WORKING PAPERS

N° 1541

April 2024

“Extreme expectile estimation for short-tailed data”

Abdelaati Daouia, Simone A. Padoan and Gilles Stupfler

Toulouse School of Economics

Extreme expectile estimation for short-tailed data

Abdelaati Daouia[a]*, Simone A. Padoan[b]† & Gilles Stupfler[c]‡

[a] Toulouse School of Economics, University of Toulouse Capitole, France
[b] Department of Decision Sciences, Bocconi University, via Roentgen 1, 20136 Milano, Italy
[c] Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France

Abstract

The use of expectiles in risk management has recently gathered remarkable momentum due to their excellent axiomatic and probabilistic properties. In particular, the class of elicitable law-invariant coherent risk measures only consists of expectiles. While the theory of expectile estimation at central levels is substantial, tail estimation at extreme levels has so far only been considered when the tail of the underlying distribution is heavy. This article is the first work to handle the short-tailed setting where the loss (*e.g.* negative log-returns) distribution of interest is bounded to the right and the corresponding extreme value index is negative. This is motivated by the assessment of long-term market risk carried by low-frequency (*e.g.* weekly) returns of equities that show evidence of being generated from short-tailed distributions. We derive an asymptotic expansion of tail expectiles in this challenging context under a general second-order extreme value condition, which allows to come up with two semi-parametric estimators of extreme expectiles, and with their asymptotic properties in a general model of strictly stationary but weakly dependent observations. We also extend the applicability of the proposed method to the regression setting. A simulation study and a real data analysis from a forecasting perspective are performed to compare the proposed competing estimation procedures.

JEL Codes: C13, C14, C18, C53, C58
Keywords: Expectiles, Extreme values, Second-order condition, Short tails, Weak dependence

1 Introduction

The class of expectiles, introduced by Newey and Powell (1987), defines useful descriptors ξ_τ of the higher ($\tau \geq \frac{1}{2}$) and lower ($\tau \leq \frac{1}{2}$) regions of the distribution of a random variable X through the asymmetric least squares minimization problem

$$\xi_\tau = \underset{\theta \in \mathbb{R}}{\arg\min}\, \mathbb{E}(\eta_\tau(X-\theta) - \eta_\tau(X)), \tag{1.1}$$

where $\eta_\tau(x) = |\tau - \mathbb{1}\{x \leq 0\}|\, x^2$, with $\mathbb{1}\{\cdot\}$ being the indicator function and $\tau \in (0,1)$ the asymmetry level. Expectiles are well-defined, finite and uniquely determined as soon the first moment of X is finite. They generalize the mean, found for $\tau = 1/2$, in the same way

*Corresponding author. E-mail: abdelaati.daouia@tse-fr.eu. Orcid: 0000-0003-2621-8860
†E-mail: simone.padoan@unibocconi.it. Orcid: 0000-0002-0417-7570
‡E-mail: gilles.stupfler@univ-angers.fr. Orcid: 0000-0003-2497-9412

quantiles generalize the median: Koenker and Bassett (1978) showed that the τth quantile q_τ of X solves the asymmetric L^1 minimization problem

$$q_\tau \in \arg\min_{\theta\in\mathbb{R}} \mathbb{E}(\varrho_\tau(X-\theta) - \varrho_\tau(X)),$$

where $\varrho_\tau(x) = |\tau - \mathbb{1}\{x \le 0\}|\,|x|$ is the so-called quantile check function. Expectiles have received renewed attention for their ability to quantify tail risk in statistical decision theory at least since the contribution of Taylor (2008). They depend on the tail realizations of X and their probability, while quantiles only depend on the frequency of tail realizations, see Kuan et al. (2009). Most importantly, Ziegel (2016) showed that expectiles are the only coherent law-invariant measure of risk which is also elicitable in the sense of Gneiting (2011), meaning that they abide by the intuitive diversification principle (Bellini et al., 2014) and that their prediction can be performed through a straightforward principled backtesting methodology. These merits and good properties have motivated the development of procedures for expectile estimation and inference over the last decade. A key, but difficult, question in any risk management setup is the estimation of the expectile ξ_τ at extreme levels, which grow to 1 as the sample size increases. This question was first tackled in Daouia et al. (2018, 2020) under the assumption that the underlying distribution is heavy-tailed, that is, its distribution function tends to 1 algebraically fast. The latest developments under this assumption have focused on, among others, bias reduction (Girard et al., 2022), accurate inference (Padoan and Stupfler, 2022), and handling more complex data in regression (Girard et al., 2021, 2022) or time series (Davison et al., 2023) setups.

The problem of estimating extreme expectiles outside of the set of heavy-tailed models is substantially more complicated from a statistical standpoint. The contribution of the present paper is precisely to build and analyze semiparametric extreme expectile estimators in the challenging short-tail model, in which the extreme value index is known to be negative. This requires employing a dedicated extrapolation relationship for population extreme expectiles. Only Mao et al. (2015) have initiated such a study at the population level. Differently from Mao et al. (2015), we first derive an asymptotic expansion of extreme expectiles without resorting to an unnecessary restriction about the link between the extreme value index and second-order parameter as in Mao et al. (2015). Based on this asymptotic expansion, we present and study two different extreme value estimators of tail expectiles. The first one builds upon the Least Asymmetrically Weighted Squares (LAWS) estimator of expectiles, namely the empirical counterpart of ξ_τ in (1.1), obtained at intermediate levels $\tau = \tau_n \to 1$ with $n(1-\tau_n) \to \infty$ as the sample size $n \to \infty$. The short-tail model assumption allows then to come up with an expectile estimator extrapolated to the far tail at arbitrarily extreme levels $\tau = 1 - p_n$ such that $(1-\tau_n)/p_n \to \infty$ as $n \to \infty$, in a semiparametric way. The second extrapolating estimator directly relies on the asymptotic expansion of ξ_τ that involves the expectation of X, the endpoint of its distribution, the quantile analog q_τ and the extreme value index, by plugging in the empirical mean and semiparametric Generalized Pareto (GP) quantile-based estimators of the tail quantities. We also discuss the extension of our LAWS approach to the extremal expectile regression problem, whereby X is assumed to depend upon a finite-dimensional covariate $\boldsymbol{Z}$ and an estimate of the extreme regression expectile $\xi_\tau(\boldsymbol{z})$ of the conditional distribution of X given $\boldsymbol{Z} = \boldsymbol{z}$ is sought. We prove the asymptotic normality of a two-step LAWS estimator built on residuals of location-scale regression models under the assumption that the residuals are sufficiently close (in an appropriate sense) to the unobserved innovations of the model.

