X} , see Ferraty et al. (2007) for a discussion on this topic. More generally, we assume that

(K) K is a function with support $[0, 1]$ and there exist $0 < c_1 < c_2 < \infty$ such that $c_1 \leq K(t) \leq c_2$ for all $t \in [0, 1]$.

One may also suppose without loss of generality that K integrates to one. In this case, K is called a type I kernel, see Definition 4.1 in Ferraty and Vieu (2006).

Second, the functional estimator of conditional quantiles $q(\alpha|\mathbf{x})$ is defined via the generalized inverse of (2.1):

$$\hat{q}_n(\alpha|\mathbf{x}) = \inf \left\{ y \in \mathbb{R} \mid \widehat{F}_n(y|\mathbf{x}) \leq 1 - \alpha \right\}. \quad (2.2)$$

In the situation where $\alpha \in (0, 1)$ is fixed, weak and strong consistency are proved respectively in Stone (1977) and Gannoun (1990) while asymptotic normality is shown in Stute (1986), Samanta (1989) and Berlinet et al. (2001) when E is finite-dimensional and by Ferraty et al. (2005) for a general metric space under dependence assumptions. In Gardes and Girard (2012), the asymptotic distribution of the estimator (2.2) is investigated when estimating extreme quantiles, i.e. when $\alpha = \alpha_n \rightarrow 1$ as $n \rightarrow \infty$.

In this paper, we propose a similar approach for functional expectile regression. For that purpose, let us recall that, from Jones (1994), the expectile $e(\alpha|\mathbf{x})$ of level α of $Y|\mathbf{X} = \mathbf{x}$ may be seen as the quantile of level α associated to the c.s.f. $\bar{E}(y|\mathbf{x})$ defined by:

$$\bar{E}(y|\mathbf{x}) = \frac{\mathbb{E} [(Y - y) \mathbb{1}_{\{Y > y\}} | \mathbf{X} = \mathbf{x}]}{\mathbb{E} [|Y - y| | \mathbf{X} = \mathbf{x}]}.$$

Introducing, for a nonnegative integer k and provided the expectations exist,

$$\psi^{(k)}(y|\mathbf{x}) = \mathbb{E} [(Y - y)^k \mathbb{1}_{\{Y > y\}} | \mathbf{X} = \mathbf{x}] \quad \text{and} \quad m^{(k)}(y|\mathbf{x}) = \mathbb{E} [(Y - y)^k | \mathbf{X} = \mathbf{x}], \quad (2.3)$$

it thus follows that $e(\alpha|\mathbf{x}) = \inf\{y \in \mathbb{R} \mid \bar{E}(y|\mathbf{x}) \leq 1 - \alpha\}$, with

$$\bar{E}(y|\mathbf{x}) = \frac{\psi^{(1)}(y|\mathbf{x})}{2\psi^{(1)}(y|\mathbf{x}) - m^{(1)}(y|\mathbf{x})}. \quad (2.4)$$

In view of (2.3), we decide to build our estimator upon the following two random quantities:

$$\begin{aligned} \hat{\psi}_n^{(k)}(y|\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n (Y_i - y)^k K \left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n} \right) \mathbb{1}_{\{Y_i > y\}} / \hat{\mu}_K^{(1)}(\mathbf{x}, h_n), \\ \hat{m}_n^{(k)}(y|\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n (Y_i - y)^k K \left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n} \right) / \hat{\mu}_K^{(1)}(\mathbf{x}, h_n). \end{aligned}$$

The quantities $\hat{\psi}_n^{(k)}(y|\mathbf{x})$ and $\hat{m}_n^{(k)}(y|\mathbf{x})$ are (under technical conditions) consistent estimators of $\psi^{(k)}(y|\mathbf{x})$ and $m^{(k)}(y|\mathbf{x})$, respectively. The c.s.f. $\bar{E}(\cdot|\mathbf{x})$ defined in (2.4) is then estimated by:

$$\hat{\bar{E}}_n(y|\mathbf{x}) = \frac{\hat{\psi}_n^{(1)}(y|\mathbf{x})}{2\hat{\psi}_n^{(1)}(y|\mathbf{x}) - \hat{m}_n^{(1)}(y|\mathbf{x})}. \quad (2.5)$$

We refer to Girard et al. (2021) for a similar approach in the case where the covariate \mathbf{X} is finite-dimensional.

In the following section and below, our aim is to estimate $e(\alpha|\mathbf{x})$ when the expectile level is extreme, i.e. $\alpha = \alpha_n \rightarrow 1$ as $n \rightarrow \infty$. We introduce several modelling and regularity assumptions in this respect. We first assume that $Y|\mathbf{X} = \mathbf{x}$ has a heavy-tailed distribution, namely, there exists $\gamma(\mathbf{x}) > 0$ such that the c.s.f. is regularly varying at infinity with index $-1/\gamma(\mathbf{x})$:

$$\forall y > 0, \lim_{t \rightarrow \infty} \frac{\bar{F}(ty|\mathbf{x})}{\bar{F}(t|\mathbf{x})} = y^{-1/\gamma(\mathbf{x})}. \quad (2.6)$$

We refer to Bingham et al. (1989) for a general account on regular variation. In our context, $\gamma(\mathbf{x})$ is referred to as the functional tail index since it tunes the tail heaviness of the conditional distribution of Y given $\mathbf{X} = \mathbf{x}$. More precisely, the heavy tail condition (2.6) is refined in the following way:

$\mathcal{C}_1(\gamma(\mathbf{x}))$ There exists $\gamma(\mathbf{x}) > 0$ such that $\bar{F}(\cdot|\mathbf{x})$ is continuously differentiable and satisfies

$$\lim_{y \rightarrow \infty} \frac{y\bar{F}'(y|\mathbf{x})}{\bar{F}(y|\mathbf{x})} = -1/\gamma(\mathbf{x}).$$

It follows from Theorem 1.1.11 in de Haan and Ferreira (2006) that condition (2.6) is indeed satisfied if $\mathcal{C}_1(\gamma(\mathbf{x}))$ holds. Another consequence of condition $\mathcal{C}_1(\gamma(\mathbf{x}))$ is that the conditional density function $f(\cdot|\mathbf{x}) = -\bar{F}'(\cdot|\mathbf{x})$ exists and is also regularly varying, with index $-1/\gamma(\mathbf{x}) - 1$. This regularity condition is unlikely to be very restrictive in practice due to the fact that all commonly used heavy-tailed models satisfy this condition. In addition, Newey and Powell (1987) point out in their Theorem 1 that continuous differentiability of the c.s.f. is a reasonable sufficient condition for expectiles to characterize the underlying distribution.

Condition $\mathcal{C}_1(\gamma(\mathbf{x}))$ will reveal useful in Section 3 to estimate extreme expectiles $e(\alpha_n|\mathbf{x})$ associated with intermediate levels α_n and located within the sample. For higher levels α_n , extreme expectiles may be located outside the sample. Extrapolation techniques are developed in Section 4, and they require a stronger condition on the distribution tail:

$\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ There exist $\gamma(\mathbf{x}) > 0$, $\rho(\mathbf{x}) \leq 0$ and a positive or negative function $A(\cdot|\mathbf{x})$ such that:

$$\forall y > 0, \lim_{t \rightarrow \infty} \frac{1}{A(t|\mathbf{x})} \left(\frac{q(1 - 1/(ty)|\mathbf{x})}{q(1 - 1/t|\mathbf{x})} - y^{\gamma(\mathbf{x})} \right) = \begin{cases} y^{\gamma(\mathbf{x})} \frac{y^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} & \text{if } \rho(\mathbf{x}) < 0, \\ y^{\gamma(\mathbf{x})} \log y & \text{if } \rho(\mathbf{x}) = 0. \end{cases} \quad (2.7)$$

According to Theorem 2.3.9 in de Haan and Ferreira (2006), this so-called second-order condition (2.7) generalizes condition (2.6), since it is equivalent to

$$\forall y > 0, \lim_{t \rightarrow \infty} \frac{1}{A(1/\bar{F}(t|\mathbf{x})|\mathbf{x})} \left(\frac{\bar{F}(ty|\mathbf{x})}{\bar{F}(t|\mathbf{x})} - y^{-1/\gamma(\mathbf{x})} \right) = y^{-1/\gamma(\mathbf{x})} \frac{y^{\rho(\mathbf{x})/\gamma(\mathbf{x})} - 1}{\gamma(\mathbf{x})\rho(\mathbf{x})}.$$

Our last assumption is a local Lipschitz condition on the conditional moments. Similar conditions are used, for instance, in Krzyzak (1986), or more recently in Daouia et al. (2013) and El Methni

et al. (2014) in the context of conditional extreme value analysis with a finite-dimensional covariate. Let us denote by $B(\mathbf{x}, r)$ the ball with center $\mathbf{x} \in E$ and radius $r > 0$, associated with the semi-metric d , and by \vee the maximum operator.

(\mathcal{L}) One has $m^{(2)}(0|\mathbf{x}) = \mathbb{E}[Y^2 | \mathbf{X} = \mathbf{x}] < \infty$ and there exist $c, r > 0$ such that

$$\forall \mathbf{x}' \in B(\mathbf{x}, r), \left| m^{(1)}(0|\mathbf{x}') - m^{(1)}(0|\mathbf{x}) \right| \vee \left| m^{(2)}(0|\mathbf{x}') - m^{(2)}(0|\mathbf{x}) \right| \leq c d(\mathbf{x}, \mathbf{x}').$$

We conclude this paragraph by introducing some notation, which will prove very useful, for the oscillation of the c.s.f. $\bar{F}(\cdot|\mathbf{x})$ above a high level y_n :

$$\omega_{h_n}(y_n|\mathbf{x}) = \sup_{\substack{z \geq y_n \\ \mathbf{x}' \in B(\mathbf{x}, h_n)}} \frac{1}{\log z} \left| \log \frac{\bar{F}(z|\mathbf{x}')}{\bar{F}(z|\mathbf{x})} \right|.$$

The quantity $\omega_{h_n}(y_n|\mathbf{x})$ measures the discrepancy between the extremes of the conditional distributions of Y at neighboring points. Similar quantities are introduced in Gardes and Stupfler (2014, 2019) and Stupfler (2013, 2016). In order to get an idea of the typical asymptotic behavior of $\omega_{h_n}(y_n|\mathbf{x})$, consider the Karamata representation of $\bar{F}(\cdot|\mathbf{x})$:

$$\forall z \geq 1, \bar{F}(z|\mathbf{x}) = z^{-1/\gamma(\mathbf{x})} \exp \left(\eta(z|\mathbf{x}) + \int_1^z \frac{\epsilon(u|\mathbf{x})}{u} du \right),$$

where $\eta(\cdot|\mathbf{x})$ and $\epsilon(\cdot|\mathbf{x})$ are measurable functions converging, respectively, to a constant and 0 at infinity, see Theorem 1.3.1 in Bingham et al. (1989). In this context, it is straightforward to prove that if there are $c, r > 0$ and $z_0 > 1$ with

$$\forall \mathbf{x}' \in B(\mathbf{x}, r), \left| \gamma(\mathbf{x}) - \gamma(\mathbf{x}') \right| \vee \sup_{z \geq z_0} \left| \frac{\eta(z|\mathbf{x}) - \eta(z|\mathbf{x}')}{\log z} \right| \vee \sup_{z \geq 1} |\epsilon(z|\mathbf{x}) - \epsilon(z|\mathbf{x}')| \leq c d(\mathbf{x}, \mathbf{x}'), \quad (2.8)$$

then $\omega_{h_n}(y_n|\mathbf{x}) = O(h_n)$.

3 Intermediate functional expectile estimation

We consider from now on the case when $\gamma(\mathbf{x}) < 1$ and $Y_- := \max(-Y, 0)$ has a finite first conditional moment. These conditions combined ensure that $\mathbb{E}(|Y| | \mathbf{X} = \mathbf{x}) < \infty$ and thus that conditional expectiles at any order exist indeed. These will be our minimal working conditions throughout this paper.

Let us consider a sequence (α_n) such that $\alpha_n \rightarrow 1$ as $n \rightarrow \infty$. It has been established in Gardes and Girard (2012, Theorem 2) that the functional estimator (2.2) of the extreme quantile $q(\alpha_n|\mathbf{x})$ is asymptotically Gaussian provided that $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ as $n \rightarrow \infty$. In this situation, $q(\alpha|\mathbf{x})$ is referred to as an intermediate conditional quantile. It is, for n large enough, located within the sample since there is then almost surely at least one sample point in the region $B(\mathbf{x}, h) \times (q(\alpha|\mathbf{x}), \infty) \subset E \times \mathbb{R}$, see Gardes and Girard (2012, Lemma 4). Condition $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ in fact ensures that there will be a growing number of observations larger than the conditional quantile to be estimated. In this section, we propose a similar approach for estimating intermediate conditional expectiles. Let

$$\hat{e}_n(\alpha_n|\mathbf{x}) = \inf \left\{ y \in \mathbb{R} \mid \widehat{E}_n(y|\mathbf{x}) \leq 1 - \alpha_n \right\}. \quad (3.1)$$

Here $\widehat{E}_n(y|\mathbf{x})$ is defined in Equation (2.5). We start by studying the joint asymptotic distribution of estimators (2.2) and (3.1), as this will be crucial in our construction of extrapolation procedures in Section 5. To this end, let us introduce

$$\sigma_n^{-1}(\mathbf{x}) = \left(n(1 - \alpha_n) \frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)} \right)^{1/2}.$$

Note that, under (\mathcal{K}) , $\sigma_n^{-1}(\mathbf{x})$ and $\sqrt{n(1 - \alpha_n)\pi(\mathbf{x}, h_n)}$ are of the same asymptotic order since

$$(c_1/c_2)^2 \pi(\mathbf{x}, h_n) \leq \frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)} \leq (c_2/c_1)^2 \pi(\mathbf{x}, h_n).$$

The following result may be seen as a functional version of Theorem 1 in Girard et al. (2021). It establishes the joint asymptotic normality of J empirical conditional intermediate expectiles $\widehat{e}_n(\alpha_{n,j}|\mathbf{x})$ with $1 - \alpha_{n,j} = \tau_j(1 - \alpha_n)$, $0 < \tau_1 < \dots < \tau_J \leq 1$, together with an empirical conditional intermediate quantile $\widehat{q}_n(a_n|\mathbf{x})$.

Theorem 1. *Assume that (\mathcal{K}) , (\mathcal{L}) and $\mathcal{C}_1(\gamma(\mathbf{x}))$ hold with $\gamma(\mathbf{x}) < 1/2$. Suppose that there exists $\delta \in (0, 1)$ such that $\mathbb{E}[Y_-^{2+\delta}|\mathbf{X} = \mathbf{x}] < \infty$. Let $\alpha_n \rightarrow 1$, $h_n \rightarrow 0$ as $n \rightarrow \infty$ and $\tau > 0$, $0 < \tau_1 < \tau_2 < \dots < \tau_J \leq 1$ with*

$$1 - a_n = \tau(1 - \alpha_n)(1 + o(1)) \quad \text{and} \quad 1 - \alpha_{n,j} = \tau_j(1 - \alpha_n),$$

for $j \in \{1, \dots, J\}$. Assume further that $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ and

$$\sigma_n^{-1}(\mathbf{x}) \log(1 - \alpha_n) \omega_{h_n}((1 - \delta)(e(\alpha_n|\mathbf{x}) \wedge q(a_n|\mathbf{x}))|\mathbf{x}) \rightarrow 0. \quad (3.2)$$

Then,

$$\sigma_n^{-1}(\mathbf{x}) \left\{ \left(\frac{\widehat{e}_n(\alpha_{n,j}|\mathbf{x})}{e(\alpha_{n,j}|\mathbf{x})} - 1 \right)_{1 \leq j \leq J}, \left(\frac{\widehat{q}_n(a_n|\mathbf{x})}{q(a_n|\mathbf{x})} - 1 \right) \right\} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{J+1}, \gamma^2(\mathbf{x})\Sigma(\mathbf{x})),$$

where $\Sigma(\mathbf{x})$ is the symmetric matrix having entries

$$\begin{aligned} \Sigma_{j,l}(\mathbf{x}) &= \tau_l^{-1} \left[\frac{1}{1 - 2\gamma(\mathbf{x})} \left(\frac{\tau_j}{\tau_l} \right)^{-\gamma(\mathbf{x})} - 1 \right], \\ \Sigma_{j,J+1}(\mathbf{x}) &= \tau_j^{-1} \left(\frac{\tau_j}{\tau} \right)^{\gamma(\mathbf{x})} \left[\left(\frac{1}{\gamma(\mathbf{x})^{-1} - 1} \wedge \frac{\tau_j}{\tau} \right)^{1-\gamma(\mathbf{x})} + \left(\frac{1}{\gamma(\mathbf{x})^{-1} - 1} \wedge \frac{\tau_j}{\tau} \right)^{-\gamma(\mathbf{x})} - \left(\frac{\tau_j}{\tau} \right)^{-\gamma(\mathbf{x})} \right], \\ \Sigma_{J+1,J+1}(\mathbf{x}) &= \tau^{-1}, \end{aligned}$$

for $(j, l) \in \{1, \dots, J\}^2$ with $j \leq l$.

In particular, Theorem 1 provides the following asymptotic normality result for $\widehat{e}_n(\alpha_n|\mathbf{x})$, which may be seen as a functional version of Theorem 2 in Daouia et al. (2018):

$$\sigma_n^{-1}(\mathbf{x}) \left(\frac{\widehat{e}_n(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3(\mathbf{x})}{1 - 2\gamma(\mathbf{x})} \right).$$

This is done under the assumptions $\gamma(\mathbf{x}) < 1/2$ and $\mathbb{E}[Y_-^{2+\delta}|\mathbf{X} = \mathbf{x}] < \infty$. These essentially guarantee that $|Y|$ has a finite conditional moment of order $2+\delta$, a reasonable assumption in this context of estimating conditional expectiles which generalize the conditional mean. In addition, condition (3.2) ensures that the bias incurred by the use of the functional kernel smoothing technique in this conditional extreme value setting is asymptotically negligible.

As we have already noted, Theorem 1 requires the condition $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ which prevents the estimation of expectiles at arbitrarily large levels α_n . In the next section, we propose to overcome this issue using an extrapolation tool.

4 Extreme functional expectile extrapolation

Our aim in this section is to estimate conditional expectiles at an extreme level β_n such that $n(1 - \beta_n)\pi(\mathbf{x}, h_n) \rightarrow c < \infty$. In extreme quantile estimation, a well-known approach is to use the Weissman extrapolation device, see Theorem 4.3.8 in de Haan and Ferreira (2006): remarking that under $\mathcal{C}_1(\gamma(\mathbf{x}))$,

$$q(\beta_n|\mathbf{x}) = \left(\frac{1 - \alpha_n}{1 - \beta_n} \right)^{\gamma(\mathbf{x})} q(\alpha_n|\mathbf{x})(1 + o(1)), \quad (4.1)$$

the extreme conditional quantile $q(\beta_n|\mathbf{x})$ can therefore be estimated using the intermediate empirical quantile $\hat{q}_n(\alpha_n|\mathbf{x})$ and a suitable estimator of $\gamma(\mathbf{x})$. We refer for instance to Daouia et al. (2011) or Gardes and Girard (2012) for extreme conditional quantile estimators developed on this basis. In the expectile case, a similar approach can be adopted based on the convergence

$$\lim_{\alpha \rightarrow 1} \frac{\bar{F}(e(\alpha|\mathbf{x})|\mathbf{x})}{1 - \alpha} = \gamma(\mathbf{x})^{-1} - 1. \quad (4.2)$$

This is a consequence of the heavy-tailed assumption (2.6): see Bellini et al. (2014, Theorem 11) and more recently Daouia et al. (2018, Proposition 1). This convergence is in fact equivalent to

$$\lim_{\alpha \rightarrow 1} \frac{e(\alpha|\mathbf{x})}{q(\alpha|\mathbf{x})} = (\gamma(\mathbf{x})^{-1} - 1)^{-\gamma(\mathbf{x})}, \quad (4.3)$$

meaning that extreme quantiles and extreme expectiles are asymptotically proportional. As a consequence, Equation (4.1) also holds for expectiles. As proposed in Girard et al. (2021) in the finite-dimensional context, we introduce a Weissman-type estimator of functional extreme expectiles:

$$\hat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x}) = \left(\frac{1 - \alpha_n}{1 - \beta_n} \right)^{\hat{\gamma}_{\alpha_n}(\mathbf{x})} \hat{e}_n(\alpha_n|\mathbf{x}), \quad (4.4)$$

where $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ is any estimator of $\gamma(\mathbf{x})$. Using the second order assumption $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$, we can state the asymptotic behavior of $\hat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})$ under high-level conditions.

Theorem 2. *Assume that (\mathcal{K}) , (\mathcal{L}) and $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ hold with $\gamma(\mathbf{x}) < 1/2$ and $\rho(\mathbf{x}) < 0$. Suppose that there exists $\delta \in (0, 1)$ such that $\mathbb{E}[Y_-^{2+\delta}|\mathbf{X} = \mathbf{x}] < \infty$. Let $\alpha_n \rightarrow 1$, $\beta_n \rightarrow 1$ and $h_n \rightarrow 0$ be such that $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ and $n(1 - \beta_n)\pi(\mathbf{x}, h_n) \rightarrow c < \infty$ as $n \rightarrow \infty$. Assume further that*

$$i) \sigma_n^{-1}(\mathbf{x}) \log(1 - \alpha_n) \omega_{h_n}((1 - \delta)e(\alpha_n|\mathbf{x})|\mathbf{x}) \rightarrow 0,$$

ii) $\sigma_n^{-1}(\mathbf{x})/\log((1-\alpha_n)/(1-\beta_n)) \rightarrow \infty$,

iii) $\sigma_n^{-1}(\mathbf{x})A((1-\alpha_n)^{-1}|\mathbf{x}) \rightarrow \lambda_1 \in \mathbb{R}$ and $\sigma_n^{-1}(\mathbf{x})/q(\alpha_n|\mathbf{x}) \rightarrow \lambda_2 \in \mathbb{R}$.

If, in addition, $\sigma_n^{-1}(\mathbf{x})(\hat{\gamma}_{\alpha_n}(\mathbf{x}) - \gamma(\mathbf{x})) \xrightarrow{d} \Gamma$, where Γ is a nondegenerate distribution, then

$$\frac{\sigma_n^{-1}(\mathbf{x})}{\log((1-\alpha_n)/(1-\beta_n))} \left(\frac{\hat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})}{e(\beta_n|\mathbf{x})} - 1 \right) \xrightarrow{d} \Gamma.$$

The Weissman-type estimator $\hat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})$ is built on the previous estimator $\hat{e}_n(\alpha_n|\mathbf{x})$ of the intermediate expectile, hence the conditions $\gamma(\mathbf{x}) < 1/2$ and $\mathbb{E}[Y_-^{2+\delta}|\mathbf{X} = \mathbf{x}] < \infty$ to ensure the asymptotic normality of $\hat{e}_n(\alpha_n|\mathbf{x})$. In order to relax these assumptions, one may exploit the asymptotic proportionality relationship (4.3) to define another Weissman-type estimator:

$$\tilde{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x}) = \left(\frac{1-\alpha_n}{1-\beta_n} \right)^{\hat{\gamma}(\mathbf{x})} \hat{q}_n(\alpha_n|\mathbf{x}) (\hat{\gamma}(\mathbf{x})^{-1} - 1)^{-\hat{\gamma}(\mathbf{x})}. \quad (4.5)$$

The following high-level result establishes the asymptotic properties of this estimator, under weaker conditions than those of Theorem 2.

Theorem 3. Assume that (\mathcal{K}) and $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ hold with $\gamma(\mathbf{x}) < 1$ and $\rho(\mathbf{x}) < 0$. Suppose that $\mathbb{E}[Y_-|\mathbf{X} = \mathbf{x}] < \infty$. Let $\alpha_n \rightarrow 1$, $\beta_n \rightarrow 1$ and $h_n \rightarrow 0$ be such that $n(1-\alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ and $n(1-\beta_n)\pi(\mathbf{x}, h_n) \rightarrow c < \infty$ as $n \rightarrow \infty$. Assume further that

i) $\sigma_n^{-1}(\mathbf{x})\log(1-\alpha_n)\omega_{h_n}((1-\delta)q(\alpha_n|\mathbf{x})|\mathbf{x}) \rightarrow 0$ for some $\delta > 0$,

ii) $\sigma_n^{-1}(\mathbf{x})/\log((1-\alpha_n)/(1-\beta_n)) \rightarrow \infty$,

iii) $\sigma_n^{-1}(\mathbf{x})A((1-\alpha_n)^{-1}|\mathbf{x}) \rightarrow \lambda_1 \in \mathbb{R}$ and $\sigma_n^{-1}(\mathbf{x})/q(\alpha_n|\mathbf{x}) \rightarrow \lambda_2 \in \mathbb{R}$.

If, in addition, $\sigma_n^{-1}(\mathbf{x})(\hat{\gamma}_{\alpha_n}(\mathbf{x}) - \gamma(\mathbf{x})) \xrightarrow{d} \Gamma$, where Γ is a nondegenerate distribution, then,

$$\frac{\sigma_n^{-1}(\mathbf{x})}{\log((1-\alpha_n)/(1-\beta_n))} \left(\frac{\tilde{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})}{e(\beta_n|\mathbf{x})} - 1 \right) \xrightarrow{d} \Gamma.$$

To illustrate these results, let us consider the case $1-\alpha_n = (n\pi(\mathbf{x}, h_n))^{-a}$ and $1-\beta_n = (n\pi(\mathbf{x}, h_n))^{-b}$ with $0 < a < 1 < b$ and $c = 0$. In this case the two conditions $n(1-\alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ and $n(1-\beta_n)\pi(\mathbf{x}, h_n) \rightarrow c = 0$ turn into the single condition $n\pi(\mathbf{x}, h_n) \rightarrow \infty$. We shall also work under Assumption (2.8) which entails $\omega_{h_n}(y_n|\mathbf{x}) = O(h_n)$ when $y_n \rightarrow \infty$. Recalling that the rate of convergence $\sigma_n^{-1}(\mathbf{x})$ grows like $\sqrt{n(1-\alpha_n)\pi(\mathbf{x}, h_n)} = (n\pi(\mathbf{x}, h_n))^{(1-a)/2}$, we find that the other conditions on α_n , β_n and h_n required in Theorem 2 and Theorem 3 can be simplified as $(n\pi(\mathbf{x}, h_n))^{(1-a)/2} \log(n\pi(\mathbf{x}, h_n))h_n \rightarrow 0$, as $n \rightarrow \infty$, and $a > 1/(1+2\xi)$ with $\xi = \min(\gamma, -\rho)$. We illustrate the implications of this discussion on two classical examples for \mathbf{X} .

Example 1. If \mathbf{X} is a fractal process of order $\tau > 0$ with respect to the semi-metric d , then $\pi(\mathbf{x}, h_n) \sim Ch_n^\tau$, as $h_n \rightarrow 0$, for some $C > 0$, see Definition 13.1 in Ferraty and Vieu (2006). This situation includes the finite p -dimensional case, where $\tau = p$. Here, the conditions $n\pi(\mathbf{x}, h_n) \rightarrow \infty$ and $(n\pi(\mathbf{x}, h_n))^{(1-a)/2} \log(n\pi(\mathbf{x}, h_n))h_n \rightarrow 0$ imply that one can choose $h_n = n^{-\theta}$ with

$$\theta \in \left(\frac{1-a}{2+\tau(1-a)}, \frac{1}{\tau} \right).$$

Let us stress that there is no constraint on b , apart from $b > 1$, and therefore extreme functional expectiles of level converging to 1 at an arbitrarily fast polynomial rate $n^{b(\tau\theta-1)}$, $b > 1$, can then be estimated with (4.4) or (4.5). Letting $a \downarrow 1/(1+2\xi)$, the bias associated with the smoothing (condition *i*) in Theorem 2 and Theorem 3) and the bias associated with the tail approximations (condition *iii*) in Theorem 2 and Theorem 3) are of the same order. It is then possible to balance the squared bias and the variance of the Weissman type estimators by choosing $\theta = \xi/(1+(2+\tau)\xi)$. This yields a polynomial rate of convergence $n^{-\xi/(1+(2+\tau)\xi)}$.

Example 2. If \mathbf{X} is an exponential-type process with orders $\tau_1, \tau_2 \geq 0$ with respect to the semi-metric d , then $\pi(\mathbf{x}, h_n) \sim C \exp(-h_n^{-\tau_1}(-\log h_n)^{\tau_2})$, as $h_n \rightarrow 0$, for some $C > 0$, see Definition 13.4 in Ferraty and Vieu (2006). We refer to Li and Shao (2001) for several examples of Gaussian processes which are exponential-type processes for the metric associated with the supremum norm in functional spaces. Here, we focus on the case $\tau_2 = 0$ for the sake of simplicity, and we let $h_n = (\log n - \kappa \log \log n)^{-1/\tau_1}$ with

$$\kappa \in \left(0, \frac{2}{\tau_1(1-a)} \right).$$

This choice yields $1 - \beta_n \sim C^{-b}(\log n)^{-b\kappa}$. The considered Weissman type estimators are, in this case, limited to the estimation of extreme functional expectiles of (arbitrary) logarithmic level. Finally, balancing the bias and variance by letting $\kappa \uparrow 2/[\tau_1(1-a)]$ yields a logarithmic rate of convergence $(\log n)^{-1/\tau_1}$. This slow rate of convergence is a consequence of the exponentially fast decay of the small ball probability $\pi(\mathbf{x}, h_n)$ as $h_n \rightarrow 0$. One way to avoid this vexing effect is to consider projection-type semi-metrics, for instance using functional Principal Component Analysis; this will be tried out in our finite-sample illustrations.

In view of Theorem 2 and Theorem 3, the asymptotic distribution of the Weissman-type estimators of $\widehat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})$ and $\widehat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})$ is exactly determined by that of the functional tail index estimator $\widehat{\gamma}_{\alpha_n}(\mathbf{x})$ used in its construction. It is therefore essential to provide estimators of $\gamma(\mathbf{x})$ having good performance. The next section is devoted to this problem from an expectile perspective.

5 Functional tail index estimation

The estimation of the tail index is a central problem in the extreme-value literature. In the unconditional heavy-tailed setting, the most popular estimator is arguably that of Hill (1975), and is based on the mean of the log-excesses. This idea of averaging log-excesses was adapted

to the conditional framework in Gardes and Girard (2012): for a given integer $J \geq 2$,

$$\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) = \frac{1}{\log(J!)} \sum_{j=1}^J \log \left(\frac{\hat{q}_n(1 - (1 - \alpha_n)/j|\mathbf{x})}{\hat{q}_n(\alpha_n|\mathbf{x})} \right). \quad (5.1)$$

The asymptotic properties of $\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})$ may be found in Theorem 4 and Corollary 1 of Gardes and Girard (2012) under mild conditions. Our first contribution here is to strengthen their results by making the bias term explicit; this will inform the construction of another of our estimators below. In the Appendix, we provide the joint asymptotic distribution of $\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})$ and estimators of extreme conditional quantiles (see Theorem 7).

Theorem 4. *Assume that (\mathcal{K}) and $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ hold. Let $\alpha_n \rightarrow 1$, $h_n \rightarrow 0$ be such that $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ and*

$$i) \sigma_n^{-1}(\mathbf{x}) \log(1 - \alpha_n) \omega_{h_n}((1 - \delta)q(\alpha_n|\mathbf{x})|\mathbf{x}) \rightarrow 0, \text{ for some } \delta > 0,$$

$$ii) \sigma_n^{-1}(\mathbf{x}) A((1 - \alpha_n)^{-1}|\mathbf{x}) \rightarrow \lambda_1 \in \mathbb{R}.$$

Then,

$$\sigma_n^{-1}(\mathbf{x}) \left(\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x}) \right) \xrightarrow{d} \mathcal{N} \left(\frac{1}{\log(J!)} \left(\sum_{j=1}^J \frac{j^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \right) \lambda_1, \frac{J(J-1)(2J-1)}{6 \log(J!)^2} \gamma^2(\mathbf{x}) \right).$$

As pointed out in Daouia et al. (2011), the choice $J = 9$ leads to the smallest asymptotic variance: in this case the factor $J(J-1)(2J-1)/(6 \log(J!)^2)$ is approximately equal to 1.245.

Second, to define an estimator of the functional tail index $\gamma(\mathbf{x})$ solely based on the use of expectiles, we introduce an alternative device based on the asymptotic relationship (4.2): this relationship implies

$$\gamma(\mathbf{x}) = \lim_{\alpha \rightarrow 1} \frac{1 - \alpha}{1 - \alpha + \bar{F}(e(\alpha|\mathbf{x})|\mathbf{x})}.$$

This leads us to consider the following estimator:

$$\hat{\gamma}_{\alpha_n}(\mathbf{x}) = \frac{1 - \alpha_n}{1 - \alpha_n + \widehat{\bar{F}}_n(\hat{e}_n(\alpha_n|\mathbf{x})|\mathbf{x})}. \quad (5.2)$$

Note that this expectile-based estimator was first introduced in Girard et al. (2021) in the finite-dimensional covariate setting. The next result gives the asymptotic distribution of the functional tail index estimator $\hat{\gamma}_{\alpha_n}(\mathbf{x})$.

Theorem 5. *Assume that (\mathcal{K}) , (\mathcal{L}) and $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ hold with $\gamma(\mathbf{x}) < 1/2$. Suppose that there exists $\delta \in (0, 1)$ such that $\mathbb{E}[Y_-^{2+\delta}|\mathbf{X} = \mathbf{x}] < \infty$. Let $\alpha_n \rightarrow 1$ and $h_n \rightarrow 0$ such that $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ as $n \rightarrow \infty$ and*

$$i) \sigma_n^{-1}(\mathbf{x}) \log(1 - \alpha_n) \omega_{h_n}((1 - \delta)e(\alpha_n|\mathbf{x})|\mathbf{x}) \rightarrow 0,$$

$$ii) \sigma_n^{-1}(\mathbf{x}) A((1 - \alpha_n)^{-1}|\mathbf{x}) \rightarrow \lambda_1 \in \mathbb{R} \text{ and } \sigma_n^{-1}(\mathbf{x})/q(\alpha_n|\mathbf{x}) \rightarrow \lambda_2 \in \mathbb{R}.$$

Then, $\sigma_n^{-1}(\mathbf{x}) (\hat{\gamma}_{\alpha_n}(\mathbf{x}) - \gamma(\mathbf{x}))$ is asymptotically Gaussian with mean

$$b(\mathbf{x}) = \frac{\gamma(\mathbf{x}) (\gamma(\mathbf{x})^{-1} - 1)^{1-\rho(\mathbf{x})}}{1 - \gamma(\mathbf{x}) - \rho(\mathbf{x})} \lambda_1 + \gamma^2(\mathbf{x}) (\gamma(\mathbf{x})^{-1} - 1)^{\gamma(\mathbf{x})+1} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}] \lambda_2$$

and variance

$$V(\mathbf{x}) = \frac{\gamma^3(\mathbf{x})(1 - \gamma(\mathbf{x}))}{1 - 2\gamma(\mathbf{x})}.$$

The main advantage of $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ compared to $\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})$ is its much smaller asymptotic variance for small values of $\gamma(\mathbf{x})$. However, the asymptotic normality of $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ requires the condition $\gamma(\mathbf{x}) < 1/2$, and the variance explodes when $\gamma(\mathbf{x})$ is close to $1/2$. To avoid this drawback, we propose to replace the “direct” estimator of the functional expectile $\hat{e}_n(\alpha_n|\mathbf{x})$ in Equation (5.2) by an “indirect” estimator designed from Equation (4.3). We thus propose a third estimator combining the two previous approaches, having a lower asymptotic variance compared to $\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})$, and asymptotically Gaussian under weaker assumptions compared to $\hat{\gamma}_{\alpha_n}(\mathbf{x})$:

$$\tilde{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) = \frac{1 - \alpha_n}{1 - \alpha_n + \hat{F}_n \left(\left(\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})^{-1} - 1 \right)^{-\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})} \hat{q}_n(\alpha_n|\mathbf{x})|\mathbf{x} \right)}. \quad (5.3)$$

The following result provides the asymptotic normality of this new estimator.

Theorem 6. Assume that (\mathcal{K}) and $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ hold with $\gamma(\mathbf{x}) < 1$. Let $\alpha_n \rightarrow 1$, $h_n \rightarrow 0$ be such that $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ as $n \rightarrow \infty$ and

i) $\sigma_n^{-1}(\mathbf{x}) \log(1 - \alpha_n) \omega_{h_n}((1 - \delta)q(\alpha_n|\mathbf{x}) \wedge e(\alpha_n|\mathbf{x})|\mathbf{x}) \rightarrow 0$, for some $\delta > 0$,

ii) $\sigma_n^{-1}(\mathbf{x}) A((1 - \alpha_n)^{-1}|\mathbf{x}) \rightarrow \lambda_1 \in \mathbb{R}$ and $\sigma_n^{-1}(\mathbf{x})/q(\alpha_n|\mathbf{x}) \rightarrow \lambda_2 \in \mathbb{R}$.

Then, $\sigma_n^{-1}(\mathbf{x}) (\tilde{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x}))$ is asymptotically Gaussian with mean

$$b_J(\mathbf{x}) = \left[h(\gamma(\mathbf{x})) \frac{1}{\log(J!)} \left(\sum_{j=1}^J \frac{j^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \right) - (1 - \gamma(\mathbf{x})) \frac{(\gamma(\mathbf{x})^{-1} - 1)^{-\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \right] \lambda_1$$

and variance $V_J(\mathbf{x}) = \gamma^2(\mathbf{x}) v_J(\mathbf{x})$, with

$$\begin{aligned} v_J(\mathbf{x}) &= h(\gamma(\mathbf{x}))^2 \frac{J(J-1)(2J-1)}{6 \log(J!)^2} - 2 \frac{(1 - \gamma(\mathbf{x}))^2}{(\gamma(\mathbf{x})^{-1} - 1) \vee 1} + (1 - \gamma(\mathbf{x})) \\ &\quad - (1 - \gamma(\mathbf{x})) h(\gamma(\mathbf{x})) \frac{2j^* ((\gamma(\mathbf{x})^{-1} - 1) \wedge 1)^{-1} + (J - j^*)(J - j^* + 1) - 2J}{\log(J!)}, \end{aligned}$$

where $h(t) = 1 - (1 - t) \log(t^{-1} - 1)$ and $j^* \in \{0, \dots, J\}$ is such that $\tau_{j^*} < \gamma(\mathbf{x})^{-1} - 1 \leq \tau_{j^*+1}$, in which we set $\tau_0 = 0$, $\tau_j = (J - j + 1)^{-1}$ for all $j \in \{1, \dots, J\}$, and $\tau_{J+1} = \infty$.

We can notice that the variance term $v_J(\mathbf{x}) = V_J(\mathbf{x})/\gamma^2(\mathbf{x})$ may also be written:

$$\begin{aligned} v_J(\mathbf{x}) &= h(\gamma(\mathbf{x}))^2 \frac{J(J-1)(2J-1)}{6 \log(J!)^2} + (1 - \gamma(\mathbf{x}))^2 \left(\frac{1}{1 - \gamma(\mathbf{x})} - \frac{2}{(\gamma(\mathbf{x})^{-1} - 1) \vee 1} \right) \\ &\quad - 2 \frac{(1 - \gamma(\mathbf{x})) h(\gamma(\mathbf{x}))}{\log(J!)} \sum_{j=1}^{J-1} \left[(\tau_j \vee (\gamma(\mathbf{x})^{-1} - 1))^{-1} - (1 \vee (\gamma(\mathbf{x})^{-1} - 1))^{-1} \right]. \end{aligned}$$

In the Appendix, the proof of Theorem 6 is provided using the above formula. Straightforward calculations lead to the equivalence of these two expressions.

It appears that this new estimator is asymptotically Gaussian under weaker conditions than those imposed on $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ and has only one source of bias (in the sense that its bias does not depend on λ_2 , coming from the condition $\sigma_n^{-1}(\mathbf{x})/q(\alpha_n|\mathbf{x}) \rightarrow \lambda_2 \in \mathbb{R}$). We investigate the asymptotic variances of the three estimators $\hat{\gamma}_{\alpha_n}^{(9)}(\mathbf{x})$, $\tilde{\gamma}_{\alpha_n}^{(9)}(\mathbf{x})$ and $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ in Figure 1. We can see here that $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ seems to have the lowest variance among all three estimators for $\gamma(\mathbf{x}) \in [0, 0.25]$. The new estimator $\tilde{\gamma}_{\alpha_n}^{(9)}(\mathbf{x})$ seems to strike a middle ground, in the sense that it appears to have the lowest variance when the conditional tail index lies between approximately 0.25 and 0.5. In addition, its variance remains relatively stable when $\gamma(\mathbf{x}) \in [0.5, 1]$, even though the quantile-based estimator $\hat{\gamma}_{\alpha_n}^{(9)}(\mathbf{x})$ has the lowest variance in this interval.