Our estimation theory for the extreme expectiles of the marginal distribution of a stationary time series is valid in a general setting of weakly dependent observations, covering very popular time series models such as ARMA models, a wide class of linear time series, and (G)ARCH processes. We explore various theoretical and practical features of extreme expectile estimation for short-tailed data, and explain why this statistical problem is more difficult than extreme quantile estimation. In particular, the extreme expectile ξ_τ is intrinsically less spread than its quantile analog q_τ, even at asymmetry levels $\tau \approx 1$ where it remains much closer to the center of the distribution than q_τ. This implies that any semi-parametric extreme value procedure for extreme expectile estimation should be expected to suffer at least from a worse bias than for extreme quantile estimation. In the extreme expectile regression setup, we show that our residual-based approach is valid in reasonably well-behaved location-scale models such as linear regression models whose error term has a continuously differentiable probability density function on its support. Importantly, this class of location-scale models does not exclude popular time series, in the sense that the covariate vector is allowed to contain lagged values of the response and/or of model error.

The problem of estimating extreme expectiles for bounded distributions appears naturally in the context of productivity analysis. When analyzing the productivity of firms, one may define their economic efficiency in terms of their ability of operating close to the geometric locus of the optimal production that can be viewed as an extreme expectile (Kokic et al., 1997). Another field of application, that we explore here, is the use of extreme expectile estimation for the assessment of long-term market risk carried by short-tailed financial data; long-term risk management makes a crucial use of low-frequency data, including in regulatory circles[1]. Our approach is motivated by the surprising finding that weekly returns of equities, commodities, or cryptocurrencies may indeed have short-tailed distributions, while standard models in both theoretical and empirical work assume heavy tails for any frequency of financial data. This is illustrated in Section 4 where we argue that Bitcoin and Netflix weekly loss returns data may be considered as short-tailed. It would thus be incorrect to assess their tail risk based on the traditional belief that the tail of losses is necessarily heavy when the mathematical theory of extreme values does not allow to reject short or light-tailedness. It is therefore important, both from a theoretical and a practical perspective, to construct an appropriate and fully data-driven estimation procedure for short-tailed data. Our methods and data have been incorporated into the `R` package `ExtremeRisks`, freely available on CRAN.

The paper is organized as follows. In Section 2, we explain in detail the short tail distributional assumption on X, state our asymptotic expansion linking extreme expectiles and quantiles, construct our two classes of extreme expectile estimators and study their asymptotic properties. A simulation study examines their finite-sample performance in Section 3, and two time series of real Bitcoin and Netflix data are analyzed in Section 4. Section A of the Appendix contains all necessary lemmas and mathematical proofs and Section B gives further simulation results.

2 Main results

2.1 Connection between extreme expectiles and quantiles

Let $F : x \mapsto \mathbb{P}(X \leq x)$ be the distribution function of the random variable of interest X and $\overline{F} = 1 - F$ be its survival function. Define the associated quantile function by $q_\tau = \inf\{x \in \mathbb{R} \,|\, F(x) \geq \tau\}$ and the tail quantile function U by $U(s) = q_{1-s^{-1}}$, $s > 1$. Differently from existing literature on extreme expectile estimation, we focus on the case

[1] See, for example, the weekly financial statements of the European Central Bank at https://www.ecb.europa.eu/press/pr/wfs/html/index.en.html.

when the distribution of X is short-tailed, or equivalently, when its extreme value index (EVI) γ is negative. According to Theorem 1.1.6 on p.10 of de Haan and Ferreira (2006), this corresponds to assuming that there is a positive function a such that

$$\forall z > 0, \lim_{s \to \infty} \frac{U(sz) - U(s)}{a(s)} = \frac{z^\gamma - 1}{\gamma}, \text{ with } \gamma < 0.$$

This assumption can be informally rewritten as

$$\forall z > 0, \; U(sz) \approx U(s) + a(s)\frac{z^\gamma - 1}{\gamma} \text{ when } s \text{ is large.} \tag{2.1}$$

This means that extreme values of X at the far tail (represented by $U(sz)$) can be recovered by extrapolating in-sample large values (represented by $U(s)$) if the scale function $a(s)$ and the shape parameter γ can be consistently estimated. The theory of the resulting extreme value estimators is usually developed under the following second-order refinement of the short-tailed model assumption above, which will be our main condition throughout (see de Haan and Ferreira, 2006, Equation (2.3.13) p.45):

Condition $\mathcal{C}_2(\gamma, a, \rho, A)$ There exist $\gamma < 0$, $\rho \le 0$, a positive function $a(\cdot)$ and a measurable function $A(\cdot)$ having constant sign and converging to 0 at infinity such that, for all $z > 0$,

$$\lim_{s \to \infty} \frac{1}{A(s)} \left(\frac{U(sz) - U(s)}{a(s)} - \frac{z^\gamma - 1}{\gamma} \right) = \int_1^z v^{\gamma - 1} \left(\int_1^v u^{\rho - 1} \mathrm{d}u \right) \mathrm{d}v.$$

This condition enables one to control the bias incurred by using the approximation (2.1) and represented by the function A. Under this condition, the right endpoint $x^\star = \sup\{x \in \mathbb{R} \,|\, F(x) < 1\}$ of X is necessarily finite (see de Haan and Ferreira, 2006, Theorem 1.2.1 p.19), that is, X is bounded to the right. This justifies calling this model a short-tailed (or bounded) model.

Suppose now that $\mathbb{E}|\min(X, 0)| < \infty$ and that condition $\mathcal{C}_2(\gamma, a, \rho, A)$ is satisfied, so that $\mathbb{E}|X| < \infty$ and expectiles of X are well-defined and finite. First, we motivate an asymptotic expansion of extreme expectiles that will be instrumental in our subsequent theory of extreme expectile estimation. Recall that the τth expectile ξ_τ satisfies

$$\xi_\tau - \mathbb{E}(X) = \frac{2\tau - 1}{1 - \tau} \mathbb{E}((X - \xi_\tau)\mathbb{1}\{X > \xi_\tau\}), \tag{2.2}$$

see Equation (12) in Bellini et al. (2014). Writing $\mathbb{E}((X - x)\mathbb{1}\{X > x\})$ as an integral of the quantiles of X above x and using condition $\mathcal{C}_2(\gamma, a, \rho, A)$ justifies the approximation

$$\mathbb{E}((X - \xi_\tau)\mathbb{1}\{X > \xi_\tau\}) \approx \frac{\overline{F}(\xi_\tau) a(1/\overline{F}(\xi_\tau))}{1 - \gamma} \quad \text{as } \tau \uparrow 1$$

(see Lemma A.2(ii) in the Appendix for a rigorous statement), and therefore

$$\lim_{\tau \uparrow 1} \frac{a(1/\overline{F}(\xi_\tau))\overline{F}(\xi_\tau)}{1 - \tau} = (x^\star - \mathbb{E}(X))(1 - \gamma). \tag{2.3}$$

The convergence $a(s)/(x^\star - U(s)) \to -\gamma$ as $s \to \infty$ (see de Haan and Ferreira, 2006, Lemma 1.2.9 p.22) then suggests

$$\lim_{\tau \uparrow 1} \frac{(x^\star - \xi_\tau)\overline{F}(\xi_\tau)}{1 - \tau} = (x^\star - \mathbb{E}(X))(1 - \gamma^{-1}). \tag{2.4}$$