Finally, let us note that Theorem 1 opens the door to the design of estimators of the functional tail index based on a combination of several empirical conditional intermediate expectiles. One could for instance consider a Pickands-type estimator (Pickands, 1975) based on the following three empirical conditional intermediate expectiles: $\hat{e}_n(1 - k_n/n|\mathbf{x})$, $\hat{e}_n(1 - k_n/(2n)|\mathbf{x})$ and $\hat{e}_n(1 - k_n/(4n)|\mathbf{x})$ with $k_n = n(1 - \alpha_n)$, which corresponds to considering three intermediate expectile estimators as in Theorem 1 with $\tau_1 = 1/4$, $\tau_2 = 1/2$ and $\tau_3 = 1$.

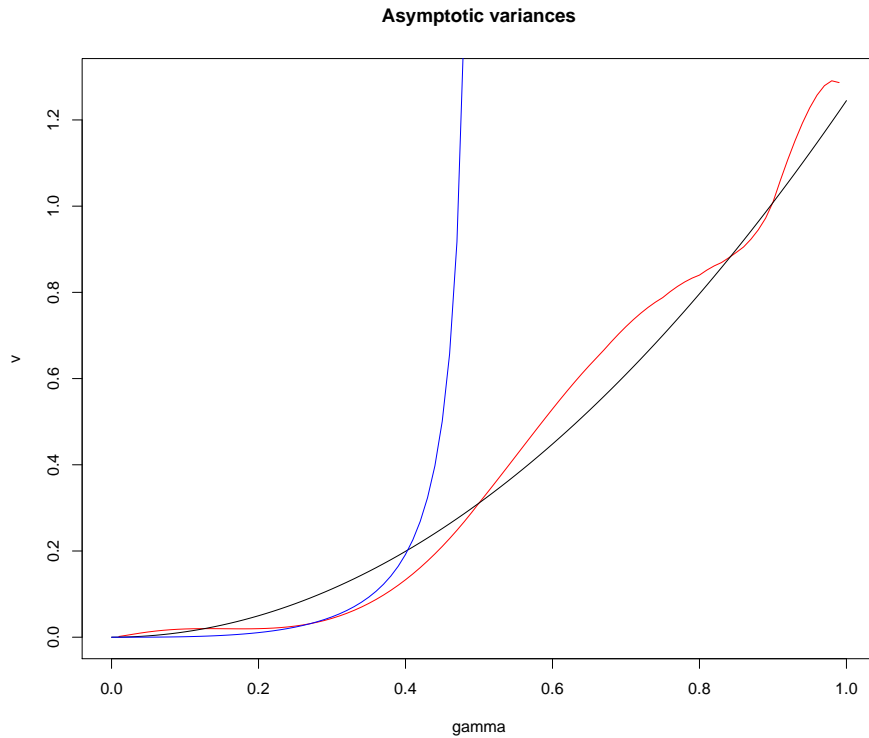


Figure 1: Asymptotic variances of $\hat{\gamma}_{\alpha_n}^{(9)}(\mathbf{x})$ (black curve), $\tilde{\gamma}_{\alpha_n}^{(9)}(\mathbf{x})$ (red curve) and $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ (blue curve) as functions of $\gamma(\mathbf{x}) \in [0, 1]$.

6 Illustration on simulated data

We briefly illustrate here the finite-sample performance of the estimators on $N = 500$ replications of an independent sample of size $n = 500$ (resp. 2,000) from a random pair (\mathbf{X}, Y) , where $\mathbf{X} \in E = L^2[0, 1]$. Here, \mathbf{X} may be three different stochastic processes:

- (\mathcal{X}_1) : A process given by $X_t = \cos(2\pi Zt)$ for all $t \in [0, 1]$ with Z a latent uniform random variable on $[1/4, 1]$, and the functional tail index is given by $\gamma(\mathbf{X}) = (8\|\mathbf{X}\|_2^2 - 3)/2.5$ with $\|\mathbf{X}\|_2^2 = \int_0^1 X_t^2 dt = \frac{1}{2} \left(1 + \frac{\sin(4\pi Z)}{4\pi Z}\right)$. A similar specification is used in Gardes and Girard (2012); in this context $\gamma(\mathbf{X}) \in [0.06, 0.61]$ and d is the standard metric on $L^2[0, 1]$.
- (\mathcal{X}_2) : A process given for all $t \in [0, 1]$ by

$$X_t = \sqrt{2} \sum_{j=1}^{50} \frac{(-1)^{j+1}}{j} Z_j \cos(j\pi t),$$

where the Z_j are i.i.d. uniform on $[-1, 1]$. Such an example is considered in Kato (2012). In this more difficult example, it is suitable to consider a (functional) Principal Component Analysis semi-metric (see Chapter 3 in Ferraty and Vieu (2006) and particularly Section 3.4), defined as

$$d^{q,\text{PCA}}(\mathbf{X}, \mathbf{X}') = \sqrt{\sum_{k=1}^q \left\{ \int_0^1 (X_t - X'_t) \varphi_k(t) dt \right\}^2},$$

where $\varphi_1, \dots, \varphi_q$ are eigenfunctions of the unobserved covariance operator $(s, t) \mapsto \mathbb{E}(X_s X_t)$ associated with its first q eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$. This semi-metric induces a semi-norm by setting $\|\mathbf{X}\|^{q,\text{PCA}} = d^{q,\text{PCA}}(\mathbf{X}, \mathbf{0})$ where $\mathbf{0}$ is the zero function. A numerical study using the R package `fdapace` (see Carroll et al. (2021)) shows that a choice of $q = 2$ leads to an explained variance around 80%. We thus take $q = 2$, estimate φ_1 and φ_2 with the R function `FPCA` in this package, and consider a conditional tail index $\gamma(\mathbf{X}) = 0.25 + 2\|\mathbf{X}\|_2$.

- (\mathcal{X}_3) : A standard Brownian bridge (X_t) on $[0, 1]$. Inspired by the previous model, we keep $\gamma(\mathbf{X}) = 0.25 + 2\|\mathbf{X}\|_2$. Here also, around 80% of the variance is typically explained with the first two eigenfunctions and we take $q = 2$ in this example too.

An overview of the shape of the three covariates considered is proposed in Figure 2 through the simulation of a few realizations. The first kind of process may be found in areas like environmental sciences (representing sound or light waves). The other two are more frequent in finance (see Metwally and Atiya (2002) for an example). Given \mathbf{X} , the random variable Y has, in all cases, a Pareto distribution with c.s.f. $\overline{F}(y|\mathbf{X}) = y^{-1/\gamma(\mathbf{X})}$, $y > 1$.

The aim of this study is to assess the quality of the estimates of the extreme functional expectile $e(\beta_n|\mathbf{x})$ with $\beta_n = 0.995$ ($1 - 2.5/n$ when $n = 500$ and $1 - 10/n$ when $n = 2,000$). For that purpose, the Weissman-type estimators (4.4) and (4.5) are used with the kernel $K(t) = (1.9 - 1.8t)\mathbb{1}_{\{0 \leq t \leq 1\}}$. The tail index is estimated with $\hat{\gamma}_{\alpha_n}^{(9)}(\mathbf{x})$ and $\tilde{\gamma}_{\alpha_n}^{(9)}(\mathbf{x})$ defined in (5.1) and (5.3)

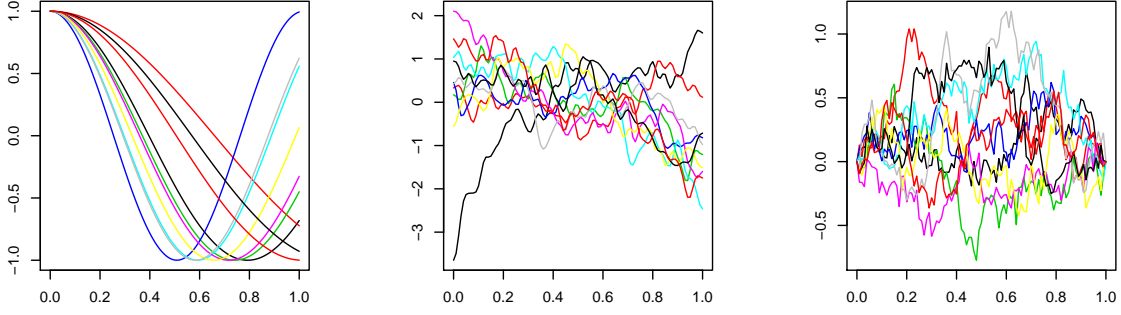


Figure 2: 10 realizations of the covariate \mathbf{X} for models (\mathcal{X}_1) (left), (\mathcal{X}_2) (middle) and (\mathcal{X}_3) (right).

respectively. Throughout this section, the L^2 -norm of a function $(x_t)_{t \in [0,1]}$ is numerically computed through its discretized form $\sqrt{m^{-1} \sum_{k=1}^m x_{t_k}^2}$, with $m = 101$ and $t_k \in \{0, 0.01, 0.02, \dots, 1\}$. Similarly, the PCA semi-norm $\|\cdot\|^{q,\text{PCA}}$ is computed as $\sqrt{\sum_{j=1}^q (m^{-1} \sum_{k=1}^m x_{t_k} \varphi_j(t_k))^2}$.

6.1 Cross-validation procedure

In an extreme regression framework, the question of selecting the bandwidth h_n and the intermediate level α_n has been considered in Gardes and Girard (2012), Daouia et al. (2013) or Girard et al. (2021) for instance. These papers proposed the following type of bandwidth h^* :

$$h^* = \arg \min_{h \in \mathcal{H}} \left\{ \sum_{i=1}^n \sum_{j=1}^n \left(\mathbb{1}_{\{Y_i \geq Y_j\}} - \widehat{F}_{n,-i}(Y_j | \mathbf{X}_i) \right)^2 \right\},$$

where $\widehat{F}_{n,-i}$ is the empirical c.s.f. computed on the $(n-1)$ -sample \mathcal{X}_{-i} where the pair (\mathbf{X}_i, Y_i) has been removed, and \mathcal{H} is a regular grid of size 15 on $[h_{\min}, h_{\max}]$ with $h_{\min} = \max_{1 \leq i \leq n} \min_{i \neq j} d(\mathbf{X}_i, \mathbf{X}_j)$ and $h_{\max} = \max_{1 \leq i, j \leq n} d(\mathbf{X}_i, \mathbf{X}_j)$. This approach returns a value h^* typically around 0.2 for $n = 500$ (0.14 for $n = 2,000$) on our first two models, and tends to return h_{\max} in the third one. Indeed, such a general choice of bandwidth not depending on $\mathbf{x} \in E$ does not take into account the heterogeneity of \mathbf{x} on its support; in particular, it may be thrown off by an isolated point \mathbf{X}_i in the data. Therefore, we propose a choice of h_n related to \mathbf{x} , which ensures that the $n \times h_n \approx 500 \times 0.2 = 100$ nearest neighbors are included in $B(\mathbf{x}, h_n)$:

$$h^*(\mathbf{x}) = \min \left\{ h > 0 \text{ such that } \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i \in B(\mathbf{x}, h)\}} = 100 \right\}.$$

This applies to the case $n = 500$; for $n = 2,000$, we consider $2,000 \times 0.14 = 280$ nearest neighbors instead. Once the bandwidth has been chosen, the intermediate level α^* is selected by minimization of a quadratic score:

$$\alpha^*(\bar{\gamma}) = \arg \min_{\alpha \in (0,1)} \sum_{i=1}^n \left(\bar{\gamma}_{\alpha,-i}(\mathbf{X}_i) - \widehat{\gamma}_{LH}(\mathbf{X}_i) \right)^2.$$

Here, $\bar{\gamma}_{\alpha, -i}$ denotes either $\hat{\gamma}_{\alpha}^{(9)}$ or $\tilde{\gamma}_{\alpha}^{(9)}$ computed on the $(n-1)$ -sample \mathcal{X}_{-i} , and $\hat{\gamma}_{LH}(\mathbf{X}_i)$ is the local Hill estimator computed on the subset of 100 values $\{Y_j^{(i)}, j = 1, \dots, 100\} := \{Y_j : 0 < d(\mathbf{X}_i, \mathbf{X}_j) \leq h^*(\mathbf{x}), j = 1, \dots, n\}$ whose covariate values are within distance $h^*(\mathbf{x})$ of \mathbf{X}_i . This estimator is defined by

$$\hat{\gamma}_{LH}(\mathbf{X}_i) = \frac{1}{k_i} \sum_{j=1}^{k_i} \log \left(\frac{Y_{100-j+1, 100}^{(i)}}{Y_{100-k_i, 100}^{(i)}} \right),$$

where we set $k_i = 20$ (50 for $n = 2,000$) and $Y_{1, 100}^{(i)} \leq \dots \leq Y_{100, 100}^{(i)}$ are the order statistics associated with $\{Y_j^{(i)}, j = 1, \dots, 100\}$, sorted in increasing order. This procedure yields selected values $\alpha^*(\hat{\gamma}^{(9)}) \simeq \alpha^*(\tilde{\gamma}^{(9)})$ typically around 0.84 (0.805 for $n = 2,000$) for (\mathcal{X}_1) , 0.80 (0.815 for $n = 2,000$) for (\mathcal{X}_2) and 0.70 (0.88 for $n = 2,000$) for (\mathcal{X}_3) .

6.2 Results

In example (\mathcal{X}_1) , Figure 3 displays the performance of the conditional tail index and conditional expectile estimators on boxplots, within the space of functions $\{x_t \mapsto \cos(2\pi zt), t \in [0, 1]\}$, depending on the latent variable $z \in [1/4, 1]$. In other words, at each z within a fine grid of points in $[1/4, 1]$, we estimate the conditional tail index and extreme conditional expectile estimators given $\mathbf{x} = (t \mapsto \cos(2\pi zt))$. The numerical results associated with $\hat{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$ and $\tilde{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$ are in line with the theoretical results illustrated in Figure 1: it appears indeed from the length of the represented boxplots that the estimator $\tilde{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$ tends to have a lower variance for lower values of $\gamma(\mathbf{x})$, and to be slightly more accurate on average, than the estimator $\hat{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$. Finally, it seems that both Weissman-type estimators $\hat{e}_{n, \alpha^*}^W(\beta_n | \mathbf{x})$ and $\tilde{e}_{n, \alpha^*}^W(\beta_n | \mathbf{x})$ generally perform similarly at the level 0.995, with an advantage for the direct estimator $\hat{e}_{n, \alpha^*}^W(\beta_n | \mathbf{x})$ in areas where the tail is heavy, and, by contrast, an advantage for the indirect estimator $\tilde{e}_{n, \alpha^*}^W(\beta_n | \mathbf{x})$ in areas where the tail is light.

In examples (\mathcal{X}_2) and (\mathcal{X}_3) , contrary to the first example, we can no longer represent the results as a function of a single univariate latent variable. To visualize our results, we thus propose to simulate a test sample $\mathbf{x}_1, \dots, \mathbf{x}_{100}$ of size 100 (fixed across all samples), and estimate the extreme expectiles $e(\beta_n | \mathbf{x}_1), \dots, e(\beta_n | \mathbf{x}_{100})$. The results are reported in Figure 4 as functions of the L^2 -norms of $\mathbf{x}_i, 1 \leq i \leq 100$. In the area where most of the data are typically concentrated, the proposed estimators provide a good approximation of the conditional tail index and the conditional extreme expectile. However, in the areas where this is not the case (for unusually small or large L^2 -norms), trusting our estimators is of course more difficult. This problem is a classical problem in the nonparametric literature. Note that the empirical curves are smoothed with the R function `loess`, and a smoothing parameter of 0.25. Note also that we computed the bias-reduced versions of our extreme expectile estimators introduced in Girard et al. (2021) (dashed lines) which improve the accuracy of the results:

$$\begin{cases} \hat{e}_{n, \alpha_n}^{W, BR}(\beta_n | \mathbf{x}) &= \hat{e}_{n, \alpha_n}^W(\beta_n | \mathbf{x}) \left(1 - \frac{\hat{m}_n(\mathbf{x}) \hat{\gamma}(\mathbf{x})}{\hat{e}_n(\alpha_n | \mathbf{x})} \right) + \hat{m}_n(\mathbf{x}) \hat{\gamma}(\mathbf{x}), \\ \tilde{e}_{n, \alpha_n}^{W, BR}(\beta_n | \mathbf{x}) &= \tilde{e}_{n, \alpha_n}^W(\beta_n | \mathbf{x}) + \hat{m}_n(\mathbf{x}) \hat{\gamma}(\mathbf{x}). \end{cases} \quad (6.1)$$

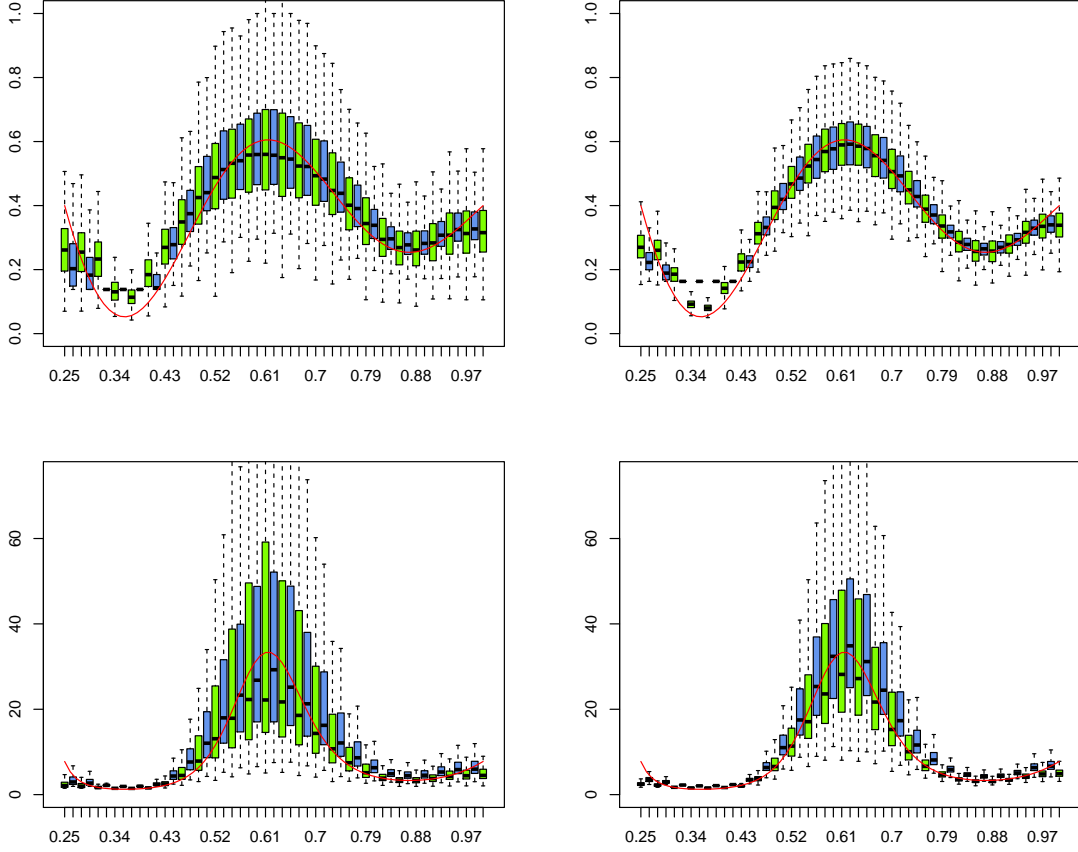


Figure 3: Simulation results on a conditional Pareto distribution with model (\mathcal{X}_1) . Top panel: tail index estimators $\hat{\gamma}_{\alpha^*}^{(9)}(\cdot)$ (green box plots) and $\tilde{\gamma}_{\alpha^*}^{(9)}(\cdot)$ (blue box plots), where at each $z \in [1/4, 1]$ we represent estimated values at $\mathbf{x} = (t \mapsto \cos(2\pi zt))$, for $n = 500$ (left) and $n = 2,000$ (right). The red curve is the true value of $\gamma(\cdot)$. Bottom panel: Indirect (green box plots) and direct (blue box plots) expectile estimators $\hat{e}_{n,\alpha^*}^W(\beta_n|\cdot)$ and $\tilde{e}_{n,\alpha^*}^W(\beta_n|\cdot)$ based on the tail index estimator $\tilde{\gamma}_{\alpha^*}^{(9)}(\cdot)$ for $n = 500$ (left) and $n = 2,000$ (right). The red curve is the true value of $e(\beta_n|\cdot)$, with $\beta_n = 0.995$.

These bias-reduced versions have the same asymptotic behavior as the original $\hat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})$ and $\tilde{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})$ but perform better in finite samples, as illustrated in Figure 4.

7 Real data example

We consider data on the price of the Bitcoin cryptocurrency between January 1, 2012 and January 8, 2018 on the exchange platform Bitstamp¹. We use our estimators to, based on hourly log-returns of the price of Bitcoin on a given day, estimate the level of extreme swings of the price of Bitcoin the next day. More specifically, our goal is to estimate extreme expectile levels of the maximum hourly log-return Y of Bitcoin price based on the information on past

¹This dataset is available at https://github.com/FutureSharks/financial-data/tree/master/pyfinancialdata/data/cryptocurrencies/bitstamp/BTC_USD and is on file with the authors.

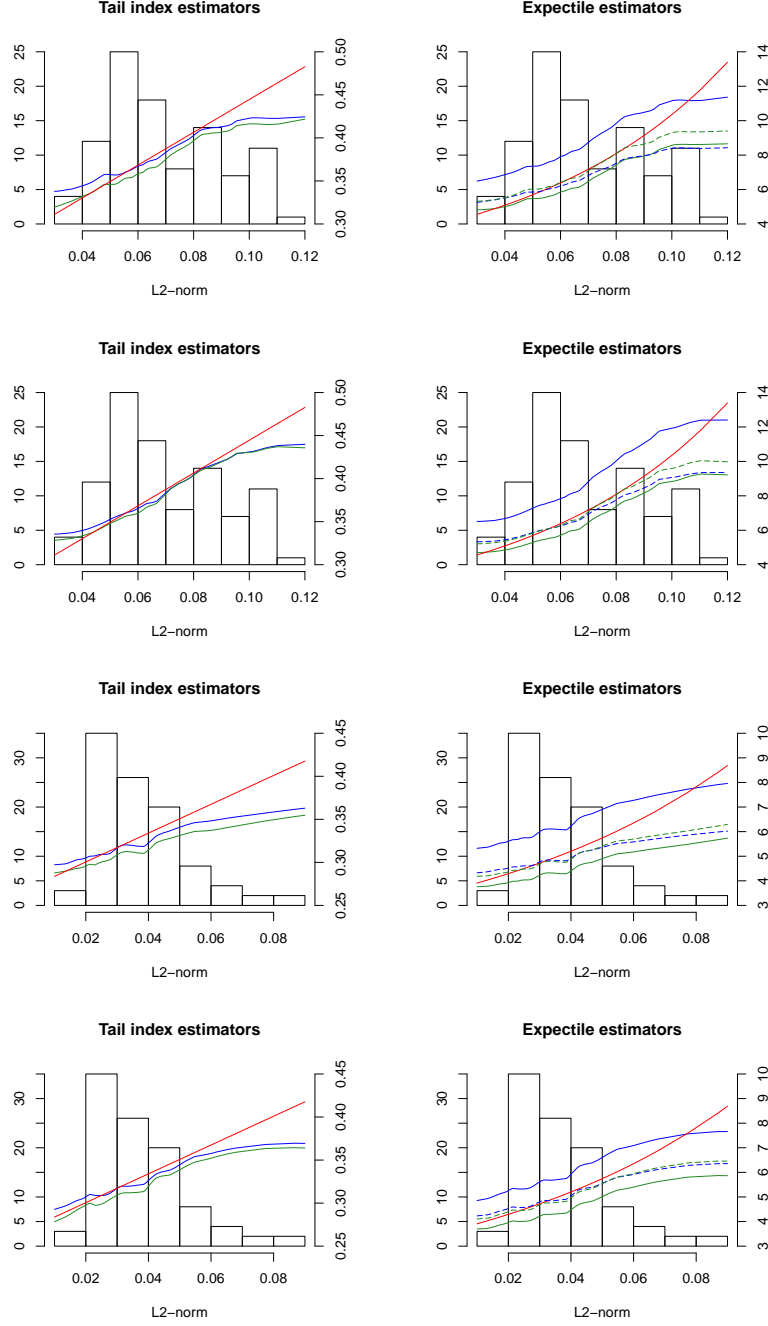


Figure 4: Histograms of the PCA-norm of the test sample $\mathbf{x}_1, \dots, \mathbf{x}_{100}$ of (\mathcal{X}_2) (first two rows) and Brownian bridges (\mathcal{X}_3) (last two rows), and the associated theoretical values of $\gamma(\mathbf{x}_i)$ and $e(\beta_n|\mathbf{x}_i)$ (red curves) as functions of $\|\mathbf{x}_i\|_2$. On the tail index panels, the median values of estimates $\hat{\gamma}_{\alpha^*}^{(9)}(\cdot)$ and $\tilde{\gamma}_{\alpha^*}^{(9)}(\cdot)$ are respectively represented with the green and blue curves. On the expectile panels, the median values of the direct and indirect estimators $\hat{e}_{n,\alpha^*}^W(\beta_n|\cdot)$ and $\tilde{e}_{n,\alpha^*}^W(\beta_n|\cdot)$ (both computed with $\tilde{\gamma}_{\alpha^*}^{(9)}(\cdot)$) are respectively represented with the green and blue curves, and their respective bias-reduced versions $\hat{e}_{n,\alpha^*}^{W,BR}(\beta_n|\cdot)$ and $\tilde{e}_{n,\alpha^*}^{W,BR}(\beta_n|\cdot)$ are the dashed lines. The sample size is $n = 500$ for the first and third rows, and $n = 2,000$ for the others.

price behavior given by its curve \mathbf{X} of hourly log-returns the day before. In order to reduce the obvious temporal dependence existing in the data, we construct our sample of data by keeping a gap of one day between observations: in other words,

- \mathbf{X}_1 and Y_1 are respectively the curve of hourly log-returns on day 1 and maximum hourly log-return on day 2,
- \mathbf{X}_2 and Y_2 are respectively the curve of hourly log-returns on day 3 and maximum hourly log-return on day 4,...

More generally, the i th data point (\mathbf{X}_i, Y_i) is obtained by considering the curve of hourly log-returns on day $2i - 1$ and maximum hourly log-return on day $2i$. Data cleaning leads to the removal of days with missing data, leading us to a final dataset $\mathcal{X} = \{(\mathbf{X}_i, Y_i), i = 1, \dots, n\}$ of size $n = 917$. The left panel of Figure 5 provides an overview of this dataset.

To confirm that the assumption of a heavy right tail is appropriate for the unconditional distribution of Y , the right panel of Figure 5 displays a quantile-quantile plot of the log-excesses $W_{i,k} = \log(Y_{n-i+1,n}/Y_{n-k,n})$, $1 \leq i \leq k$ against the quantiles of the standard exponential distribution. It is indeed known that, if Y is heavy-tailed with tail index γ and if k , corresponding to the sample fraction of the highest data points considered as “extreme”, is chosen sufficiently small, then the $W_{i,k}$ ’s are approximately order statistics of an exponential random variable with mean γ , see for instance Equation (10) in Beirlant et al. (1999). The linear trend appearing in the right panel of Figure 5, for which $k = 100$ (thus taking into account approximately the top 11% of the data) constitutes graphical evidence that the assumption of a heavy right tail makes sense, with an estimated unconditional tail index around $1/2$.

We now discuss the choice of the parameters h and α . In the considered situation where the \mathbf{X}_i ’s are widely scattered (with respect to the L^2 or PCA semi-distances d), it is sensible to use the uniform kernel $K(t) = \mathbb{1}_{\{0 \leq t \leq 1\}}$, and to select a local bandwidth h_i^* for each day i so as to guarantee that enough data will be available for our local estimation. As in the simulation study, we choose here h_i^* so that there are always 100 covariate values in the h_i^* – neighbourhood of \mathbf{X}_i , that is, $\#\{j \in \{1, \dots, n\} : d(\mathbf{X}_i, \mathbf{X}_j) < h_i^*\} = 100$ for all $i \in \{1, \dots, n\}$. We also use the PCA norm $\|\cdot\|^{2,\text{PCA}}$, as this allows to explain an estimated 75% in the variance, see the bottom left panel of Figure 5. The cross-validation procedure introduced in Section 6 (with 25 points used in the local Hill estimator) yields intermediate levels $\alpha^*(\hat{\gamma}^{(9)}) = 1 - 230/917 \simeq 0.749$, $\alpha^*(\tilde{\gamma}^{(9)}) = 1 - 230/917 \simeq 0.749$ and $\alpha^*(\hat{\gamma}) = 1 - 120/917 \simeq 0.869$.

It is then interesting, as a first step in the analysis, to compare the behavior of the three functional tail index estimators $\hat{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$, $\hat{\gamma}_{\alpha^*}(\mathbf{x})$, $\tilde{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$ and the local Hill estimator introduced in Section 6. Note that in this application, we use a (partly) bias-reduced version of $\hat{\gamma}_{\alpha^*}(\mathbf{x})$:

$$\hat{\gamma}_{\alpha^*}^{BR}(\mathbf{x}) = \hat{\gamma}_{\alpha^*}(\mathbf{x}) \left(1 - \frac{\hat{m}_n(\mathbf{x}) (1 - \hat{\gamma}_{\alpha^*}(\mathbf{x}))}{\hat{e}_n(\alpha^*|\mathbf{x})} \right).$$

Indeed, according to Theorem 5, this estimator features a bias term linked to $1/q(\alpha_n|\mathbf{x})$, which can be eliminated in a straightforward way (we use here a procedure similar to that of Corollary 2 in Girard et al. (2021), and we omit the details). The unconditional Hill estimate used in the

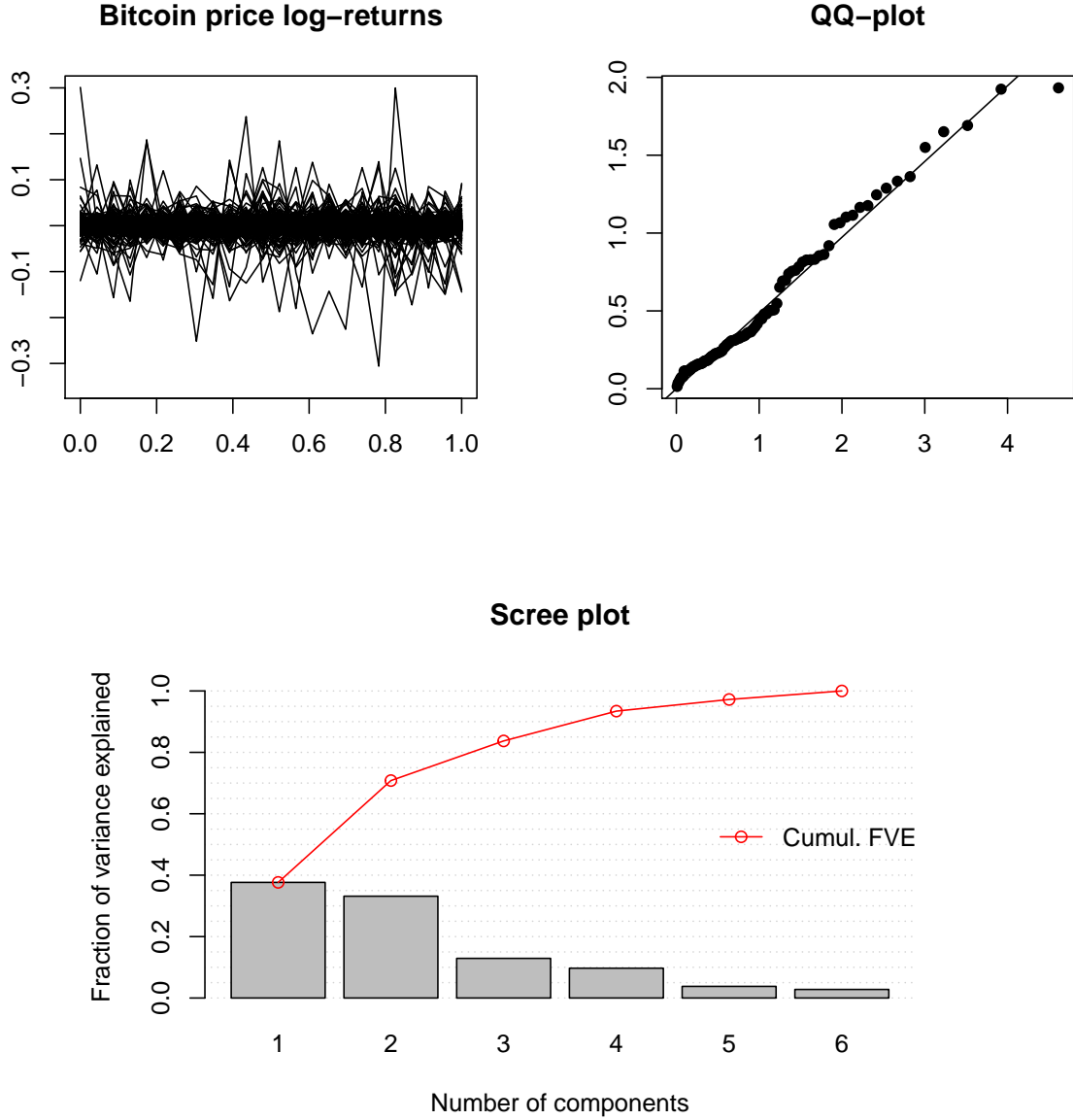


Figure 5: Financial data on the price of Bitcoin. Top left panel: Hourly log-returns of Bitcoin price. Top right panel: Exponential quantile-quantile plot of the log-excesses $\log(Y_{n-i+1,n}/Y_{n-k,n})$ (vertically) versus $\log((k+1)/i)$ (horizontally), for $1 \leq i \leq k$, with $k = 100$. The straight line has a slope computed using the Hill estimator on the top k Y_i 's. Bottom panel: scree plot following the functional PCA procedure, obtained via the FPCA function of the R package `fdapace`.

QQ-plot of Figure 5 is slightly below $1/2$; it is likely, from Theorem 6 and Figure 1, that this is a favorable case for our proposed estimator $\hat{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$, due to our estimator having lowest variance in this area.

Using these selected parameters, our aim is now, for each $i \in \{1, \dots, n\}$, to predict the expectile of level β_n of Y_i , the maximum hourly log-return of the price of Bitcoin on day $i + 1$, given \mathbf{X}_i , the curve of hourly log-returns on day i , using the dataset $\mathcal{X}_{-i} = \{(\mathbf{X}_j, Y_j), j \neq i\}$. We propose to consider the extreme level $\beta_n = 1 - 1/n \approx 0.999$. To give a reasonable idea of how extreme conditional expectiles vary in this functional context, we represent the bias-reduced estimated expectiles $\hat{e}_{n,\alpha^*}^{W,BR}(\beta_n|\mathbf{X}_i)$ (introduced in Equation (6.1) and improving upon (4.4)), versus the PCA norm $\|\mathbf{X}_i\|^{2,\text{PCA}}$ and the daily volatility $s^2(\mathbf{X}_i) = \text{mean}((\mathbf{X}_i - \bar{\mathbf{X}}_i)^2)$ (here $\bar{\mathbf{X}}_i$ denotes the average hourly log-return) in Figures 6 and 7. One can observe that the expectile curves seem to increase with respect to both the norm and the volatility, and that all the proposed estimators return similar values. As a sanity check, we recall that from (4.2), as $n \rightarrow \infty$, one has $n \mathbb{P}(Y > e(1 - 1/n|\mathbf{x})|\mathbf{x}) \rightarrow (\gamma(\mathbf{x})^{-1} - 1)$. In this example, based on a value of the functional tail index around 0.4, we would therefore expect no more than a handful of exceedances above the estimated conditional expectile $\hat{e}_{n,\alpha^*}^{W,BR}(\beta_n|\mathbf{X}_i) = \hat{e}_{n,\alpha^*}^{W,BR}(1 - 1/n|\mathbf{X}_i)$. A graphical inspection yields 2 exceedances for the estimate based on $\hat{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$, and 4 exceedances for the estimates based on either $\hat{\gamma}_{\alpha^*}^{(9)}(\mathbf{x})$ or $\hat{\gamma}_{\alpha^*}^{BR}(\mathbf{x})$. Finally, in order to confirm our previous observation on the trend of extreme expectiles depending on volatility, we plot in Figure 8 our expectile predictions sorted by volatility observed the previous day. The estimated expectile curves generally behave as the observed values of the Y_i ; in comparison with the other estimates, $\hat{\gamma}_{\alpha^*}^{BR}(\mathbf{x})$ seems to lead to lower estimates of the extreme expectiles. All in all, it appears that our extreme expectile method gives a reasonable account of large movements in the price of Bitcoin; they in particular give an indication that a large volatility on a given day leads to potentially more extreme swings in price the following day.

Acknowledgements

The authors acknowledge an anonymous Associate Editor and an anonymous reviewer for their comments that led to an improved version of this paper. This research was supported by the French National Research Agency under the grant ANR-19-CE40-0013/ExtremReg project. S. Girard gratefully acknowledges the support of the Chair Stress Test, Risk Management and Financial Steering, led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas, and the support of the French National Research Agency in the framework of the Investissements d'Avenir programme (ANR-15-IDEX-02). G. Stupfler also acknowledges support from an AXA Research Fund Award on "Mitigating risk in the wake of the COVID-19 pandemic".

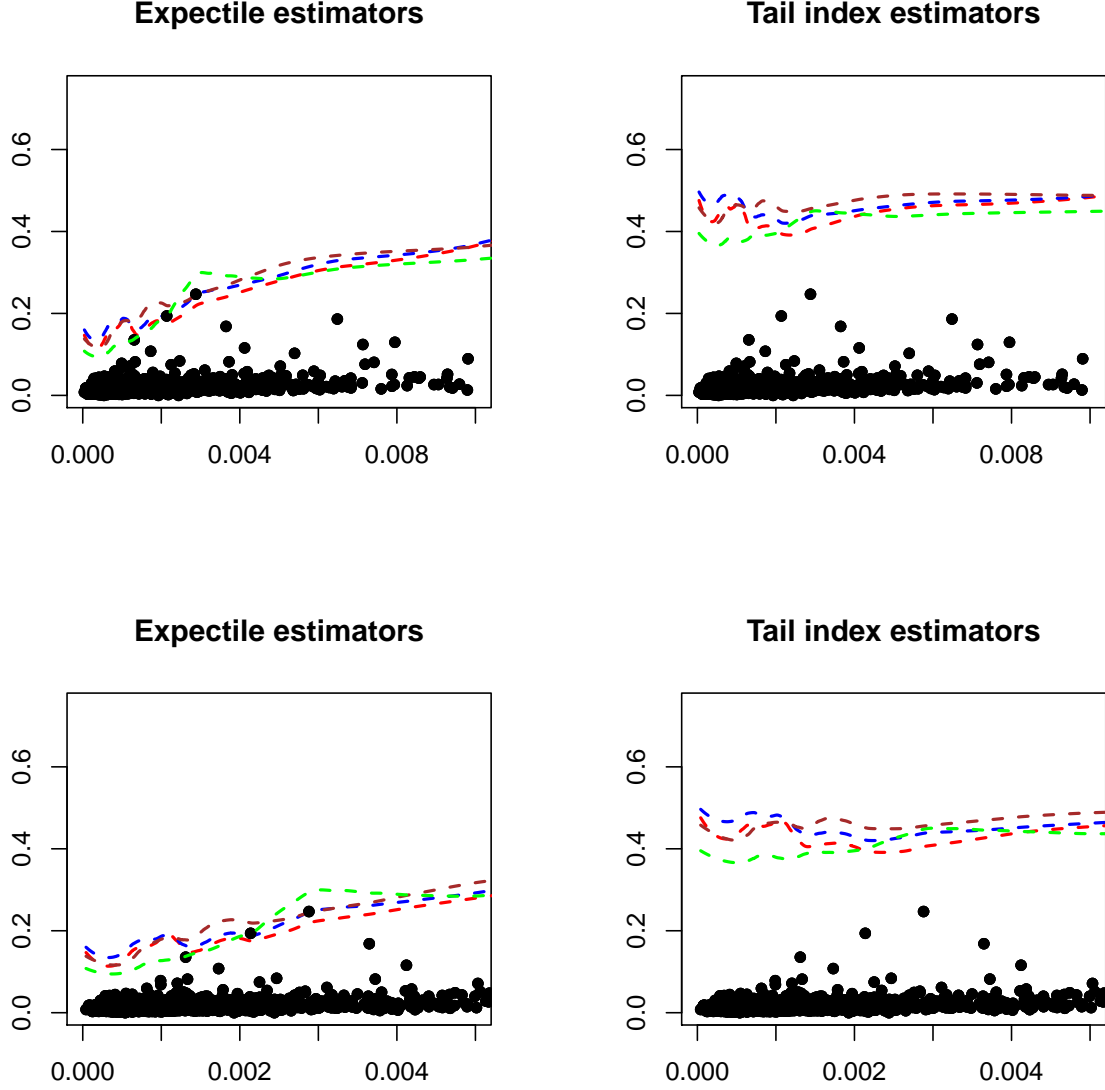


Figure 6: Left panels: Direct expectile estimates $\hat{e}_{n,\alpha^*}^{W,BR}(\beta_n|\cdot)$ based on the four tail index estimators $\hat{\gamma}_{\alpha_n}^{(9)}(\cdot)$ (red), $\hat{\gamma}_{\alpha^*}^{BR}(\cdot)$ (green), $\tilde{\gamma}_{\alpha_n}^{(9)}(\cdot)$ (blue) and a local Hill estimator (brown). Right panels: Tail index estimates, with the same color code. On all panels, curves are smoothed with the R function `loess` (and a smoothing parameter of 0.25). The black dots are the pairs $(\|\mathbf{X}_i\|^{2,\text{PCA}}, Y_i)$. The bottom panels zoom in on the region $\|\mathbf{x}\|^{2,\text{PCA}} \in [0, 0.005]$.