The approximations $\overline{F}(\xi_\tau)/(1-\tau) \approx \overline{F}(\xi_\tau)/\overline{F}(q_\tau) \approx (x^\star - \xi_\tau)^{-1/\gamma}/(x^\star - q_\tau)^{-1/\gamma}$ motivated by the regular variation property of $x \mapsto \overline{F}(x^\star - 1/x)$ (see de Haan and Ferreira, 2006, Theorem 1.2.1.2 p.19) finally entail

$$\lim_{\tau \uparrow 1} \frac{x^\star - \xi_\tau}{(x^\star - q_\tau)^{1/(1-\gamma)}} = [(x^\star - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}. \tag{2.5}$$

Consequently, extreme expectiles can be extrapolated from their quantile analogs, in conjunction with endpoint and tail index estimation. Analyzing the asymptotic properties of extreme expectile estimators built in this way will require quantifying the difference between the ratio $(x^\star - \xi_\tau)/(x^\star - q_\tau)^{1/(1-\gamma)}$ and its limit in (2.5). This is the focus of our first main result below.

Proposition 1. *Suppose that* $\mathbb{E}|\min(X, 0)| < \infty$ *and condition* $\mathcal{C}_2(\gamma, a, \rho, A)$ *holds with* $\rho < 0$*, and let* $x^\star$ *be the finite right endpoint of* F*. Then*

$$\begin{aligned} x^\star - \xi_\tau &= [(x^\star - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}(x^\star - q_\tau)^{1/(1-\gamma)} \\ &\quad \times \Bigg(1 - [(x^\star - \mathbb{E}(X))(1-\gamma^{-1})]^{-1/(1-\gamma)}(x^\star - q_\tau)^{1/(1-\gamma)}(1+\mathrm{o}(1)) \\ &\qquad + \frac{\gamma[(x^\star - \mathbb{E}(X))(1-\gamma^{-1})]^{-\rho/(1-\gamma)}}{\rho(\gamma+\rho)(1-\gamma-\rho)} A((1-\tau)^{-1}(x^\star - q_\tau)^{1/(1-\gamma)})(1+\mathrm{o}(1))\Bigg) \end{aligned}$$

as $\tau \uparrow 1$*. In particular*

$$\begin{aligned} x^\star - \xi_\tau &= [(x^\star - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}(x^\star - q_\tau)^{1/(1-\gamma)} \\ &\qquad \times \left(1 + \mathrm{O}((1-\tau)^{-\gamma/(1-\gamma)}) + \mathrm{O}(|A((1-\tau)^{-1/(1-\gamma)})|)\right). \end{aligned}$$

The additional condition $\rho < 0$ in Proposition 1 is very mild and satisfied in all standard short-tailed models, see Beirlant et al. (2004, Table 2.2 p.68). This result is an extension, tailored to our general semiparametric GP setting and extended second-order regular variation assumption $\mathcal{C}_2(\gamma, a, \rho, A)$, of Proposition 3.4 in Mao et al. (2015). The latter result is formulated under a different, nonstandard second-order regular variation condition on $\overline{F}$ when X belongs to the domain of attraction of a Generalized Extreme Value distribution. It is readily checked by straightforward but tedious calculations that their quantities c, γ, ρ and $A(s)$ respectively correspond to (with the notation of Lemma A.3 in Section A.1) $C^{1/\gamma}$, $-1/\gamma$, $-\rho/\gamma$ and $-C^{-\rho/\gamma}A(s^{-1/\gamma})/(\gamma(\gamma+\rho))$ of the present paper. In particular, when their asymptotic expansion applies, it coincides with ours, but we lift an unnecessary restriction on the second-order parameter ρ that features in their result.

An immediate consequence of Equation (2.5) is that $(x^\star - \xi_\tau)/(x^\star - q_\tau) \to \infty$ as $\tau \uparrow 1$, that is, extreme quantiles are closer to the endpoint of a short-tailed distribution than extreme expectiles. It is therefore unsurprising that the bias due to the approximation of tail expectiles by their quantile analogs under the second-order framework, which is asymptotically proportional to $A((1-\tau)^{-1}(x^\star - q_\tau)^{1/(1-\gamma)})$, converges more slowly to 0 than the corresponding bias term in the heavy-tailed setting, whose order is $A((1-\tau)^{-1})$, see Proposition 1(i) in Daouia et al. (2020). As a second consequence, at least as far as handling bias is concerned, estimating extreme expectiles under short-tailed models using a semiparametric extreme value methodology should be expected to be much harder than under heavy-tailed models.

We conclude this section by drawing a useful corollary from Proposition 1; see Equation (A.15) in the proof of the latter result.

Corollary 1. *Under the conditions of Proposition 1,*

$$x^\star - \xi_\tau = [(x^\star - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}(x^\star - q_\tau)^{1/(1-\gamma)}$$
$$\times \left(1 - \frac{x^\star - \xi_\tau}{(x^\star - \mathbb{E}(X))(1-\gamma^{-1})}(1+\mathrm{o}(1)) + \frac{\gamma}{\rho(\gamma+\rho)(1-\gamma-\rho)}A(1/\overline{F}(\xi_\tau))(1+\mathrm{o}(1))\right).$$

It should be noted that the quantity $x^\star - \mathbb{E}(X)$, which is a measure of the spread of the distribution tail, appears in the asymptotic equivalent of $(x^\star - \xi_\tau)/(x^\star - q_\tau)$ and in both of the remainder terms of the asymptotic expansion for $x^\star - \xi_\tau$. By contrast, no measures of spread appear in the asymptotic connection between extreme expectiles and quantiles of heavy-tailed distributions, although the expectation $\mathbb{E}(X)$, which can be understood as a location parameter, appears in an error term proportional to $1/q_\tau$, as can be seen from Proposition 1 in Daouia et al. (2020).

With Proposition 1 and Corollary 1 at our disposal, we can now construct and study two classes of extreme expectile estimators. The first one, in Section 2.2 below, is built upon asymmetric least squares minimization, while the second one, in Section 2.3, is directly obtained by plugging in Equation (2.5) estimators of $\mathbb{E}(X)$ and of the tail quantities γ, $x^\star$ and q_τ.

2.2 Asymmetric least squares estimation

Suppose that the available data has been generated from the random variables $X_1, \ldots, X_n$ with common distribution function F, and let $\tau_n \uparrow 1$ (as $n \to \infty$) be a high asymmetry level at which the target unknown expectile ξ_{τ_n} is to be estimated. A first solution is to construct the estimator minimizing the empirical counterpart of problem (1.1). This produces the Least Asymmetrically Weighted Squares (LAWS) estimator

$$\widehat{\xi}_{\tau_n} = \underset{\theta\in\mathbb{R}}{\arg\min}\, \frac{1}{n}\sum_{t=1}^{n} \eta_{\tau_n}(X_t - \theta) - \eta_{\tau_n}(X_t) = \underset{\theta\in\mathbb{R}}{\arg\min} \sum_{t=1}^{n} \eta_{\tau_n}(X_t - \theta). \tag{2.6}$$

Our theoretical analysis of this estimator hinges upon the following observation made by Jones (1994): the τth expectile of F is actually the τth quantile of the distribution function $E = 1 - \overline{E}$, where

$$\overline{E}(x) = \frac{\mathbb{E}(|X-x|\mathbb{1}\{X > x\})}{\mathbb{E}(|X-x|)}.$$

This survival function can equivalently be rewritten as

$$\overline{E}(x) = \frac{\varphi^{(1)}(x)}{2\varphi^{(1)}(x) + x - \mathbb{E}(X)}, \text{ with } \varphi^{(\kappa)}(x) = \mathbb{E}((X-x)^\kappa \mathbb{1}\{X > x\}).$$

Since $\widehat{\xi}_\tau$ is the τth expectile of the empirical distribution function $\widehat{F}_n = 1 - \widehat{\overline{F}}_n$ defined as

$$\widehat{\overline{F}}_n(x) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{X_t > x\},$$

it must therefore be the τth quantile of the distribution function $\widehat{E}_n = 1 - \widehat{\overline{E}}_n$ defined as

$$\widehat{\overline{E}}_n(x) = \frac{\widehat{\varphi}_n^{(1)}(x)}{2\widehat{\varphi}_n^{(1)}(x) + x - \overline{X}_n}, \text{ where } \widehat{\varphi}_n^{(\kappa)}(x) = \frac{1}{n}\sum_{t=1}^{n}(X_t - x)^\kappa \mathbb{1}\{X_t > x\},$$

with $\overline{X}_n$ being the sample mean. Intuitively, to derive the asymptotic behavior of $\widehat{\xi}_{\tau_n} - \xi_{\tau_n}$, it suffices to obtain the asymptotic behavior of $\widehat{\overline{E}}_n(x)/\overline{E}(x)$ at a level $x = x_n$ close to ξ_{τ_n} in an appropriate sense and to apply a suitable inversion argument.

We do so in a general framework of strictly stationary, weakly dependent random variables. Recall that a strictly stationary sequence $(X_t)_{t\geq 1}$ is said to be $\alpha-$mixing (or strongly mixing) if $\alpha(l) = \sup_{m\geq 1} \alpha_m(l) \to 0$, where

$$\forall l \geq 1,\ \alpha_m(l) = \sup_{\substack{A\in\mathcal{F}_{1,m} \\ B\in\mathcal{F}_{m+l,\infty}}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

with $\mathcal{F}_{1,m} = \sigma(X_1,\ldots,X_m)$ and $\mathcal{F}_{m+l,\infty} = \sigma(X_{m+l}, X_{m+l+1},\ldots)$ denoting the past and future $\sigma-$algebras. The $\alpha-$mixing condition is one of the weakest dependence assumptions in the mixing time series literature: more restrictive conditions include $\beta-$, $\rho-$, $\phi-$ and $\psi-$mixing, see Bradley (2005). We make the following assumption about the mixing rate.

Condition $\mathcal{M}$ There exist sequences of positive integers (l_n) and (r_n), both tending to infinity, such that $l_n/r_n \to 0$, $r_n/n \to 0$ and $n\,\alpha(l_n)/r_n \to 0$, as $n\to\infty$.

The sequences (l_n) and (r_n) are respectively interpreted as "small-block" and "big-block" sequences, and are used to develop a big-block/small-block argument as a prerequisite to evaluating the asymptotic variance of $\widehat{\xi}_{\tau_n}$. Condition $\mathcal{M}$ has already been used in the literature on the extreme values of time series, see *e.g.* Rootzén et al. (1998, Equation (2.1)). We also require the following tail dependence condition on the joint extreme behavior of $(X_t)_{t\geq 1}$ at different time points.

Condition $\mathcal{D}$ For any integer $t\geq 1$, there exists a function R_t on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ such that

$$\forall (x,y)\in [0,\infty]^2 \setminus \{(\infty,\infty)\},\ \lim_{s\to\infty} s\,\mathbb{P}(\overline{F}(X_1)\leq x/s, \overline{F}(X_{t+1})\leq y/s) = R_t(x,y),$$

and there exist a constant $K\geq 0$ and a nonnegative summable sequence $(\rho(t))_{t\geq 1}$ such that, for s large enough,

$$\forall t\geq 1,\ \forall x,y\in (0,1],\ s\,\mathbb{P}(\overline{F}(X_1)\leq x/s, \overline{F}(X_{t+1})\leq y/s) \leq \rho(t)\sqrt{xy} + \frac{K}{s}xy.$$

The function R_t, called the tail copula of (X_1, X_{t+1}) (see Schmidt and Stadtmüller, 2006), finely quantifies the degree of asymptotic dependence between X_1 and X_{t+1}. The first half of Condition $\mathcal{D}$ ensures that the probability of a joint extreme value of X_1 and X_{t+1} is at most of the same order of magnitude as the probability of an extreme value of X_1, meaning that clusters of extreme values across time cannot form too often. The second half of Condition $\mathcal{D}$ guarantees that a variant of the dominated convergence theorem can be applied in correlation calculations prior to using central limit theory for the asymptotic normality of our estimators. A similar anti-clustering assumption is made in Drees (2003), see conditions (C2) and (C3) therein.

Under these temporal dependence assumptions and using our insight about the link between the LAWS estimator and the empirical estimator of $\overline{E}$, we can prove the following result on the joint asymptotic normality of the LAWS estimator and an empirical quantile having the same order of magnitude, *i.e.* an order statistic $\widehat{q}_{\pi_n} = X_{\lceil n\pi_n\rceil,n}$ with $\overline{F}(\xi_{\tau_n})/(1-\pi_n)\to 1$, where $X_{1,n}\leq X_{2,n}\leq\cdots\leq X_{n,n}$ is the ordered version of $(X_1,\ldots,X_n)$.