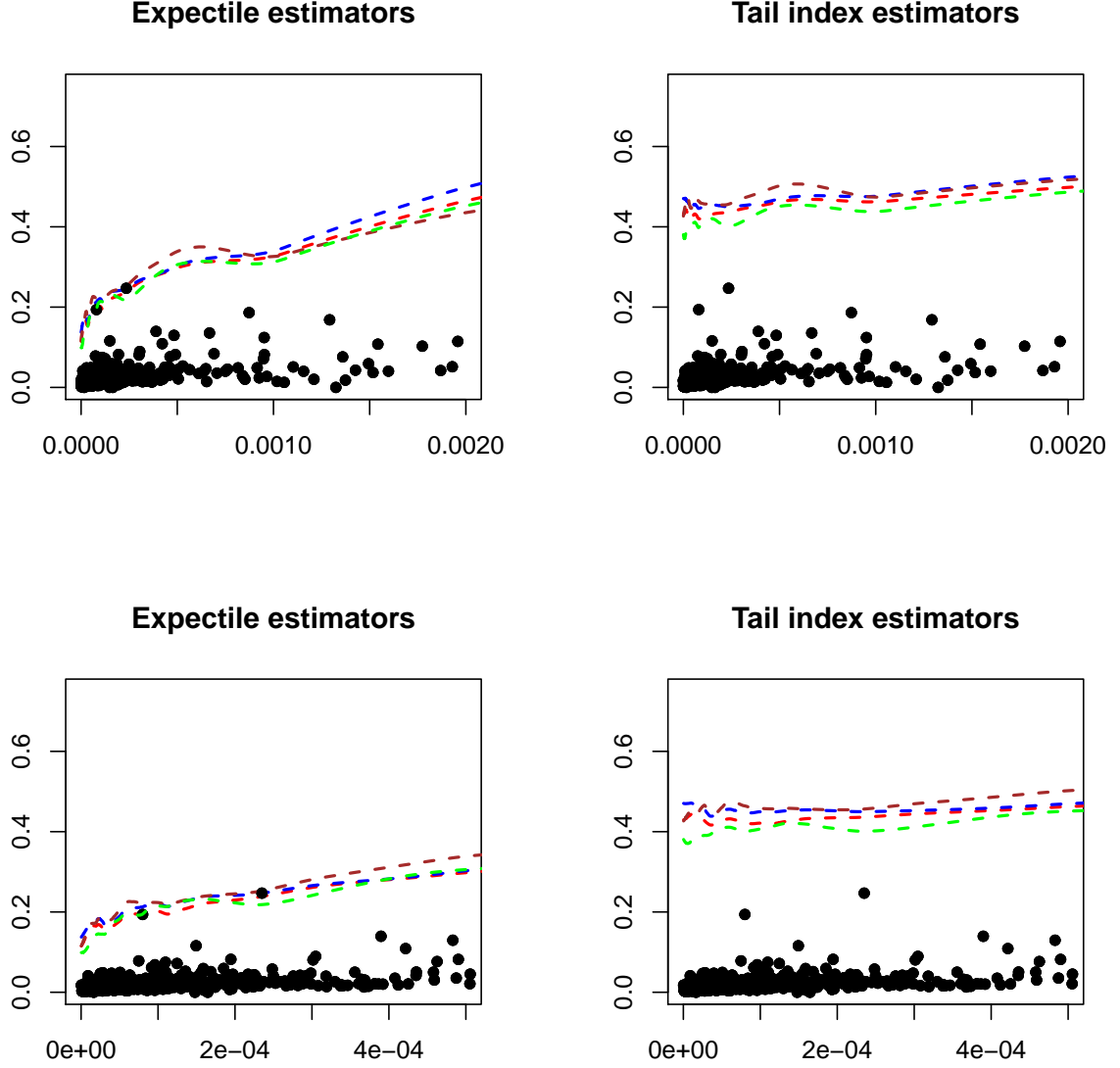


Figure 7: Left panels: Direct expectile estimates $\hat{e}_{n,\alpha^*}^{W,BR}(\beta_n|\cdot)$ based on the four tail index estimators $\hat{\gamma}_{\alpha_n}^{(9)}(\cdot)$ (red), $\hat{\gamma}_{\alpha^*}^{BR}(\cdot)$ (green), $\tilde{\gamma}_{\alpha_n}^{(9)}(\cdot)$ (blue) and a local Hill estimator (brown). Right panels: Tail index estimates, with the same color code. On all panels, curves are smoothed with the R function `loess` (and a smoothing parameter of 0.25). The black dots are the pairs $(s^2(\mathbf{X}_i), Y_i)$. The bottom panels zoom in on the region $s^2(\mathbf{x}) \in [0, 0.0005]$.

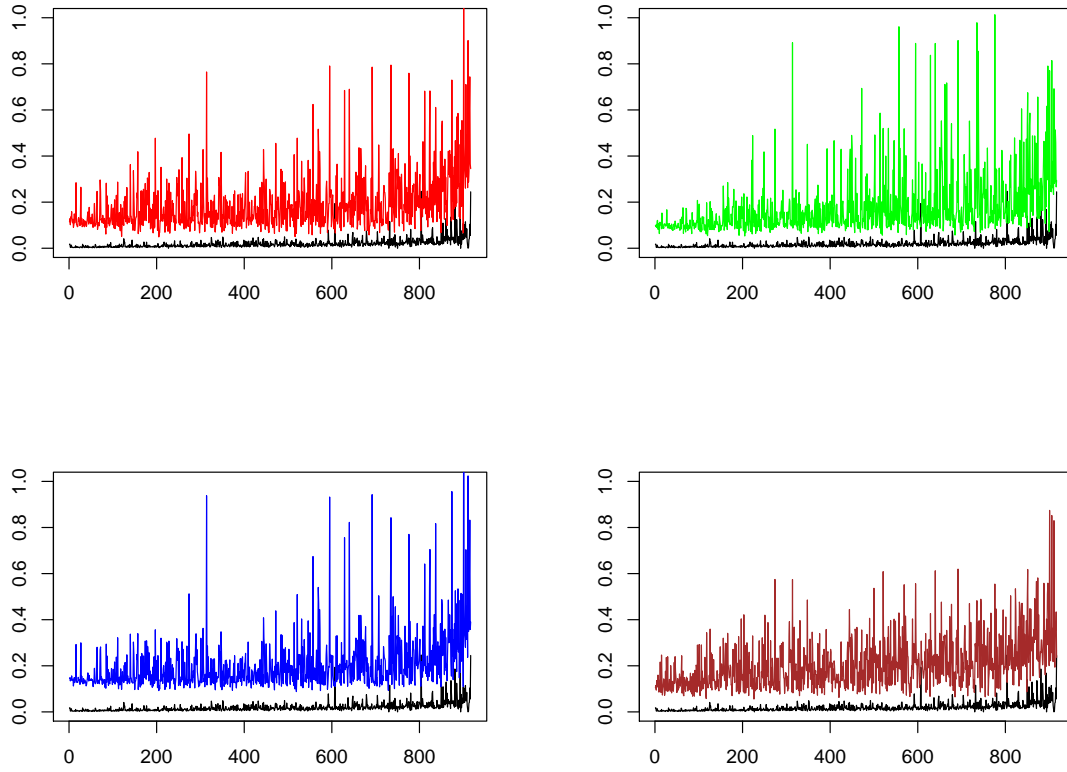


Figure 8: Extreme expectile estimates $\hat{e}_{n,\alpha^*}^{W,BR}(\beta_n|\mathbf{X}_i)$ based on $\hat{\gamma}_{\alpha_n}^{(9)}(\cdot)$ (top left), $\hat{\gamma}_{\alpha^*}^{BR}(\cdot)$ (top right), $\hat{\gamma}_{\alpha_n}^{(9)}(\cdot)$ (bottom left) and a local Hill estimator (bottom right). The black curves represent the pairs (j_i, Y_{j_i}) , where the j_i are such that $s^2(\mathbf{X}_{j_1}) < \dots < s^2(\mathbf{X}_{j_n})$, and the colored curves represent the predicted extreme expectiles $(j_i, \hat{e}_{n,\alpha^*}^{W,BR}(\beta_n|\mathbf{X}_{j_i}))$.

References

- Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, 2(2):177–200.
- Bellini, F., Klar, B., Müller, A., and Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54:41–48.
- Berlinet, A., Gannoun, A., and Matzner-Løber, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles. *Statistics*, 35(2):139–169.
- Billingsley, P. (1995). *Probability and Measure (Third edition)*. John Wiley & Sons.
- Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular Variation*. Cambridge University Press.
- Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika*, 75(4):761–771.
- Cai, J. and Weng, C. (2016). Optimal reinsurance with expectile. *Scandinavian Actuarial Journal*, 2016(7):624–645.
- Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., Ji, H., Mueller, H.-G., and Wang, J.-L. (2021). *fdapace: Functional Data Analysis and Empirical Dynamics*. R package version 0.5.6.
- Chang, J. T. and Pollard, D. (1997). Conditioning as disintegration. *Statistica Neerlandica*, 51(3):287–317.
- Chen, Z. (1996). Conditional L_p -quantiles and their application to the testing of symmetry in non-parametric regression. *Statistics & Probability Letters*, 29(2):107–115.
- Daouia, A., Gardes, L., and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B):2557–2589.
- Daouia, A., Gardes, L., Girard, S., and Lekina, A. (2011). Kernel estimators of extreme level curves. *TEST*, 20:311–333.
- Daouia, A., Girard, S., and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B*, 80(2):263–292.
- Daouia, A., Girard, S., and Stupfler, G. (2019). Extreme M-quantiles as risk measures: from L^1 to L^p optimization. *Bernoulli*, 25(1):264–309.

- Daouia, A., Girard, S., and Stupfler, G. (2020). Tail expectile process and risk assessment. *Bernoulli*, 26(1):531–556.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag, New York.
- El Methni, J., Gardes, L., and Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, 41(4):988–1012.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.
- Ferraty, F., Laksaci, A., and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, 9(1):47–76.
- Ferraty, F., Mas, A., and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Australian & New Zealand Journal of Statistics*, 49(3):267–286.
- Ferraty, F., Rabhi, A., and Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic *El Niño* phenomenon. *Sankhya: The Indian Journal of Statistics*, 67(2):378–398.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer.
- Gannoun, A. (1990). Estimation non paramétrique de la médiane conditionnelle. Médiagramme et méthode du noyau. Application à la prévision des processus. *Publications de l’Institut de Statistique de l’Université de Paris*, XXXV(1):11–22.
- Gardes, L. and Girard, S. (2012). Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics*, 6:1715–1744.
- Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17(1):45–75.
- Gardes, L. and Stupfler, G. (2019). An integrated functional Weissman estimator for conditional extreme quantiles. *REVSTAT: Statistical Journal*, 17(1):109–144.
- Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021). Nonparametric extreme conditional expectile estimation. *Scandinavian Journal of Statistics*. To appear.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174.

- Jones, M. C. (1994). Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters*, 20(2):149–153.
- Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics*, 40(6):3108–3136.
- Koenker, R. and Bassett, G. J. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.
- Krzyzak, A. (1986). The rates of convergence of kernel regression estimates and classification rules. *IEEE Transactions on Information Theory*, 32(5):668–679.
- Kuan, C.-M., Yeh, J.-H., and Hsu, Y.-C. (2009). Assessing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics*, 150(2):261–270.
- Li, W. V. and Shao, Q.-M. (2001). Gaussian processes: inequalities, small ball probabilities and applications. In Rao, C. and Shanbhag, D., editors, *Stochastic processes, Theory and Methods. Handbook of Statistics*, volume 19, pages 533–597. Elsevier, Amsterdam.
- Metwally, S. A. and Atiya, A. F. (2002). Using Brownian bridge for fast simulation of jump-diffusion processes and barrier options. *Journal of Derivatives*, 10(1):43–54.
- Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.
- Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.
- Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics & Probability Letters*, 7(5):407–412.
- Stone, C. J. (1977). Consistent nonparametric regression (with discussion). *The Annals of Statistics*, 5(4):595–645.
- Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics*, 7:2298–2343.
- Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis*, 144:1–24.
- Stute, W. (1986). Conditional empirical processes. *The Annals of Statistics*, 14(2):638–647.
- Taylor, J. W. (2008). Estimating Value at Risk and Expected Shortfall Using Expectiles. *Journal of Financial Econometrics*, 6(2):231–252.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815.
- Ziegel, J. (2016). Coherence and elicibility. *Mathematical Finance*, 26(4):901–918.

A Appendix: proofs

This Appendix is organized as follows: Section A.1 provides some preliminary results useful for the proofs of the main results in Section A.2. Define

$$\begin{aligned}\tilde{\psi}_n^{(k)}(y|\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n (Y_i - y)^k K\left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n}\right) \mathbb{1}_{\{Y_i > y\}} / \mu_K^{(1)}(\mathbf{x}, h_n) \\ \text{and } \tilde{m}_n^{(k)}(y|\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n (Y_i - y)^k K\left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n}\right) / \mu_K^{(1)}(\mathbf{x}, h_n).\end{aligned}$$

The quantities $\tilde{\psi}_n^{(k)}(y|\mathbf{x})$ and $\tilde{m}_n^{(k)}(y|\mathbf{x})$ are the pseudo-estimator counterparts of $\hat{\psi}_n^{(k)}(y|\mathbf{x})$ and $\hat{m}_n^{(k)}(y|\mathbf{x})$ when the smoothed small-ball probability $\mu_K^{(1)}(\mathbf{x}, h_n)$ is assumed to be known.

A.1 Preliminary results

Lemma 1. *Assume that (\mathcal{K}) holds. Then for any $\mathbf{x} \in E$, $h > 0$ and $b > 0$,*

$$c_1^b \pi(\mathbf{x}, h) \leq \mu_K^{(b)}(\mathbf{x}, h) \leq c_2^b \pi(\mathbf{x}, h).$$

In particular, for any $b, b' > 0$ and $\delta > 0$, $[\mu_K^{(b')}(\mathbf{x}, h)]^{1+\delta} / \mu_K^{(b)}(\mathbf{x}, h) \rightarrow 0$ as $h \rightarrow 0$.

Proof. The double inequality is an obvious consequence of assumption (\mathcal{K}) . The desired convergence follows because, by monotone convergence, $\pi(\mathbf{x}, h) \rightarrow \mathbb{P}(\mathbf{X} = \mathbf{x}) = 0$ as $h \rightarrow 0$ (since the probability distribution of \mathbf{X} is non-atomic). \square

Lemma 2. *Assume that (\mathcal{L}) holds. Let $y_n \rightarrow \infty$ and $h_n \rightarrow 0$. Then, uniformly in $\mathbf{x}' \in B(\mathbf{x}, h_n)$,*

$$m^{(1)}(y_n|\mathbf{x}') - m^{(1)}(y_n|\mathbf{x}) = O(h_n) \quad \text{and} \quad m^{(2)}(y_n|\mathbf{x}') - m^{(2)}(y_n|\mathbf{x}) = O(y_n h_n).$$

Proof. Remarking that

$$\begin{aligned}m^{(1)}(y_n|\mathbf{x}') - m^{(1)}(y_n|\mathbf{x}) &= m^{(1)}(0|\mathbf{x}') - m^{(1)}(0|\mathbf{x}), \\ m^{(2)}(y_n|\mathbf{x}') - m^{(2)}(y_n|\mathbf{x}) &= m^{(2)}(0|\mathbf{x}') - m^{(2)}(0|\mathbf{x}) - 2y_n \left(m^{(1)}(0|\mathbf{x}') - m^{(1)}(0|\mathbf{x}) \right),\end{aligned}$$

the result is a straightforward consequence of (\mathcal{L}) . \square

Lemma 3. *Assume (\mathcal{K}) holds. Let $y_n \rightarrow \infty$ and $h_n \rightarrow 0$ be such that $n\pi(\mathbf{x}, h_n) \rightarrow \infty$ as $n \rightarrow \infty$. Then,*

$$\mathbb{E} \left[\tilde{m}_n^{(0)}(y_n|\mathbf{x}) \right] = 1 \quad \text{and} \quad \text{Var} \left[\tilde{m}_n^{(0)}(y_n|\mathbf{x}) \right] = \frac{\mu_K^{(2)}(\mathbf{x}, h_n)}{n\mu_K^{(1)}(\mathbf{x}, h_n)^2} (1 + o(1)).$$

If moreover (\mathcal{L}) holds then

$$\mathbb{E} \left[\tilde{m}_n^{(1)}(y_n|\mathbf{x}) \right] = m^{(1)}(y_n|\mathbf{x}) + O(h_n) \quad \text{and} \quad \text{Var} \left[\tilde{m}_n^{(1)}(y_n|\mathbf{x}) \right] = \frac{\mu_K^{(2)}(\mathbf{x}, h_n)}{n\mu_K^{(1)}(\mathbf{x}, h_n)^2} m^{(2)}(y_n|\mathbf{x}) (1 + o(1)).$$

Proof. By definition,

$$\mathbb{E} \left[\tilde{m}_n^{(0)}(y_n | \mathbf{x}) \right] = \frac{\mathbb{E} \left[K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \right]}{\mu_K^{(1)}(\mathbf{x}, h_n)} = 1,$$

hence the first result. The second one is obtained through a similar calculation:

$$\text{Var} \left[\tilde{m}_n^{(0)}(y_n | \mathbf{x}) \right] = \frac{\mathbb{E} \left[K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \right] - \left(\mathbb{E} \left[K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \right] \right)^2}{n \mu_K^{(1)}(\mathbf{x}, h_n)^2} = \frac{\mu_K^{(2)}(\mathbf{x}, h_n) - \mu_K^{(1)}(\mathbf{x}, h_n)^2}{n \mu_K^{(1)}(\mathbf{x}, h_n)^2}.$$

Using Lemma 1 proves the result. Then,

$$\mathbb{E} \left[\tilde{m}_n^{(1)}(y_n | \mathbf{x}) \right] = \frac{\mathbb{E} \left[m^{(1)}(y_n | \mathbf{X}) K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \right]}{\mu_K^{(1)}(\mathbf{x}, h_n)} = m^{(1)}(y_n | \mathbf{x}) + O(h_n)$$

from Lemma 2. The third result is proved. The fourth one is obtained through a similar calculation:

$$\begin{aligned} \text{Var} \left[\tilde{m}_n^{(1)}(y_n | \mathbf{x}) \right] &= \\ \frac{\mathbb{E} \left[\left(m^{(2)}(y_n | \mathbf{X}) - m^{(2)}(y_n | \mathbf{x}) \right) K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right)^2 \right] + m^{(2)}(y_n | \mathbf{x}) \mu_K^{(2)}(\mathbf{x}, h_n)}{n \mu_K^{(1)}(\mathbf{x}, h_n)^2} &- \frac{m^{(1)}(y_n | \mathbf{x})^2}{n} (1 + o(1)). \end{aligned}$$

Clearly $m^{(1)}(y_n | \mathbf{x}) = y_n(1 + o(1))$ and $m^{(2)}(y_n | \mathbf{x}) = y_n^2(1 + o(1))$. Combining the results of Lemmas 1 and 2 concludes the proof. \square

In the next lemma and throughout, $\mathcal{B}(\cdot, \cdot)$ denotes the Beta function. The proofs of Lemmas 4 and 5 are omitted: they are straightforward adaptations of corresponding results (Lemma 2i) and Lemma 3 respectively) in Girard et al. (2021), whose proofs are written based on a multivariate covariate (their finite-dimensional nature not playing any role whatsoever).