Theorem 1. *Assume that X satisfies condition $\mathcal{C}_2(\gamma,a,\rho,A)$. Let $\tau_n,\pi_n\uparrow 1$ be such that $n\overline{F}(\xi_{\tau_n})\to\infty$, $\overline{F}(\xi_{\tau_n})/(1-\pi_n)\to 1$ and $\sqrt{n\overline{F}(\xi_{\tau_n})}A(1/\overline{F}(\xi_{\tau_n})) = \mathrm{O}(1)$.*

(i) Suppose that $(X_t)_{t\geq 1}$ is a strictly stationary sequence of copies of X, whose distribution function F is continuous, satisfying conditions $\mathcal{M}$ and $\mathcal{D}$. Assume that $r_n\overline{F}(\xi_{\tau_n}) \to 0$, and that there is $\delta > 0$ such that

$$\mathbb{E}(|\min(X,0)|^{2+\delta}) < \infty, \ \sum_{l\geq 1} l^{2/\delta}\alpha(l) < \infty \text{ and } r_n \left(\frac{r_n}{\sqrt{n\overline{F}(\xi_{\tau_n})}}\right)^{\delta} \to 0.$$

Then

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}, \widehat{q}_{\pi_n} - q_{\pi_n}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{V}(\gamma) + 2\,\boldsymbol{C}(\gamma, R))$$

where the 2×2 symmetric matrices $\boldsymbol{V}(\gamma)$ and $\boldsymbol{C}(\gamma,R)$ are defined elementwise as $V_{11}(\gamma) = 2/[(1-\gamma)(1-2\gamma)]$, $V_{12}(\gamma) = 1/(1-\gamma)$ and $V_{22}(\gamma) = 1$,

$$C_{11}(\gamma,R) = \frac{1}{\gamma^2}\iint_{(0,1]^2}\sum_{t=1}^{\infty} R_t(x^{-1/\gamma}, y^{-1/\gamma})\,\mathrm{d}x\,\mathrm{d}y$$

$$C_{12}(\gamma,R) = -\frac{1}{2\gamma}\int_0^1\sum_{t=1}^{\infty}[R_t(x^{-1/\gamma},1) + R_t(1,x^{-1/\gamma})]\,\mathrm{d}x$$

$$\text{and } C_{22}(\gamma,R) = \sum_{t=1}^{\infty} R_t(1,1).$$

(ii) If the X_i are i.i.d. copies of X and $\mathbb{E}(|\min(X,0)|^2) < \infty$, then the above asymptotic normality result holds with $R_t \equiv 0$ for any $t \geq 1$, that is,

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}, \widehat{q}_{\pi_n} - q_{\pi_n}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{V}(\gamma)).$$

If X is bounded, then assumption $\sum_{l\geq 1} l^{2/\delta}\alpha(l) < \infty$ in (i) can be weakened to $\sum_{l\geq 1}\alpha(l) < \infty$, and no integrability assumption on X is necessary.

In Theorem 1, condition $n\overline{F}(\xi_{\tau_n}) \to \infty$ requires τ_n to be intermediate, *i.e.* not too large. Assumption $\sqrt{n\overline{F}(\xi_{\tau_n})}A(1/\overline{F}(\xi_{\tau_n})) = \mathrm{O}(1)$ is a bias condition which corresponds exactly to the usual bias condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$ in extreme quantile estimation when replacing ξ_{τ_n} with its quantile analog q_{τ_n}, see Theorem 2.4.1 on p.50 in de Haan and Ferreira (2006). In fact, an inspection of the proof of Theorem 1(i) reveals that the bias condition $\sqrt{n\overline{F}(\xi_{\tau_n})}A(1/\overline{F}(\xi_{\tau_n})) = \mathrm{O}(1)$ is only needed for the asymptotic normality of $\widehat{q}_{\pi_n} - q_{\pi_n}$, and is thus unnecessary for the validity of the asymptotic normality of $\widehat{\xi}_{\tau_n} - \xi_{\tau_n}$ alone. The conditions on r_n in Theorem 1(i) are similar to those of Theorem 3.1 in Davison et al. (2023) in heavy-tailed models, taking into account that $\overline{F}(\xi_{\tau_n})$ is asymptotically proportional to $1-\tau_n$ in the latter setting. The integrability assumption on X and the condition on the mixing rate $\alpha(l)$ ensure that a central limit theorem applies to $\overline{X}_n$, as part of the proof of the asymptotic normality of $\widehat{\overline{E}}_n(x)/\overline{E}(x)$ at high levels $x = x_n$ close to ξ_{τ_n}.

It is natural, and instructive, to compare Theorem 1 with results one may obtain in the i.i.d. setting. It is, first of all, obvious that the asymptotic variance $\boldsymbol{V}(\gamma) + 2\,\boldsymbol{C}(\gamma,R)$ obtained in our mixing framework is always greater than or equal to the asymptotic variance $\boldsymbol{V}(\gamma)$ in the i.i.d. setup. This can be viewed as a consequence of positive extremal dependence between bivariate margins of the time series (X_t) under condition $\mathcal{D}$, in the sense that for any $x, y \geq 0$ and $t \geq 1$,

$$s\Big\{\mathbb{P}(\overline{F}(X_1) \leq x/s, \overline{F}(X_{t+1}) \leq y/s) - \mathbb{P}(\overline{F}(X_1) \leq x/s)\mathbb{P}(\overline{F}(X_{t+1}) \leq y/s)\Big\} \to R_t(x,y) \geq 0$$

as $s \to \infty$. This is nothing but a weaker version of the classical positive quadrant dependence assumption between pairs (X_1, X_{t+1}) which itself is a fairly weak assumption on the family of bivariate copulas of these pairs, see the discussion on p.200 in Nelsen (2006). The positive quadrant dependence assumption is satisfied in particular if these copulas are extreme value copulas, see Gudendorf and Segers (2010).

We then compare our result with an asymptotic normality result for intermediate quantile estimation at level τ_n by its direct empirical counterpart $\widehat{q}_{\tau_n} = X_{\lceil n\tau_n \rceil, n}$. According to Theorem 2.4.1 on p.50 in de Haan and Ferreira (2006), when the X_i are i.i.d.,

$$\frac{\sqrt{n\overline{F}(q_{\tau_n})}}{a(1/\overline{F}(q_{\tau_n}))}(\widehat{q}_{\tau_n} - q_{\tau_n}) \xrightarrow{d} \mathcal{N}(0,1).$$

Observe that, by a combination of Lemma 1.2.9 on p.22 in de Haan and Ferreira (2006) and Lemma A.1 in Section A.1, $a(1/\overline{F}(x))/(x^\star - x) \to -\gamma$ as $x \uparrow x^\star$, and therefore

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} \Bigg/ \frac{\sqrt{n\overline{F}(q_{\tau_n})}}{a(1/\overline{F}(q_{\tau_n}))} = \sqrt{\frac{\overline{F}(\xi_{\tau_n})}{\overline{F}(q_{\tau_n})}} \times \frac{x^\star - q_{\tau_n}}{x^\star - \xi_{\tau_n}}(1 + \mathrm{o}(1)).$$

By (2.4) and (2.5) this ratio is asymptotically proportional to $(x^\star - q_{\tau_n})^{-(\gamma+1/2)/(1-\gamma)}$ under the mild further condition $\rho < 0$. In other words, the intermediate LAWS estimator $\widehat{\xi}_{\tau_n}$ converges faster than $\widehat{q}_{\tau_n}$ when $\gamma > -1/2$, has the same rate of convergence if $\gamma = -1/2$, and converges at a slower rate if $\gamma < -1/2$.