Lemma 4. Suppose $\mathcal{C}_1(\gamma(\mathbf{x}))$ holds. Then, for all $a \in [0, 1/\gamma(\mathbf{x}))$,

$$\psi^{(a)}(y | \mathbf{x}) = \frac{\mathcal{B}(a + 1, \gamma(\mathbf{x})^{-1} - a)}{\gamma(\mathbf{x})} y^a \bar{F}(y | \mathbf{x}) (1 + o(1)) \text{ as } y \rightarrow \infty.$$

Lemma 5. Assume $\mathcal{C}_1(\gamma(\mathbf{x}))$ holds. Let $y_n \rightarrow \infty$ and $h_n \rightarrow 0$ such that $\omega_{h_n}(y_n | \mathbf{x}) \log(y_n) \rightarrow 0$. Then, uniformly in $\mathbf{x}' \in B(\mathbf{x}, h_n)$ and for all $a \in [0, 1/\gamma(\mathbf{x}))$,

$$\frac{\psi^{(a)}(y_n | \mathbf{x}')}{\psi^{(a)}(y_n | \mathbf{x})} - 1 = O(\omega_{h_n}(y_n | \mathbf{x}) \log(y_n)).$$

Lemma 6. Assume that (\mathcal{K}) and $\mathcal{C}_1(\gamma(\mathbf{x}))$ hold. Let $y_n \rightarrow \infty$ and $h_n \rightarrow 0$ be such that $\omega_{h_n}(y_n | \mathbf{x}) \log(y_n) \rightarrow 0$ as $n \rightarrow \infty$. Then, for all $a \in [0, 1/\gamma(\mathbf{x}))$ and $b > 0$,

$$\mathbb{E} \left[(Y - y_n)^a K^b \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_n\}} \right] = \psi^{(a)}(y_n | \mathbf{x}) \mu_K^{(b)}(\mathbf{x}, h_n) (1 + O(\omega_{h_n}(y_n | \mathbf{x}) \log(y_n))).$$

In particular,

$$\forall a \in [0, 1/\gamma(\mathbf{x})), \quad \mathbb{E} \left[\tilde{\psi}_n^{(a)}(y_n | \mathbf{x}) \right] = \psi^{(a)}(y_n | \mathbf{x}) (1 + O(\omega_{h_n}(y_n | \mathbf{x}) \log(y_n)))$$

$$\text{and } \forall a \in [0, 1/(2\gamma(\mathbf{x}))), \quad \text{Var} \left[\tilde{\psi}_n^{(a)}(y_n | \mathbf{x}) \right] = \psi^{(2a)}(y_n | \mathbf{x}) \left(\frac{\mu_K^{(2)}(\mathbf{x}, h_n)}{n \mu_K^{(1)}(\mathbf{x}, h_n)^2} \right) (1 + o(1)).$$

Proof. The first identity is immediate by noting that

$$\mathbb{E} \left[(Y - y_n)^a K^b \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_n\}} \right] = \mathbb{E} \left[\psi^{(a)}(y_n | \mathbf{X}) K^b \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \right]$$

and applying Lemma 5. The second one follows immediately because

$$\mathbb{E} \left[\tilde{\psi}_n^{(a)}(y_n | \mathbf{x}) \right] = \frac{1}{\mu_K^{(1)}(\mathbf{x}, h_n)} \mathbb{E} \left[(Y - y_n)^a K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_n\}} \right].$$

The third result can be obtained by similar calculations, which yield

$$\mathbb{V}\text{ar} \left[\tilde{\psi}_n^{(a)}(y_n | \mathbf{x}) \right] = \frac{\psi^{(2a)}(y_n | \mathbf{x}) \mu_K^{(2)}(\mathbf{x}, h_n)}{n \mu_K^{(1)}(\mathbf{x}, h_n)^2} (1 + o(1)) - \frac{\psi^{(a)}(y_n | \mathbf{x})^2}{n} (1 + o(1)),$$

Combining the results of Lemmas 1 and 4 completes the proof. \square

Lemma 7. Assume (\mathcal{K}) and $\mathcal{C}_1(\gamma(\mathbf{x}))$ hold with $\gamma(\mathbf{x}) < 1/2$. Suppose also that there exists $\delta \in (0, 1)$ such that $\mathbb{E}[Y_-^{2+\delta} | \mathbf{X} = \mathbf{x}] < \infty$. Let $y_n \rightarrow \infty$, $h_n \rightarrow 0$ and $z_n = \theta y_n(1 + o(1))$, with $\theta > 0$, such that $n\bar{F}(y_n | \mathbf{x})\pi(\mathbf{x}, h_n) \rightarrow \infty$ and

$$\sqrt{n\bar{F}(y_n | \mathbf{x})\pi(\mathbf{x}, h_n) \log(y_n) \omega_{h_n}} ((1 - \delta)(\theta \wedge 1)y_n | \mathbf{x}) \rightarrow 0.$$

If, for all $j \in \{1, \dots, J\}$, $y_{n,j} = \tau_j^{-\gamma(\mathbf{x})} y_n(1 + o(1))$ with $0 < \tau_1 < \tau_2 < \dots < \tau_J \leq 1$, then

$$\sqrt{n\bar{F}(y_n | \mathbf{x}) \frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)}} \left\{ \left(\frac{\tilde{\psi}_n^{(1)}(y_{n,j} | \mathbf{x})}{\psi^{(1)}(y_{n,j} | \mathbf{x})} - 1 \right)_{1 \leq j \leq J}, \left(\frac{\tilde{\psi}_n^{(0)}(z_n | \mathbf{x})}{\psi^{(0)}(z_n | \mathbf{x})} - 1 \right) \right\} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{J+1}, \mathbf{V}(\mathbf{x}))$$

where $\mathbf{V}(\mathbf{x})$ is the symmetric matrix having entries:

$$\begin{cases} V_{j,l}(\mathbf{x}) = \frac{1 - \gamma(\mathbf{x})}{\gamma(\mathbf{x})} \tau_l^{-1} \left[\frac{1}{1 - 2\gamma(\mathbf{x})} \left(\frac{\tau_j}{\tau_l} \right)^{-\gamma(\mathbf{x})} - 1 \right], \quad (j, l) \in \{1, \dots, J\}^2, \quad j \leq l, \\ V_{j,J+1}(\mathbf{x}) = \frac{\gamma(\mathbf{x})\theta^{1/\gamma(\mathbf{x})} \left(\theta \vee \tau_j^{-\gamma(\mathbf{x})} \right)^{1-1/\gamma(\mathbf{x})} + (1 - \gamma(\mathbf{x})) \left(\theta \vee \tau_j^{-\gamma(\mathbf{x})} - \tau_j^{-\gamma(\mathbf{x})} \right)}{\gamma(\mathbf{x})\tau_j^{1-\gamma(\mathbf{x})}}, \quad j \in \{1, \dots, J\}, \\ V_{J+1,J+1}(\mathbf{x}) = \theta^{1/\gamma(\mathbf{x})}. \end{cases}$$

If in addition (\mathcal{L}) holds, then

$$\sqrt{n\bar{F}(y_n | \mathbf{x}) \frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)}} \left\{ \left(\frac{\widehat{E}_n(y_{n,j} | \mathbf{x})}{\bar{E}(y_{n,j} | \mathbf{x})} - 1 \right)_{1 \leq j \leq J}, \left(\frac{\widehat{F}_n(z_n | \mathbf{x})}{\bar{F}(z_n | \mathbf{x})} - 1 \right) \right\} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{J+1}, \mathbf{V}(\mathbf{x})).$$

Proof. Let $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J, \beta_{J+1}) \in \mathbb{R}^{J+1}$ and $v_n(\mathbf{x}) = \sqrt{n\bar{F}(y_n | \mathbf{x}) \frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)}}$. One has:

$$\begin{aligned} & v_n(\mathbf{x}) \left\{ \sum_{j=1}^J \beta_j \left(\frac{\tilde{\psi}_n^{(1)}(y_{n,j} | \mathbf{x})}{\psi^{(1)}(y_{n,j} | \mathbf{x})} - 1 \right) + \beta_{J+1} \left(\frac{\tilde{\psi}_n^{(0)}(z_n | \mathbf{x})}{\psi^{(0)}(z_n | \mathbf{x})} - 1 \right) \right\} \\ &= v_n(\mathbf{x}) \left\{ \sum_{j=1}^J \beta_j \left(\frac{\tilde{\psi}_n^{(1)}(y_{n,j} | \mathbf{x}) - \mathbb{E}[\tilde{\psi}_n^{(1)}(y_{n,j} | \mathbf{x})]}{\psi^{(1)}(y_{n,j} | \mathbf{x})} \right) + \beta_{J+1} \left(\frac{\tilde{\psi}_n^{(0)}(z_n | \mathbf{x}) - \mathbb{E}[\tilde{\psi}_n^{(0)}(z_n | \mathbf{x})]}{\psi^{(0)}(z_n | \mathbf{x})} \right) \right\} \\ &+ v_n(\mathbf{x}) \left\{ \sum_{j=1}^J \beta_j \left(\frac{\mathbb{E}[\tilde{\psi}_n^{(1)}(y_{n,j} | \mathbf{x})]}{\psi^{(1)}(y_{n,j} | \mathbf{x})} - 1 \right) + \beta_{J+1} \left(\frac{\mathbb{E}[\tilde{\psi}_n^{(0)}(z_n | \mathbf{x})]}{\psi^{(0)}(z_n | \mathbf{x})} - 1 \right) \right\}. \end{aligned}$$

According to Lemma 6,

$$\begin{cases} \mathbb{E} \left[\tilde{\psi}_n^{(0)}(z_n|\mathbf{x}) \right] = \psi^{(0)}(z_n|\mathbf{x})(1 + O(\omega_{h_n}(z_n|\mathbf{x}) \log(z_n))), \\ \mathbb{E} \left[\tilde{\psi}_n^{(1)}(y_{n,j}|\mathbf{x}) \right] = \psi^{(1)}(y_{n,j}|\mathbf{x})(1 + O(\omega_{h_n}(y_{n,j}|\mathbf{x}) \log(y_{n,j}))). \end{cases}$$

Noticing that for n large enough, $y_{n,j} > y_n(1 - \delta)$, we obtain $\omega_{h_n}(y_{n,j}|\mathbf{x}) \leq \omega_{h_n}((1 - \delta)y_n|\mathbf{x})$, $\forall j \in \{1, \dots, J\}$. Similarly $z_n > \theta y_n(1 - \delta)$ and thus $\omega_{h_n}(z_n|\mathbf{x}) \leq \omega_{h_n}((1 - \delta)\theta y_n|\mathbf{x})$. Moreover, $\log(y_{n,j}) = O(\log(y_n))$ for any $j \in \{1, \dots, J\}$ and $\log(z_n) = O(\log(y_n))$. Therefore

$$v_n(\mathbf{x}) \left\{ \sum_{j=1}^J \beta_j \left(\frac{\mathbb{E} \left[\tilde{\psi}_n^{(1)}(y_{n,j}|\mathbf{x}) \right]}{\psi^{(1)}(y_{n,j}|\mathbf{x})} - 1 \right) + \beta_{J+1} \left(\frac{\mathbb{E} \left[\tilde{\psi}_n^{(0)}(z_n|\mathbf{x}) \right]}{\psi^{(0)}(z_n|\mathbf{x})} - 1 \right) \right\} = o(1).$$

We now focus on the asymptotic distribution of

$$Z_n = v_n(\mathbf{x}) \left\{ \sum_{j=1}^J \beta_j \left(\frac{\tilde{\psi}_n^{(1)}(y_{n,j}|\mathbf{x}) - \mathbb{E} \left[\tilde{\psi}_n^{(1)}(y_{n,j}|\mathbf{x}) \right]}{\psi^{(1)}(y_{n,j}|\mathbf{x})} \right) + \beta_{J+1} \left(\frac{\tilde{\psi}_n^{(0)}(z_n|\mathbf{x}) - \mathbb{E} \left[\tilde{\psi}_n^{(0)}(z_n|\mathbf{x}) \right]}{\psi^{(0)}(z_n|\mathbf{x})} \right) \right\}.$$

We clearly have $\mathbb{E}[Z_n] = 0$. In addition, $\text{Var}[Z_n] = \bar{F}(y_n|\mathbf{x}) \boldsymbol{\beta}^\top \mathbf{B}^{(n)} \boldsymbol{\beta}$, where $\mathbf{B}^{(n)}$ is the symmetric matrix-valued sequence having entries

$$\begin{cases} B_{j,l}^{(n)} = \frac{\text{cov} \left((Y - y_{n,j}) K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}}, (Y - y_{n,l}) K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,l}\}} \right)}{\mu_K^{(2)}(\mathbf{x}, h_n) \psi^{(1)}(y_{n,j}|\mathbf{x}) \psi^{(1)}(y_{n,l}|\mathbf{x})}, \quad j \leq l \in \{1, \dots, J\}, \\ B_{j,J+1}^{(n)} = \frac{\text{cov} \left((Y - y_{n,j}) K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}}, K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > z_n\}} \right)}{\mu_K^{(2)}(\mathbf{x}, h_n) \psi^{(1)}(y_{n,j}|\mathbf{x}) \psi^{(0)}(z_n|\mathbf{x})}, \quad j \in \{1, \dots, J\}, \\ B_{J+1,J+1}^{(n)} = \frac{\text{Var} \left[K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > z_n\}} \right]}{\mu_K^{(2)}(\mathbf{x}, h_n) [\psi^{(0)}(z_n|\mathbf{x})]^2}. \end{cases}$$

Let us first focus, for $j \leq l$, on the term $A_{j,l}^{(n)} = B_{j,l}^{(n)} \psi^{(1)}(y_{n,j}|\mathbf{x}) \psi^{(1)}(y_{n,l}|\mathbf{x})$. Since $y_{n,j} > y_{n,l}$ for n large enough, we find:

$$\begin{aligned} A_{j,l}^{(n)} &= \frac{1}{\mu_K^{(2)}(\mathbf{x}, h_n)} \mathbb{E} \left[(Y - y_{n,j})(Y - y_{n,l}) K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}} \right] \\ &\quad - \frac{1}{\mu_K^{(2)}(\mathbf{x}, h_n)} \mathbb{E} \left[(Y - y_{n,j}) K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}} \right] \mathbb{E} \left[(Y - y_{n,l}) K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,l}\}} \right]. \end{aligned}$$

According to Lemma 6, the second term is equal to $\mu_K^{(1)}(\mathbf{x}, h_n)^2 \psi^{(1)}(y_{n,j}|\mathbf{x}) \psi^{(1)}(y_{n,l}|\mathbf{x}) / \mu_K^{(2)}(\mathbf{x}, h_n) (1 + o(1))$. It thus remains to focus on the first term of $A_{j,l}^{(n)}$ which we rewrite as

$$\frac{1}{\mu_K^{(2)}(\mathbf{x}, h_n)} \left(\mathbb{E} \left[(Y - y_{n,j})^2 K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}} \right] + (y_{n,j} - y_{n,l}) \mathbb{E} \left[(Y - y_{n,j}) K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}} \right] \right).$$

Using Lemma 6, we get

$$\frac{1}{\mu_K^{(2)}(\mathbf{x}, h_n)} \mathbb{E} \left[(Y - y_{n,j})^2 K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}} \right] = \psi^{(2)}(y_{n,j}|\mathbf{x}) (1 + o(1))$$

and

$$\frac{(y_{n,j} - y_{n,l})}{\mu_K^{(2)}(\mathbf{x}, h_n)} \mathbb{E} \left[(Y - y_{n,j}) K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}} \right] = \left(\tau_j^{-\gamma(\mathbf{x})} - \tau_l^{-\gamma(\mathbf{x})} \right) \psi^{(1)}(y_{n,j}|\mathbf{x}) y_n (1 + o(1)).$$

Besides, Lemma 4 and the regular variation property of $\bar{F}(\cdot|\mathbf{x})$ provide, for any j ,

$$\begin{cases} \psi^{(1)}(y_{n,j}|\mathbf{x}) = \frac{\gamma(\mathbf{x})}{1 - \gamma(\mathbf{x})} \tau_j^{-\gamma(\mathbf{x})} y_n \tau_j \bar{F}(y_n|\mathbf{x}) (1 + o(1)), \\ \psi^{(2)}(y_{n,j}|\mathbf{x}) = \frac{2\gamma(\mathbf{x})^2}{(1 - 2\gamma(\mathbf{x}))(1 - \gamma(\mathbf{x}))} \tau_j^{-2\gamma(\mathbf{x})} y_n^2 \tau_j \bar{F}(y_n|\mathbf{x}) (1 + o(1)). \end{cases}$$

Consequently the first term in $A_{j,l}^{(n)}$ is of order $y_n^2 \bar{F}(y_n|\mathbf{x})$ and thus dominates, by Lemma 1. Straightforward calculations then yield:

$$B_{j,l}^{(n)} = \frac{1 - \gamma(\mathbf{x})}{\gamma(\mathbf{x})} \tau_l^{-1} \left[\frac{1}{1 - 2\gamma(\mathbf{x})} \frac{\tau_j^{-\gamma(\mathbf{x})}}{\tau_l^{-\gamma(\mathbf{x})}} - 1 \right] \frac{1}{\bar{F}(y_n|\mathbf{x})} (1 + o(1)).$$

We now deal with $B_{j,J+1}^{(n)}$ for $j \in \{1, \dots, J\}$, which can be rewritten as

$$\frac{\mathbb{E} \left[(Y - y_{n,j}) K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j} \vee z_n\}} \right] - \mathbb{E} \left[(Y - y_{n,j}) K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j}\}} \right] \mathbb{E} \left[K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > z_n\}} \right]}{\mu_K^{(2)}(\mathbf{x}, h_n) \psi^{(1)}(y_{n,j}|\mathbf{x}) \psi^{(0)}(z_n|\mathbf{x})}.$$

Using Lemma 6, the second term in the numerator equals $\mu_K^{(1)}(\mathbf{x}, h_n)^2 \psi^{(1)}(y_{n,j}|\mathbf{x}) \psi^{(0)}(z_n|\mathbf{x}) (1 + o(1))$ and the first term can be rewritten

$$\begin{aligned} & \mathbb{E} \left[(Y - y_{n,j}) K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > y_{n,j} \vee z_n\}} \right] \\ &= \mu_K^{(2)}(\mathbf{x}, h_n) \left[\psi^{(1)}(y_{n,j} \vee z_n|\mathbf{x}) + (y_{n,j} \vee z_n - y_{n,j}) \psi^{(0)}(y_{n,j} \vee z_n|\mathbf{x}) \right] (1 + o(1)). \end{aligned}$$

Combining Lemma 1 and Lemma 4 with the asymptotic equivalent $\bar{F}(y_{n,j} \vee z_n|\mathbf{x}) / \bar{F}(y_n|\mathbf{x}) = (\theta \vee \tau_j^{-\gamma(\mathbf{x})})^{-1/\gamma(\mathbf{x})} (1 + o(1))$, we get that the first term dominates again. From straightforward calculations,

$$\begin{aligned} B_{j,J+1}^{(n)} &= \frac{\psi^{(1)}(y_{n,j} \vee z_n|\mathbf{x}) + (y_{n,j} \vee z_n - y_{n,j}) \psi^{(0)}(y_{n,j} \vee z_n|\mathbf{x})}{\psi^{(1)}(y_{n,j}|\mathbf{x}) \psi^{(0)}(z_n|\mathbf{x})} (1 + o(1)) \\ &= \frac{\gamma(\mathbf{x}) \left(\theta \vee \tau_j^{-\gamma(\mathbf{x})} \right)^{1-1/\gamma(\mathbf{x})} + (1 - \gamma(\mathbf{x})) \left(\theta \vee \tau_j^{-\gamma(\mathbf{x})} - \tau_j^{-\gamma(\mathbf{x})} \right) \left(\theta \vee \tau_j^{-\gamma(\mathbf{x})} \right)^{-1/\gamma(\mathbf{x})}}{\gamma(\mathbf{x}) \tau_j^{1-\gamma(\mathbf{x})} \theta^{-1/\gamma(\mathbf{x})}} \frac{1}{\bar{F}(y_n|\mathbf{x})} (1 + o(1)) \\ &= \frac{\gamma(\mathbf{x}) \left(\theta \vee \tau_j^{-\gamma(\mathbf{x})} \right)^{1-1/\gamma(\mathbf{x})} + (1 - \gamma(\mathbf{x})) \left(\theta \vee \tau_j^{-\gamma(\mathbf{x})} - \tau_j^{-\gamma(\mathbf{x})} \right) \theta^{-1/\gamma(\mathbf{x})}}{\gamma(\mathbf{x}) \tau_j^{1-\gamma(\mathbf{x})} \theta^{-1/\gamma(\mathbf{x})}} \frac{1}{\bar{F}(y_n|\mathbf{x})} (1 + o(1)). \end{aligned}$$

Finally, combining Lemmas 1, 4 and 6, the variance term $B_{J+1,J+1}^{(n)}$ is clearly

$$\frac{\mathbb{E} \left[K^2 \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > z_n\}} \right] - \left\{ \mathbb{E} \left[K \left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n} \right) \mathbb{1}_{\{Y > z_n\}} \right] \right\}^2}{\mu_K^{(2)}(\mathbf{x}, h_n) [\psi^{(0)}(z_n|\mathbf{x})]^2} = \theta^{1/\gamma(\mathbf{x})} \frac{1}{\bar{F}(y_n|\mathbf{x})} (1 + o(1)).$$

Therefore, $\text{Var}[Z_n] \rightarrow \beta^\top \mathbf{V}(\mathbf{x})\beta$, where $\mathbf{V}(\mathbf{x})$ is given in the statement of the lemma. It only remains to prove the asymptotic normality of Z_n . We thus write $Z_n = \sum_{i=1}^n Z_{i,n}$, where

$$\begin{aligned} Z_{i,n} &= \frac{v_n(\mathbf{x})}{n\mu_K^{(1)}(\mathbf{x}, h_n)} \sum_{j=1}^J \beta_j \frac{(Y_i - y_{n,j})K\left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n}\right) \mathbb{1}_{\{Y_i > y_{n,j}\}} - \mathbb{E}\left[(Y_i - y_{n,j})K\left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n}\right) \mathbb{1}_{\{Y_i > y_{n,j}\}}\right]}{\psi^{(1)}(y_{n,j}|\mathbf{x})} \\ &+ \frac{v_n(\mathbf{x})}{n\mu_K^{(1)}(\mathbf{x}, h_n)} \beta_{J+1} \frac{K\left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n}\right) \mathbb{1}_{\{Y_i > z_n\}} - \mathbb{E}\left[K\left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h_n}\right) \mathbb{1}_{\{Y_i > z_n\}}\right]}{\psi^{(0)}(z_n|\mathbf{x})}. \end{aligned}$$

We prove that there is $\delta > 0$ such that $n\mathbb{E}|Z_{1,n}|^{2+\delta} \rightarrow 0$ as $n \rightarrow \infty$; the result then follows by the Lyapunov central limit theorem (see Theorem 27.3 p.362 of Billingsley, 1995). For that purpose, note that, if T_1, \dots, T_q have finite $(2 + \delta)$ -moments,

$$\left\{ \mathbb{E} \left| \sum_{i=1}^q [T_i - \mathbb{E}(T_i)] \right|^{2+\delta} \right\}^{1/(2+\delta)} \leq \sum_{i=1}^q \{\mathbb{E}|T_i|^{2+\delta}\}^{1/(2+\delta)} + \mathbb{E}|T_i| \leq 2q \max_{1 \leq i \leq q} \{\mathbb{E}|T_i|^{2+\delta}\}^{1/(2+\delta)}$$

by the triangle inequality for the standard norm in the space of random variables with finite $(2 + \delta)$ -moment, and Jensen's inequality. Therefore, $\mathbb{E}|Z_{1,n}|^{2+\delta} =$

$$O \left(\left(\frac{\bar{F}(y_n|\mathbf{x})}{n\mu_K^{(2)}(\mathbf{x}, h_n)} \right)^{1+\delta/2} \left\{ \max_{1 \leq j \leq J} \mathbb{E} \left| \frac{(Y - y_{n,j})K\left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n}\right) \mathbb{1}_{\{Y > y_{n,j}\}}}{\psi^{(1)}(y_{n,j}|\mathbf{x})} \right|^{2+\delta} + \mathbb{E} \left| \frac{K\left(\frac{d(\mathbf{x}, \mathbf{X})}{h_n}\right) \mathbb{1}_{\{Y > z_n\}}}{\psi^{(0)}(z_n|\mathbf{x})} \right|^{2+\delta} \right\} \right).$$

Combine now, for δ small enough, Lemmas 1, 4 and 6 with the asymptotic equivalences $\bar{F}(z_n|\mathbf{x}) = \theta^{-1/\gamma(\mathbf{x})}\bar{F}(y_n|\mathbf{x})(1 + o(1))$ and $\bar{F}(y_{n,j}|\mathbf{x}) = \tau_j\bar{F}(y_n|\mathbf{x})(1 + o(1))$ to get

$$n\mathbb{E}|Z_{1,n}|^{2+\delta} = O \left([n\pi(\mathbf{x}, h_n)\bar{F}(y_n|\mathbf{x})]^{-\delta/2} \right),$$

which clearly tends to 0 as $n \rightarrow \infty$. The first result is proved. To show the second result, set

$$R_{n,j} := \frac{2 \left(\tilde{\psi}_n^{(1)}(y_{n,j}|\mathbf{x}) - \psi^{(1)}(y_{n,j}|\mathbf{x}) \right) - \left(\tilde{m}_n^{(1)}(y_{n,j}|\mathbf{x}) - m^{(1)}(y_{n,j}|\mathbf{x}) \right)}{2\psi^{(1)}(y_{n,j}|\mathbf{x}) - m^{(1)}(y_{n,j}|\mathbf{x})}.$$

We may write

$$\begin{aligned} &v_n(\mathbf{x}) \sum_{j=1}^J \beta_j \left(\frac{\widehat{E}_n(y_{n,j}|\mathbf{x})}{\bar{E}(y_{n,j}|\mathbf{x})} - 1 \right) + v_n(\mathbf{x}) \beta_{J+1} \left(\frac{\widehat{F}_n(z_n|\mathbf{x})}{\bar{F}(z_n|\mathbf{x})} - 1 \right) \\ &= v_n(\mathbf{x}) \sum_{j=1}^J \beta_j \left(\frac{\tilde{\psi}_n^{(1)}(y_{n,j}|\mathbf{x})}{\psi^{(1)}(y_{n,j}|\mathbf{x})} - 1 \right) (1 + R_{n,j})^{-1} + v_n(\mathbf{x}) \beta_{J+1} \left(\frac{\tilde{\psi}_n^{(0)}(z_n|\mathbf{x})}{\psi^{(0)}(z_n|\mathbf{x})} - 1 \right) \frac{1}{\tilde{m}_n^{(0)}(z_n|\mathbf{x})} \\ &+ v_n(\mathbf{x}) \sum_{j=1}^J \beta_j ((1 + R_{n,j})^{-1} - 1) + v_n(\mathbf{x}) \beta_{J+1} \left(\frac{1}{\tilde{m}_n^{(0)}(z_n|\mathbf{x})} - 1 \right). \end{aligned}$$

Combining Lemmas 1, 3 and 6 directly entails

$$\forall j \in \{1, \dots, J\}, R_{n,j} = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n\pi(\mathbf{x}, h_n)}} \right) = o_{\mathbb{P}} \left(\frac{1}{\sqrt{n\pi(\mathbf{x}, h_n)\bar{F}(y_n|\mathbf{x})}} \right) \xrightarrow{\mathbb{P}} 0$$

and the result then follows by applying the first result of the lemma. \square

Lemma 8. Assume (\mathcal{K}) and $\mathcal{C}_1(\gamma(\mathbf{x}))$ hold. Let $y_n \rightarrow \infty$, $h_n \rightarrow 0$ and $z_n = \theta y_n(1 + o(1))$, with $\theta > 0$, be such that $n\bar{F}(y_n|\mathbf{x})\pi(\mathbf{x}, h_n) \rightarrow \infty$ and

$$\exists \delta > 0, \sqrt{n\bar{F}(y_n|\mathbf{x})\pi(\mathbf{x}, h_n)} \log(y_n)\omega_{h_n}((1-\delta)(\theta \wedge 1)y_n|\mathbf{x}) \rightarrow 0.$$

Then if, for all $j \in \{1, \dots, J\}$, $y_{n,j} = \tau_j^{-\gamma(\mathbf{x})} y_n(1 + o(1))$ with $0 < \tau_1 < \tau_2 < \dots < \tau_J \leq 1$, one has

$$\sqrt{n\bar{F}(y_n|\mathbf{x}) \frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)}} \left\{ \left(\frac{\widehat{F}_n(y_{n,j}|\mathbf{x})}{\bar{F}(y_{n,j}|\mathbf{x})} - 1 \right)_{1 \leq j \leq J}, \left(\frac{\widehat{F}_n(z_n|\mathbf{x})}{\bar{F}(z_n|\mathbf{x})} - 1 \right) \right\} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{J+1}, \mathbf{\Sigma}(\mathbf{x})),$$

where $\mathbf{\Sigma}(\mathbf{x})$ is the symmetric matrix having entries:

$$\begin{cases} \Sigma_{j,l}(\mathbf{x}) = \tau_l^{-1}, (j,l) \in \{1, \dots, J\}^2, j \leq l, \\ \Sigma_{j,J+1}(\mathbf{x}) = (\tau_j \vee \theta^{-1/\gamma(\mathbf{x})})^{-1}, j \in \{1, \dots, J\}, \\ \Sigma_{J+1,J+1}(\mathbf{x}) = \theta^{1/\gamma(\mathbf{x})}. \end{cases}$$

Proof. The proof is entirely similar to that of Lemma 7 and is thus omitted. \square

A.2 Proofs of main results

Proof of Theorem 1 Let us recall that $\sigma_n^{-1}(\mathbf{x}) = \sqrt{n(1 - \alpha_n) \frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)}}$, introduce $\mathbf{z} = (z_1, \dots, z_J, t)$ and focus on the probability

$$\begin{aligned} \Phi_n(\mathbf{z}) &= \mathbb{P} \left(\bigcap_{j=1}^J \left\{ \sigma_n^{-1}(\mathbf{x}) \left(\frac{\widehat{e}_n(\alpha_{n,j}|\mathbf{x})}{e(\alpha_{n,j}|\mathbf{x})} - 1 \right) \leq z_j \right\} \cap \left\{ \sigma_n^{-1}(\mathbf{x}) \left(\frac{\widehat{q}_n(a_n|\mathbf{x})}{q(a_n|\mathbf{x})} - 1 \right) \leq t \right\} \right) \\ &= \mathbb{P} \left(\bigcap_{j=1}^J \left\{ \widehat{e}_n(\alpha_{n,j}|\mathbf{x}) \leq e(\alpha_{n,j}|\mathbf{x}) (1 + z_j \sigma_n(\mathbf{x})) \right\} \cap \left\{ \widehat{q}_n(a_n|\mathbf{x}) \leq q(a_n|\mathbf{x}) (1 + t \sigma_n(\mathbf{x})) \right\} \right). \end{aligned}$$

By noticing that $1 - \alpha_{n,j} = \bar{E}(e(\alpha_{n,j}|\mathbf{x})|\mathbf{x})$ and $1 - a_n = \bar{F}(q(a_n|\mathbf{x})|\mathbf{x})$, and using that for all y , $\alpha, \widehat{e}_n(\alpha|\mathbf{x}) \leq y \Leftrightarrow \widehat{E}_n(y|\mathbf{x}) \leq 1 - \alpha$ and $\widehat{q}_n(\alpha|\mathbf{x}) \leq y \Leftrightarrow \widehat{F}_n(y|\mathbf{x}) \leq 1 - \alpha$, we find that $\Phi_n(\mathbf{z})$ is equal to

$$\mathbb{P} \left(\bigcap_{j=1}^J \left\{ \widehat{E}_n(e(\alpha_{n,j}|\mathbf{x})(1 + z_j \sigma_n(\mathbf{x}))|\mathbf{x}) \leq \bar{E}(e(\alpha_{n,j}|\mathbf{x})|\mathbf{x}) \right\} \cap \left\{ \widehat{F}_n(q(a_n|\mathbf{x})(1 + t \sigma_n(\mathbf{x}))|\mathbf{x}) \leq \bar{F}(q(a_n|\mathbf{x})|\mathbf{x}) \right\} \right).$$

Letting $y_n = e(\alpha_n|\mathbf{x})$, $y_{n,j} = e(\alpha_{n,j}|\mathbf{x})(1 + z_j \sigma_n(\mathbf{x}))$, $z_n = q(a_n|\mathbf{x})(1 + t \sigma_n(\mathbf{x}))$ and $v_n(\mathbf{x}) = \sqrt{n\bar{F}(y_n|\mathbf{x}) \frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)}}$, we have

$$\begin{aligned} \Phi_n(\mathbf{z}) &= \mathbb{P} \left(\bigcap_{j=1}^J \left\{ v_n(\mathbf{x}) \left(\frac{\widehat{E}_n(y_{n,j}|\mathbf{x})}{\bar{E}(y_{n,j}|\mathbf{x})} - 1 \right) \leq v_n(\mathbf{x}) \left(\frac{\bar{E}(e(\alpha_{n,j}|\mathbf{x})|\mathbf{x})}{\bar{E}(y_{n,j}|\mathbf{x})} - 1 \right) \right\} \right. \\ &\quad \left. \cap \left\{ v_n(\mathbf{x}) \left(\frac{\widehat{F}_n(z_n|\mathbf{x})}{\bar{F}(z_n|\mathbf{x})} - 1 \right) \leq v_n(\mathbf{x}) \left(\frac{\bar{F}(q(a_n|\mathbf{x})|\mathbf{x})}{\bar{F}(z_n|\mathbf{x})} - 1 \right) \right\} \right). \end{aligned}$$

By Corollary 3 of Girard et al. (2021), $\bar{E}(\cdot|\mathbf{x})$ is continuously differentiable with

$$\lim_{y \rightarrow \infty} \frac{y \bar{E}'(y|\mathbf{x})}{\bar{E}(y|\mathbf{x})} = \lim_{y \rightarrow \infty} \frac{y \bar{F}'(y|\mathbf{x})}{\bar{F}(y|\mathbf{x})} = -\frac{1}{\gamma(\mathbf{x})}.$$

Thus, for any $j \in \{1, \dots, J\}$, a Taylor expansion shows that there exists $\theta_{n,j} \in [0, 1]$ with:

$$\begin{aligned} \bar{E}(y_{n,j}|\mathbf{x}) &= 1 - \alpha_{n,j} + z_j \sigma_n(\mathbf{x}) e(\alpha_{n,j}|\mathbf{x}) \bar{E}'(e(\alpha_{n,j}|\mathbf{x}) + \theta_{n,j} z_j \sigma_n(\mathbf{x}) e(\alpha_{n,j}|\mathbf{x})|\mathbf{x}) \\ &= 1 - \alpha_{n,j} + z_j \sigma_n(\mathbf{x}) e(\alpha_{n,j}|\mathbf{x}) \bar{E}'(e(\alpha_{n,j}|\mathbf{x})|\mathbf{x})(1 + o(1)) \\ &= (1 - \alpha_{n,j}) \left(1 - \frac{z_j \sigma_n(\mathbf{x})}{\gamma(\mathbf{x})} (1 + o(1)) \right) = \bar{E}(e(\alpha_{n,j}|\mathbf{x})|\mathbf{x}) \left(1 - \frac{z_j \sigma_n(\mathbf{x})}{\gamma(\mathbf{x})} (1 + o(1)) \right). \end{aligned}$$

In view of (4.2), $v_n(\mathbf{x}) = \sigma_n^{-1}(\mathbf{x}) \sqrt{\gamma(\mathbf{x})^{-1} - 1} (1 + o(1))$ and therefore

$$\forall j \in \{1, \dots, J\}, \quad v_n(\mathbf{x}) \left(\frac{\bar{E}(e(\alpha_{n,j}|\mathbf{x})|\mathbf{x})}{\bar{E}(y_{n,j}|\mathbf{x})} - 1 \right) = \frac{z_j}{\gamma(\mathbf{x})} \sqrt{\frac{1 - \gamma(\mathbf{x})}{\gamma(\mathbf{x})}} (1 + o(1)).$$

Similarly

$$v_n(\mathbf{x}) \left(\frac{\bar{F}(q(a_n|\mathbf{x})|\mathbf{x})}{\bar{F}(z_n|\mathbf{x})} - 1 \right) = \frac{t}{\gamma(\mathbf{x})} \sqrt{\frac{1 - \gamma(\mathbf{x})}{\gamma(\mathbf{x})}} (1 + o(1)).$$

Finally, $y_{n,j} = \tau_j^{-\gamma(\mathbf{x})} y_n (1 + o(1))$ and $z_n = (\gamma(\mathbf{x})^{-1} - 1)^{\gamma(\mathbf{x})} \tau^{-\gamma(\mathbf{x})} y_n (1 + o(1))$, since, in view of (4.3), $e(\cdot|\mathbf{x})$ is regularly varying. Thus, for n large enough, $\omega_{h_n}(y_{n,j}|\mathbf{x}) \leq \omega_{h_n}((1 - \delta)e(\alpha_n|\mathbf{x})|\mathbf{x})$ and $\omega_{h_n}(z_n|\mathbf{x}) \leq \omega_{h_n}((1 - \delta)q(a_n|\mathbf{x})|\mathbf{x})$. Moreover, $\log(y_{n,j}) = O(\log(y_n))$ and $\log(z_n) = O(\log(y_n))$. Noting that $\log(y_n) = O(|\log(1 - \alpha_n)|)$ by the regular variation property of the tail expectile function $t \mapsto e(1 - t^{-1}|\mathbf{x})$, Lemma 7 can then be applied with $\theta = \tau^{-\gamma(\mathbf{x})} (\gamma(\mathbf{x})^{-1} - 1)^{\gamma(\mathbf{x})}$ to conclude the proof. \square

Proof of Theorem 2 The key is to write

$$\log \left(\frac{\hat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})}{e(\beta_n|\mathbf{x})} \right) = \log \left(\frac{1 - \alpha_n}{1 - \beta_n} \right) (\hat{\gamma}_{\alpha_n}(\mathbf{x}) - \gamma(\mathbf{x})) + \log \left(\frac{\hat{e}_n(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} \right) + \log \left(\left[\frac{1 - \alpha_n}{1 - \beta_n} \right]^{\gamma(\mathbf{x})} \frac{e(\alpha_n|\mathbf{x})}{e(\beta_n|\mathbf{x})} \right).$$

The first term is controlled by our assumptions on $\hat{\gamma}_{\alpha_n}(\mathbf{x})$, the second term by Theorem 1 and the third one by Lemma 8 of Girard et al. (2021) (whose proof carries over to the functional case without modification). The delta-method concludes the proof. \square

Proof of Theorem 3 The key is to write

$$\begin{aligned} \log \left(\frac{\hat{e}_{n,\alpha_n}^W(\beta_n|\mathbf{x})}{e(\beta_n|\mathbf{x})} \right) &= \log \left(\frac{1 - \alpha_n}{1 - \beta_n} \right) (\hat{\gamma}_{\alpha_n}(\mathbf{x}) - \gamma(\mathbf{x})) + \log \left(\frac{\hat{q}_n(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} \right) \\ &+ \log \left(\left[\frac{1 - \alpha_n}{1 - \beta_n} \right]^{\gamma(\mathbf{x})} \frac{q(\alpha_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} \right) \\ &+ \log \left(\frac{(\hat{\gamma}_{\alpha_n}(\mathbf{x})^{-1} - 1)^{-\hat{\gamma}_{\alpha_n}(\mathbf{x})}}{(\gamma(\mathbf{x})^{-1} - 1)^{-\gamma(\mathbf{x})}} \right) + \log \left(\frac{q(\beta_n|\mathbf{x}) (\gamma(\mathbf{x})^{-1} - 1)^{-\gamma(\mathbf{x})}}{e(\beta_n|\mathbf{x})} \right). \end{aligned}$$

Under our conditions, the second term is a $O_{\mathbb{P}}(\sigma_n(\mathbf{x}))$. According to Theorem 4.3.8 in de Haan and Ferreira (2006) and condition *iii*), the third term is a $O(A((1 - \alpha_n)^{-1}|\mathbf{x}))$. Similarly, Proposition 1 in Daouia et al. (2018) entails that the last term is a $O(A((1 - \alpha_n)^{-1}|\mathbf{x})) + O(q(\alpha_n|\mathbf{x})^{-1})$. Finally, our assumptions on $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ imply that the fourth term is a $O_{\mathbb{P}}(\sigma_n(\mathbf{x}))$. The result follows. \square

Extension and proof of Theorem 4 The proof of Theorem 4 hinges on the result below.