One may also compare Theorem 1, devoted to short-tailed data, with the corresponding result one obtains for i.i.d. heavy-tailed data. If X has a heavy right tail, that is, $U(sz)/U(s) \to z^\gamma$ as $s \to \infty$ for any $z > 0$, where $0 < \gamma < 1/2$, and under the assumptions that $\mathbb{E}(|\min(X,0)|^{2+\delta}) < \infty$ for some $\delta > 0$, $\tau_n \uparrow 1$ and $n(1-\tau_n) \to \infty$, one has, by Theorem 2 in Daouia et al. (2018),

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma^3}{1-2\gamma}\right).$$

In this same setting, $(U(sz) - U(s))/a(s) \to (z^\gamma - 1)/\gamma$ as $s \to \infty$, with $a(s) = \gamma U(s)$, and $\overline{F}(\xi_{\tau_n})/(1-\tau_n) \to \gamma^{-1} - 1 = (1-\gamma)/\gamma$ (this was first shown by Bellini et al., 2014, Theorem 11). Therefore, when X has a heavy right tail,

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \approx \frac{\sqrt{1-\gamma}}{\gamma^{3/2}} \times \sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2(1-\gamma)}{1-2\gamma}\right).$$

It follows that the rates of convergence of the LAWS estimator look similar in both the heavy and bounded tail settings, but there is a phase transition in terms of asymptotic variance: the term $1-\gamma$ appears in its numerator for heavy tails, while it appears in the denominator for short tails, as established in Theorem 1. Interestingly, the two asymptotic variances in the heavy and short-tailed settings converge to 2, and therefore exactly match in the light-tailed scenario, when $\gamma \to 0$.

As a corollary of Theorem 1, we obtain the asymptotic normality of the empirical estimator $\widehat{\overline{F}}_n(\widehat{\xi}_{\tau_n})$ of $\overline{F}(\xi_{\tau_n})$, on which the rate of convergence of $\widehat{\xi}_{\tau_n}$ crucially depends.

Corollary 2. *Work under the conditions of Theorem 1. Then*

$$\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\widehat{\overline{F}}_n(\widehat{\xi}_{\tau_n})}{\overline{F}(\xi_{\tau_n})}-1\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{2\gamma^2+\gamma+1}{(1-\gamma)(1-2\gamma)}\right.$$
$$\left.+2\iint_{(0,1]^2}\sum_{t=1}^{\infty}\left(\frac{1}{\gamma^2}R_t(x^{-1/\gamma},y^{-1/\gamma})+\frac{1}{\gamma}[R_t(x^{-1/\gamma},1)+R_t(1,x^{-1/\gamma})]+R_t(1,1)\right)\mathrm{d}x\,\mathrm{d}y\right).$$

The rate of convergence of $\widehat{\overline{F}}_n(\widehat{\xi}_{\tau_n})$ is rather natural: for a sequence (u_n) tending to $x^\star$ such that $n\overline{F}(u_n)\to\infty$, Lemma A.5 states that

$$\sqrt{n\overline{F}(u_n)}\left(\frac{\widehat{\overline{F}}_n(u_n)}{\overline{F}(u_n)}-1\right)\xrightarrow{d}\mathcal{N}\left(0,1+2\sum_{t=1}^{\infty}R_t(1,1)\right).$$

It is worth noticing that the asymptotic variance of $\widehat{\overline{F}}_n(\widehat{\xi}_{\tau_n})$ does not coincide with the variance that would be obtained if ξ_{τ_n} were known, namely, if $\widehat{\overline{F}}_n(\xi_{\tau_n})$ were considered instead. This is due to the asymptotic dependence existing between $\widehat{\xi}_{\tau_n}$ and high order statistics of the sample (and therefore between $\widehat{\xi}_{\tau_n}$ and $\widehat{\overline{F}}_n$), see Theorem 1 and the proof of Corollary 2.

We now have the tools necessary to construct an extreme value estimator of a properly extreme expectile ξ_{1-p_n}, where $p_n\downarrow 0$ at any possible rate as $n\to\infty$. Recall, first of all, that condition $n\overline{F}(\xi_{\tau_n})\to\infty$, ensuring that $\widehat{\xi}_{\tau_n}$ is an asymptotically normal estimator of ξ_{τ_n}, requires τ_n to be intermediate. In particular, $\widehat{\xi}_{1-p_n}$ is not going to be an asymptotically normal estimator of ξ_{1-p_n} whatever the choice of $p_n\downarrow 0$ is; the construction of an appropriate estimator for ξ_{1-p_n} with p_n arbitrarily close to 1 requires extrapolating the intermediate LAWS estimator $\widehat{\xi}_{\tau_n}$ using the extreme value condition $\mathcal{C}_2(\gamma,a,\rho,A)$. Using (2.1) with $s=1/\overline{F}(\xi_{\tau_n})$ and $z=\overline{F}(\xi_{\tau_n})/\overline{F}(\xi_{1-p_n})$ motivates the approximation

$$\xi_{1-p_n}\approx\xi_{\tau_n}+a(1/\overline{F}(\xi_{\tau_n}))\frac{(\overline{F}(\xi_{\tau_n})/\overline{F}(\xi_{1-p_n}))^\gamma-1}{\gamma}.$$

By Theorem 1, ξ_{τ_n} is estimated by the LAWS estimator $\widehat{\xi}_{\tau_n}$ at rate $a(1/\overline{F}(\xi_{\tau_n}))/\sqrt{n\overline{F}(\xi_{\tau_n})}$. The scale parameter $a(1/\overline{F}(\xi_{\tau_n}))$ and shape parameter γ can be estimated by a variety of techniques such as:

- The (pseudo-)Generalized Pareto maximum likelihood (GPML) estimators, that is, if $k=k_n\to\infty$ is a sequence of integers such that $k/n\to 0$,

$$(\widehat{a}^{\mathrm{ML}}(n/k),\widehat{\gamma}_n^{\mathrm{ML}})=\underset{\sigma>0,\,\gamma>-1/2}{\arg\max}\prod_{i=1}^{k}h(X_{n-i+1,n}-X_{n-k,n}|\sigma,\gamma)$$

  where the GP probability density function $h(\cdot|\sigma,\gamma)$ is defined as

$$h(x|\sigma,\gamma)=\frac{1}{\sigma}\left(1+\frac{\gamma x}{\sigma}\right)^{-1/\gamma-1}\text{ for all }x>0\text{ with }1+\frac{\gamma x}{\sigma}>0.$$

- The Moment-type estimators of Dekkers et al. (1989), defined as

$$(\widehat{a}^{\mathrm{Mom}}(n/k),\widehat{\gamma}_n^{\mathrm{Mom}})=(X_{n-k,n}M_k^{(1)}(1-\widehat{\gamma}_n^{(-)}),M_k^{(1)}+\widehat{\gamma}_n^{(-)})$$

where

$$\widehat{\gamma}_n^{(-)} = 1 - \frac{1}{2}\left(1 - \frac{(M_k^{(1)})^2}{M_k^{(2)}}\right)^{-1}$$

$$\text{and} \quad M_k^{(j)} = \frac{1}{k}\sum_{i=1}^{k}(\log X_{n-i+1,n} - \log X_{n-k,n})^j, \quad \text{for} \quad j = 1, 2.$$

Typically, estimators of the scale function $a(1/\overline{F}(u_n))$ converge on the relative scale at the rate $1/\sqrt{n\overline{F}(u_n)}$ when $u_n \uparrow x^\star$ is such that $n\overline{F}(u_n) \to \infty$; see Sections 3.4 and 4.2 in de Haan and Ferreira (2006) in the i.i.d. case, and Section 6 in Drees (2003) in the dependent data setup. Since, by Corollary 2, the (unknown) quantity $\overline{F}(u_n) = \overline{F}(\xi_{\tau_n})$ can be consistently estimated at the rate $1/\sqrt{n\overline{F}(\xi_{\tau_n})}$, we therefore expect to be able to estimate $a(1/\overline{F}(\xi_{\tau_n}))$ at this rate on the relative scale. Finally, given an intermediate level τ_n, it is customary to estimate the extreme value index γ at the rate $1/\sqrt{n(1-\tau_n)}$ when the top $k = \lfloor n(1-\tau_n)\rfloor$ values in the data are used, see Sections 3.3, 3.4, 3.5 and 3.6 in de Haan and Ferreira (2006) in the i.i.d. case, and again Section 6 in Drees (2003) when the data points are serially dependent. It remains to find a way to estimate $\overline{F}(\xi_{\tau_n})/\overline{F}(\xi_{1-p_n})$, which depends on the target quantity ξ_{1-p_n} itself. A combination of Equations (2.4) and (2.5) with the fact that the function $s \mapsto x^\star - U(s)$ is regularly varying with index γ (see de Haan and Ferreira, 2006, Corollary 1.2.10 p.23) suggests that

$$\begin{aligned}\frac{\overline{F}(\xi_{\tau_n})}{\overline{F}(\xi_{1-p_n})} &\approx \frac{1-\tau_n}{p_n} \times \frac{x^\star - \xi_{1-p_n}}{x^\star - \xi_{\tau_n}} \approx \frac{1-\tau_n}{p_n}\left(\frac{x^\star - q_{1-p_n}}{x^\star - q_{\tau_n}}\right)^{1/(1-\gamma)} \\ &\approx \frac{1-\tau_n}{p_n}\left(\frac{1-\tau_n}{p_n}\right)^{\gamma/(1-\gamma)} = \left(\frac{1-\tau_n}{p_n}\right)^{1/(1-\gamma)} \qquad (2.7)\end{aligned}$$

which in turn leads to the expectile-specific approximation

$$\xi_{1-p_n} \approx \xi_{\tau_n} + a(1/\overline{F}(\xi_{\tau_n}))\frac{((1-\tau_n)/p_n)^{\gamma/(1-\gamma)} - 1}{\gamma}.$$

Consequently, like extreme quantiles, extreme expectiles can be extrapolated from their values at lower levels, but their values are not (in the appropriate sense) asymptotically equivalent to those of intermediate expectiles, since

$$\frac{\sqrt{n(1-\tau_n)}}{a(1/\overline{F}(\xi_{\tau_n}))}(\xi_{1-p_n} - \xi_{\tau_n}) \to +\infty$$

when $n(1-\tau_n) \to \infty$ and $p_n/(1-\tau_n) \to 0$. Given estimators $\widehat{\sigma}_n$ and $\widehat{\gamma}_n$ of $a(1/\overline{F}(\xi_{\tau_n}))$ and γ, respectively, one can then construct the following estimator of ξ_{1-p_n}:

$$\widehat{\xi}^\star_{1-p_n} = \widehat{\xi}_{\tau_n} + \widehat{\sigma}_n\frac{((1-\tau_n)/p_n)^{\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - 1}{\widehat{\gamma}_n}. \qquad (2.8)$$

Since $(1-\tau_n)/\overline{F}(\xi_{\tau_n}) \to 0$, the parameter γ is estimated at a slower rate than the other quantities, so we expect the asymptotic behavior of $\widehat{\gamma}_n$ to govern that of $\widehat{\xi}^\star_{1-p_n}$. The last theorem of this section makes this intuition rigorous. Its proof crucially relies on Theorem 1 and on Proposition 1 in order to quantify the bias in the approximation (2.7).

Theorem 2. *Work under the conditions of Theorem 1. If moreover $\rho < 0$, $n(1-\tau_n) \to \infty$, $(1-\tau_n)/p_n \to \infty$, $\sqrt{n(1-\tau_n)}/\log((1-\tau_n)/p_n) \to \infty$, $\sqrt{n(1-\tau_n)}(x^\star - q_{\tau_n})^{1/(1-\gamma)} = \mathrm{O}(1)$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$, $\widehat{\sigma}_n$ and $\widehat{\gamma}_n$ are such that*

$$\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\widehat{\sigma}_n}{a(1/\overline{F}(\xi_{\tau_n}))} - 1\right) = \mathrm{O}_{\mathbb{P}}(1) \quad \text{and} \quad \sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma) \xrightarrow{d} \Gamma,$$

where Γ is a nondegenerate limit, then

$$\frac{\sqrt{n(1-\tau_n)}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}^\star_{1-p_n} - \xi_{1-p_n}) \xrightarrow{d} \frac{\Gamma}{\gamma^2}.$$

2.3 Quantile-based estimation

We use here Proposition 1 to present an alternative estimator of extreme expectiles, purely based on quantiles, and to develop its asymptotic theory. Similarly to the setup of extreme quantile estimation in Section 4.3 of de Haan and Ferreira (2006), assume that $k = k_n \to \infty$ is a sequence of positive integers such that $k/n \to 0$ and that estimators $\widehat{\gamma}_n$, $\widehat{a}(n/k)$ and $X_{n-k,n}$ of γ, $a(n/k)$ and $U(n/k)$, respectively, are given such that

$$\sqrt{k}\left(\widehat{\gamma}_n - \gamma, \frac{\widehat{a}(n/k)}{a(n/k)} - 1, \frac{X_{n-k,n} - U(n/k)}{a(n/k)}\right) \xrightarrow{d} (\Gamma, \Lambda, B) \tag{2.9}$$

where (Γ, Λ, B) is a nontrivial trivariate weak limit. This assumption is satisfied by the moment and Generalized Pareto maximum likelihood (GPML) estimators of the shape and scale parameters presented in Section 2.2, among others, see an overview in Section 4.3 of de Haan and Ferreira (2006) in the case where the X_i are independent random variables. It is also satisfied when $(X_t)_{t\geq 1}$ is a strictly stationary but serially dependent sequence: this is for example the case when the data points are $\beta-$mixing and satisfy an anti-clustering condition similar to the tail dependence assumption $\mathcal{D}$, as a consequence of the powerful results of Drees (2003).