Recall that $\sigma_n(\mathbf{x}) = 1/\sqrt{n(1-\alpha_n)\frac{\mu_K^{(1)}(\mathbf{x}, h_n)^2}{\mu_K^{(2)}(\mathbf{x}, h_n)}}$.

Proposition 1. Assume (\mathcal{K}) and $\mathcal{C}_1(\gamma(\mathbf{x}))$ hold. Let $\alpha_n \rightarrow 1$, $h_n \rightarrow 0$ and $1 - a_n = \tau(1 - \alpha_n)(1 + o(1))$ be such that $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ and

$$\exists \delta > 0, \sqrt{n(1 - \alpha_n)\pi(\mathbf{x}, h_n)}\omega_{h_n}((1 - \delta)(\tau^{-\gamma(\mathbf{x})} \wedge 1)q(\alpha_n|\mathbf{x})|\mathbf{x}) \log(1 - \alpha_n) \rightarrow 0$$

Then, if $1 - \alpha_{n,j} = \tau_j(1 - \alpha_n)$ for some $0 < \tau_1 < \tau_2 < \dots < \tau_J \leq 1$, one has

$$\sigma_n^{-1}(\mathbf{x}) \left\{ \left(\frac{\widehat{q}_n(\alpha_{n,j}|\mathbf{x})}{q(\alpha_{n,j}|\mathbf{x})} - 1 \right)_{1 \leq j \leq J}, \left(\frac{\widehat{q}_n(a_n|\mathbf{x})}{q(a_n|\mathbf{x})} - 1 \right) \right\} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{J+1}, \gamma^2(\mathbf{x})\Sigma(\mathbf{x})),$$

where $\Sigma(\mathbf{x})$ is defined in Lemma 8 (with $\theta = \tau^{-\gamma(\mathbf{x})}$).

Proof. The proof is entirely similar to that of Theorem 1, applying Lemma 8 rather than Lemma 7 at the final step. We omit the details. \square

Theorem 7. Assume (\mathcal{K}) and $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ hold. Let $\alpha_n \rightarrow 1$, $h_n \rightarrow 0$ be such that $n(1 - \alpha_n)\pi(\mathbf{x}, h_n) \rightarrow \infty$ and introduce $1 - a_n = \tau(1 - \alpha_n)(1 + o(1))$, where $\tau > 0$. Assume further that

$$i) \sigma_n^{-1}(\mathbf{x}) \log(1 - \alpha_n) \omega_{h_n}((1 - \delta)(\tau^{-\gamma(\mathbf{x})} \wedge 1)q(\alpha_n|\mathbf{x})|\mathbf{x}) \rightarrow 0 \text{ for some } \delta > 0,$$

$$ii) \sigma_n^{-1}(\mathbf{x}) A((1 - \alpha_n)^{-1}|\mathbf{x}) \rightarrow \lambda_1 \in \mathbb{R}.$$

Then

$$\sigma_n^{-1}(\mathbf{x}) \left(\widehat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x}), \frac{\widehat{q}_n(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} - 1, \frac{\widehat{q}_n(a_n|\mathbf{x})}{q(a_n|\mathbf{x})} - 1 \right)$$

converges in distribution to a Gaussian random vector with expectation

$$\left(\frac{1}{\log(J!)} \left(\sum_{j=2}^J \frac{j^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \right) \lambda_1, 0, 0 \right)$$

and covariance matrix

$$\gamma^2(\mathbf{x}) \begin{pmatrix} \frac{J(J-1)(2J-1)}{6 \log(J!)^2} & 0 & \frac{1}{\log(J!)} \sum_{j=1}^{J-1} [(\tau_j \vee \tau)^{-1} - (1 \vee \tau)^{-1}] \\ 0 & 1 & (\tau \vee 1)^{-1} \\ \frac{1}{\log(J!)} \sum_{j=1}^{J-1} [(\tau_j \vee \tau)^{-1} - (1 \vee \tau)^{-1}] & (\tau \vee 1)^{-1} & \tau^{-1} \end{pmatrix}$$

where $\tau_j = (J - j + 1)^{-1}$ for all $j \in \{1, \dots, J\}$.

Proof. Recall that $\tau_j = 1/(J - j + 1)$ (in particular $\tau_J = 1$) and therefore

$$\begin{aligned}
& \sigma_n^{-1}(\widehat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x})) \\
&= \sigma_n^{-1}(\mathbf{x}) \left(\frac{1}{\log(J!)} \sum_{j=1}^{J-1} \log \left(\frac{\widehat{q}_n(1 - \tau_j(1 - \alpha_n)|\mathbf{x})}{\widehat{q}_n(1 - \tau_J(1 - \alpha_n)|\mathbf{x})} \right) - \gamma(\mathbf{x}) \right) \\
&= \sigma_n^{-1}(\mathbf{x}) \left(\frac{1}{\log(J!)} \sum_{j=1}^{J-1} \log \left(\frac{\widehat{q}_n(1 - \tau_j(1 - \alpha_n)|\mathbf{x})}{q(1 - \tau_j(1 - \alpha_n)|\mathbf{x})} \right) - \frac{J-1}{\log(J!)} \log \left(\frac{\widehat{q}_n(1 - \tau_J(1 - \alpha_n)|\mathbf{x})}{q(1 - \tau_J(1 - \alpha_n)|\mathbf{x})} \right) \right) \\
&+ \sigma_n^{-1}(\mathbf{x}) \left(\frac{1}{\log(J!)} \sum_{j=1}^{J-1} \log \left(\frac{q(1 - \tau_j(1 - \alpha_n)|\mathbf{x})}{q(\alpha_n|\mathbf{x})} \right) - \gamma(\mathbf{x}) \right) =: A_n + B_n.
\end{aligned}$$

The bias term B_n is controlled by applying Theorem 2.3.9 in de Haan and Ferreira (2006): since

$$\forall u > 0, \quad \frac{q(1 - (1 - \alpha_n)/u|\mathbf{x})}{q(\alpha_n|\mathbf{x})} = u^{\gamma(\mathbf{x})} \left(1 + A((1 - \alpha_n)^{-1}|\mathbf{x}) \left[\frac{u^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} + o(1) \right] \right),$$

it is easily seen that, by condition *ii*),

$$B_n \rightarrow \frac{1}{\log(J!)} \left(\sum_{j=1}^{J-1} \frac{\tau_j^{-\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \right) \lambda_1 = \frac{1}{\log(J!)} \left(\sum_{j=2}^J \frac{j^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \right) \lambda_1 \quad \text{as } n \rightarrow \infty.$$

Besides, by Proposition 1 and a Taylor expansion,

$$A_n = \sigma_n^{-1}(\mathbf{x}) \frac{1}{\log(J!)} (1, \dots, 1, -(J-1)) \begin{pmatrix} \frac{\widehat{q}_n(1 - \tau_1(1 - \alpha_n)|\mathbf{x})}{q(1 - \tau_1(1 - \alpha_n)|\mathbf{x})} - 1 \\ \vdots \\ \frac{\widehat{q}_n(1 - \tau_{J-1}(1 - \alpha_n)|\mathbf{x})}{q(1 - \tau_{J-1}(1 - \alpha_n)|\mathbf{x})} - 1 \\ \frac{\widehat{q}_n(1 - \tau_J(1 - \alpha_n)|\mathbf{x})}{q(1 - \tau_J(1 - \alpha_n)|\mathbf{x})} - 1 \end{pmatrix} + o_{\mathbb{P}}(1).$$

By Proposition 1 again, it follows that $\Delta_n = \sigma_n^{-1}(\mathbf{x}) \left(\widehat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x}), \frac{\widehat{q}_n(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} - 1, \frac{\widehat{q}_n(a_n|\mathbf{x})}{q(a_n|\mathbf{x})} - 1 \right)$ is asymptotically Gaussian, with asymptotic mean

$$\left(\frac{1}{\log(J!)} \left(\sum_{j=2}^J \frac{j^{\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \right) \lambda_1, 0, 0 \right).$$

It remains to calculate the covariance matrix of this limiting Gaussian distribution. Note that, by Proposition 1 and straightforward calculations,

$$\text{Var}(A_n) \rightarrow \frac{J(J-1)(2J-1)}{6 \log(J!)^2}$$

(see also Gardes and Girard (2012)) and

$$\begin{aligned}
\text{cov} \left(A_n, \frac{\widehat{q}_n(a_n|\mathbf{x})}{q(a_n|\mathbf{x})} - 1 \right) &\rightarrow \frac{1}{\log(J!)} \left(\sum_{j=1}^{J-1} (\tau_j \vee \tau)^{-1} - (J-1)(\tau_J \vee \tau)^{-1} \right) \\
&= \frac{1}{\log(J!)} \sum_{j=1}^{J-1} [(\tau_j \vee \tau)^{-1} - (1 \vee \tau)^{-1}].
\end{aligned}$$

It now follows directly from Proposition 1 and straightforward calculations that the covariance matrix is

$$\gamma^2(\mathbf{x}) \begin{pmatrix} \frac{J(J-1)(2J-1)}{6 \log(J!)^2} & 0 & \frac{1}{\log(J!)} \sum_{j=1}^{J-1} [(\tau_j \vee \tau)^{-1} - (1 \vee \tau)^{-1}] \\ 0 & 1 & (\tau \vee 1)^{-1} \\ \frac{1}{\log(J!)} \sum_{j=1}^{J-1} [(\tau_j \vee \tau)^{-1} - (1 \vee \tau)^{-1}] & (\tau \vee 1)^{-1} & \tau^{-1} \end{pmatrix}$$

as required. \square

Proof of Theorem 5 Let $\tau = \gamma(\mathbf{x})^{-1} - 1$. For any $z \in \mathbb{R}$, let us focus on the probability

$$\Phi_n(z) = \mathbb{P} \left(\sigma_n^{-1}(\mathbf{x}) \left(\frac{\hat{F}_n(\hat{e}_n(\alpha_n|\mathbf{x})|\mathbf{x})}{1 - \alpha_n} - \tau \right) \leq z \right) = \mathbb{P} \left(\hat{F}_n(\hat{e}_n(\alpha_n|\mathbf{x})|\mathbf{x}) \leq (1 - \alpha_n)(\tau + z\sigma_n(\mathbf{x})) \right).$$

Equivalently, $\Phi_n(z) = \mathbb{P}(\hat{e}_n(\alpha_n|\mathbf{x}) \geq \hat{q}_n(\beta_n|\mathbf{x}))$, where $\beta_n = 1 - (1 - \alpha_n)(\tau + z\sigma_n(\mathbf{x}))$ and therefore $\Phi_n(z)$ is given by

$$\mathbb{P} \left(\sigma_n^{-1}(\mathbf{x}) \left(\frac{\hat{e}_n(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right) \geq \sigma_n^{-1}(\mathbf{x}) \left(\frac{\hat{q}_n(\beta_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} - 1 \right) \frac{q(\beta_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} + \sigma_n^{-1}(\mathbf{x}) \left(\frac{q(\beta_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right) \right).$$

From de Haan and Ferreira (2006, Theorem 2.3.9), condition $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$ entails

$$\frac{q(\beta_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} = \tau^{-\gamma(\mathbf{x})} \left[1 + \frac{\tau^{-\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} A((1 - \alpha_n)^{-1}|\mathbf{x}) (1 + o(1)) - \frac{\gamma(\mathbf{x})z\sigma_n(\mathbf{x})}{\tau} (1 + o(1)) \right]. \quad (\text{A.1})$$

In addition, under $\mathcal{C}_2(\gamma(\mathbf{x}), \rho(\mathbf{x}), A(\cdot|\mathbf{x}))$, Daouia et al. (2018, Corollary 1) yields

$$\frac{e(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} = \tau^{-\gamma(\mathbf{x})} \left[1 + r(\alpha_n|\mathbf{x}) + \frac{\tau^{-\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} A((1 - \alpha_n)^{-1}|\mathbf{x}) \right], \quad (\text{A.2})$$

where

$$r(\alpha_n|\mathbf{x}) = \frac{\gamma(\mathbf{x})\tau^{\gamma(\mathbf{x})}}{q(\alpha_n|\mathbf{x})} (\mathbb{E}[Y|\mathbf{X} = \mathbf{x}] + o(1)) + \left(\frac{\tau^{-\rho(\mathbf{x})}}{1 - \gamma(\mathbf{x}) - \rho(\mathbf{x})} + o(1) \right) A((1 - \alpha_n)^{-1}|\mathbf{x}).$$

Combining (A.1) and (A.2) yields

$$\sigma_n^{-1}(\mathbf{x}) \left(\frac{q(\beta_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right) = -z \frac{\gamma(\mathbf{x})}{\tau} + o(1) - \sigma_n^{-1}(\mathbf{x}) r(\alpha_n|\mathbf{x}) (1 + o(1)).$$

In view of conditions *ii*) and *iii*), one has

$$\sigma_n^{-1}(\mathbf{x}) r(\alpha_n|\mathbf{x}) \rightarrow \lambda = \frac{(\gamma(\mathbf{x})^{-1} - 1)^{-\rho(\mathbf{x})}}{1 - \gamma(\mathbf{x}) - \rho(\mathbf{x})} \lambda_1 + \gamma(\mathbf{x}) (\gamma(\mathbf{x})^{-1} - 1)^{\gamma(\mathbf{x})} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}] \lambda_2 \quad \text{as } n \rightarrow \infty.$$

This entails

$$\sigma_n^{-1}(\mathbf{x}) \left(\frac{q(\beta_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right) = -z \frac{\gamma(\mathbf{x})}{\tau} - \lambda + o(1). \quad (\text{A.3})$$

Then, $\Phi_n(z)$ can be rewritten as

$$\mathbb{P} \left(\frac{\tau}{\gamma(\mathbf{x})} \left[-\sigma_n^{-1}(\mathbf{x}) \left(\frac{\hat{e}_n(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right) + \sigma_n^{-1}(\mathbf{x}) \left(\frac{\hat{q}_n(\beta_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} - 1 \right) (1 + o(1)) \right] - \lambda \frac{\tau}{\gamma(\mathbf{x})} + o(1) \leq z \right),$$

and therefore the asymptotic distribution of $\sigma_n^{-1}(\mathbf{x}) \left(\frac{\hat{F}_n(\hat{e}_n(\alpha_n|\mathbf{x})|\mathbf{x})}{1-\alpha_n} - \tau \right)$ is the same as that of

$$\frac{1-\gamma(\mathbf{x})}{\gamma^2(\mathbf{x})} (-1, 1) \sigma_n^{-1}(\mathbf{x}) \left(\frac{\frac{\hat{e}_n(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1}{\frac{\hat{q}_n(\beta_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} - 1} \right) - \lambda \frac{1-\gamma(\mathbf{x})}{\gamma^2(\mathbf{x})}.$$

Finally, $\hat{\gamma}_{\alpha_n}(\mathbf{x})$ being obtained by applying the function $u \mapsto (1+u)^{-1}$ to $\hat{F}_n(\hat{e}_n(\alpha_n|\mathbf{x})|\mathbf{x})/(1-\alpha_n)$, its asymptotic distribution is the limit in distribution of

$$(1-\gamma(\mathbf{x})) (1, -1) \sigma_n^{-1}(\mathbf{x}) \left(\frac{\frac{\hat{e}_n(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1}{\frac{\hat{q}_n(\beta_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} - 1} \right) + \lambda(1-\gamma(\mathbf{x}))$$

by the delta-method. Apply Theorem 1 to conclude the proof. \square

Proof of Theorem 6 Let $\tau = \gamma(\mathbf{x})^{-1} - 1$, $\tilde{e}_n(\alpha_n|\mathbf{x}) = \left(\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})^{-1} - 1 \right)^{-\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})} \hat{q}_n(\alpha_n|\mathbf{x})$ and for any $z \in \mathbb{R}$, focus on the probability

$$\Phi_n(z) = \mathbb{P} \left(\sigma_n(\mathbf{x})^{-1} \left(\frac{\hat{F}_n(\tilde{e}_n(\alpha_n|\mathbf{x})|\mathbf{x})}{1-\alpha_n} - \tau \right) \leq z \right) = \mathbb{P} \left(\hat{F}_n(\tilde{e}_n(\alpha_n|\mathbf{x})|\mathbf{x}) \leq (1-\alpha_n)(\tau + z\sigma_n(\mathbf{x})) \right).$$

Similarly to the previous proof, and using (A.3), $\Phi_n(z)$ can be rewritten as

$$\mathbb{P} \left(\frac{\tau}{\gamma(\mathbf{x})} \left[-\sigma_n^{-1}(\mathbf{x}) \left(\frac{\tilde{e}_n(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right) + \sigma_n^{-1}(\mathbf{x}) \left(\frac{\hat{q}_n(\beta_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} - 1 \right) (1 + o(1)) \right] - \lambda \frac{\tau}{\gamma(\mathbf{x})} + o(1) \leq z \right)$$

where $\beta_n = 1 - (1 - \alpha_n)(\tau + z\sigma_n(\mathbf{x}))$. Consider the expansion:

$$\begin{aligned} & \sigma_n(\mathbf{x})^{-1} \left(\frac{\tilde{e}_n(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right) \\ &= \sigma_n(\mathbf{x})^{-1} \left(\left(\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})^{-1} - 1 \right)^{-\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})} - \tau^{-\gamma(\mathbf{x})} \right) \frac{q(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} \frac{\hat{q}_n(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} \\ &+ \sigma_n(\mathbf{x})^{-1} \left(\frac{\hat{q}_n(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} - 1 \right) \frac{\tau^{-\gamma(\mathbf{x})} q(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} + \sigma_n(\mathbf{x})^{-1} \left(\frac{\tau^{-\gamma(\mathbf{x})} q(\alpha_n|\mathbf{x})}{e(\alpha_n|\mathbf{x})} - 1 \right). \end{aligned}$$

By (A.2) above, Proposition 1, Theorem 4, and Slutsky's lemma, the asymptotic distribution of $\sigma_n(\mathbf{x})^{-1} (\tilde{e}_n(\alpha_n|\mathbf{x})/e(\alpha_n|\mathbf{x}) - 1)$ is that of

$$\sigma_n(\mathbf{x})^{-1} \left(\frac{1}{1-\gamma(\mathbf{x})} - \log(\gamma(\mathbf{x})^{-1} - 1), 1 \right) \left(\frac{\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x})}{\frac{\hat{q}_n(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} - 1} \right) - \lambda - \frac{\tau^{-\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \lambda_1.$$

Recall then (A.3) to find that $\sigma_n(\mathbf{x})^{-1} \left(\hat{F}_n(\tilde{e}_n(\alpha_n|\mathbf{x})|\mathbf{x})/(1-\alpha_n) - \tau \right)$ has the asymptotic distribution of

$$-\frac{\tau}{\gamma(\mathbf{x})} \sigma_n(\mathbf{x})^{-1} \left(\frac{1}{1-\gamma(\mathbf{x})} - \log(\gamma(\mathbf{x})^{-1} - 1), 1, -1 \right) \left(\frac{\frac{\hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x})}{\frac{\hat{q}_n(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} - 1}}{\frac{\hat{q}_n(\beta_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} - 1} \right) + \frac{\tau}{\gamma(\mathbf{x})} \frac{\tau^{-\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \lambda_1.$$

Finally, $\tilde{\gamma}_{\alpha_n}^{(J)}(\mathbf{x})$ being obtained by applying the function $u \mapsto (1+u)^{-1}$ to $\hat{\tilde{F}}_n(\tilde{e}_n(\alpha_n|\mathbf{x})|\mathbf{x})/(1-\alpha_n)$, its asymptotic distribution is the limit in distribution of

$$(1-\gamma(\mathbf{x}))\sigma_n(\mathbf{x})^{-1} \left(\frac{1}{1-\gamma(\mathbf{x})} - \log(\gamma(\mathbf{x})^{-1} - 1), 1, -1 \right) \begin{pmatrix} \hat{\gamma}_{\alpha_n}^{(J)}(\mathbf{x}) - \gamma(\mathbf{x}) \\ \frac{\hat{q}_n(\alpha_n|\mathbf{x})}{q(\alpha_n|\mathbf{x})} - 1 \\ \frac{\hat{q}_n(\beta_n|\mathbf{x})}{q(\beta_n|\mathbf{x})} - 1 \end{pmatrix} - (1-\gamma(\mathbf{x})) \frac{\tau^{-\rho(\mathbf{x})} - 1}{\rho(\mathbf{x})} \lambda_1$$

by the delta-method. Apply Theorem 7 to conclude the proof. \square