Let $p_n \downarrow 0$ with $k/(np_n) \to \infty$, so that the level $1 - p_n$ is much more extreme than $1 - k/n$. Following Section 4.3 of de Haan and Ferreira (2006), the extreme quantile q_{1-p_n} and the right endpoint $x^\star$ can be estimated by

$$\widehat{q}^\star_{1-p_n} = X_{n-k,n} + \widehat{a}(n/k)\frac{(k/(np_n))^{\widehat{\gamma}_n} - 1}{\widehat{\gamma}_n} \quad \text{and} \quad \widehat{x}^\star = X_{n-k,n} - \frac{\widehat{a}(n/k)}{\widehat{\gamma}_n}. \tag{2.10}$$

According to Proposition 1, an estimator of ξ_{1-p_n} follows then as

$$\widetilde{\xi}^\star_{1-p_n} = \widehat{x}^\star - [(\widehat{x}^\star - \overline{X}_n)(1 - \widehat{\gamma}_n^{-1})]^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)}(\widehat{x}^\star - \widehat{q}^\star_{1-p_n})^{1/(1-\widehat{\gamma}_n)}. \tag{2.11}$$

The next result provides its asymptotic properties, where a sequence (u_n) is said to be asymptotically proportional to another sequence (v_n) if (u_n/v_n) converges to a finite positive limit as $n \to \infty$.

Theorem 3. *Suppose that $\mathbb{E}|\min(X,0)| < \infty$ and condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds with $\rho < 0$. Assume that condition (2.9) holds true and that $\sqrt{k}(\overline{X}_n - \mathbb{E}(X)) \xrightarrow{\mathbb{P}} 0$ with $k = k_n$ being asymptotically proportional to n^χ, for some $\chi \in (0,1)$. Let p_n be asymptotically proportional to $n^{-\omega}$ where $\omega > 0$ is such that $\chi + \omega - 1 > 0$. If moreover $\sqrt{k}A(n/k) \to$*

$\lambda \in \mathbb{R}$, *then we have, up to changing probability spaces and with appropriate versions of the estimators involved,*

$$\begin{aligned}\widetilde{\xi}^{\star}_{1-p_n} - \xi_{1-p_n} &= \frac{a(n/k)}{\sqrt{k}}\frac{1}{\gamma^2}\left(\Gamma + \gamma^2 B - \gamma\Lambda - \lambda\frac{\gamma}{\gamma+\rho} + \mathrm{o}_{\mathbb{P}}(1)\right) \\ &+ [a(n/k)(k/(np_n))^{\gamma}]^{1/(1-\gamma)}(-\gamma)^{-1}(1-\gamma)^{-1/(1-\gamma)}(x^\star - \mathbb{E}(X))^{-\gamma/(1-\gamma)} \\ &\times \left(\frac{\log(np_n^{1/(1-\gamma)}/k)}{\sqrt{k}}\Gamma + \mathrm{o}_{\mathbb{P}}\left(\frac{\log n}{\sqrt{k}}\right)\right) \\ &+ \mathrm{O}(n^{\omega\gamma/(1-\gamma)}(n^{\omega\gamma/(1-\gamma)} + |A(n^{\omega/(1-\gamma)})|)).\end{aligned}$$

Let us discuss the assumptions made in Theorem 3. Condition $\sqrt{k}(\overline{X}_n - \mathbb{E}(X)) \xrightarrow{\mathbb{P}} 0$ is satisfied in practice if $\sqrt{n}(\overline{X}_n - \mathbb{E}(X)) = \mathrm{O}_{\mathbb{P}}(1)$, which is in particular true when a central limit theorem applies. As already highlighted below Theorem 1, this will be the case if $\mathbb{E}(|\min(X,0)|^2) < \infty$ when the X_i are independent, or if there is $\delta > 0$ such that $\mathbb{E}(|\min(X,0)|^{2+\delta}) < \infty$ and $\sum_{l\geq 1} l^{2/\delta}\alpha(l) < \infty$ when $(X_t)_{t\geq 1}$ is $\alpha-$mixing. In particular, when the data mixes geometrically fast, then $\sqrt{n}(\overline{X}_n - \mathbb{E}(X)) = \mathrm{O}_{\mathbb{P}}(1)$ as soon as X has a finite moment of order $2+\delta$, for some $\delta > 0$. Besides, the assumption that $k = k_n$ is asymptotically equivalent to a positive and finite multiple of n^{χ}, is only very slightly stronger than the usual pair of extreme value conditions $k \to \infty$ and $k/n \to 0$. The only difference is that our assumption does not allow to take k growing to infinity logarithmically fast; such sequences produce, however, very small values of k in practice and would therefore yield estimators having very large variances. We also note that in standard settings such as those of Beirlant et al. (2004, Table 2.2 p.68), $A(s)$ is asymptotically proportional to s^{ρ}, in which case the optimal choices of k satisfying the usual bias-variance tradeoff for extreme value index estimation would fulfill $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}\setminus\{0\}$, that is, k should be asymptotically proportional to $n^{-2\rho/(1-2\rho)}$. In other words, it is reasonable to expect that optimal choices of k in practice have to be asymptotically equivalent to a positive and finite multiple of a fractional power of n.

It follows from Theorem 3 that the asymptotic behavior of the extreme expectile estimator $\widetilde{\xi}^{\star}_{1-p_n}$ is more complex than that of the extreme quantile estimator $\widehat{q}^{\star}_{1-p_n}$: while, from Theorem 4.3.1 on p.134 and Theorem 4.5.1 on p.146 of de Haan and Ferreira (2006), $\widehat{q}^{\star}_{1-p_n} - q_{1-p_n}$ converges to the same distribution $\frac{1}{\gamma^2}\left(\Gamma + \gamma^2 B - \gamma\Lambda - \lambda\frac{\gamma}{\gamma+\rho}\right)$ as $\widehat{x}^\star - x^\star$ at the rate $a(n/k)/\sqrt{k}$ for $\gamma < 0$, the asymptotic distribution of $\widetilde{\xi}^{\star}_{1-p_n} - \xi_{1-p_n}$ may be a nonstandard mixture of the two limiting distributions of $\widehat{x}^\star - x^\star$ and $\widehat{\gamma}_n - \gamma$. In particular, Corollary 3 shows that when, for example, $\omega = 1$ (containing the typical setting $p_n = 1/n$) and χ (and hence k) is chosen small enough, it is in fact the asymptotic distribution Γ of $\widehat{\gamma}_n - \gamma$ that dominates in $\widetilde{\xi}^{\star}_{1-p_n} - \xi_{1-p_n}$, while Corollary 4 examines what can otherwise be said.

Corollary 3. *Under the assumptions of Theorem 3, if moreover* $\chi < 1 - \omega/(1-\gamma)$ *and* $\chi < 2\omega\min(-\gamma, -\rho)/(1-\gamma)$, *then*

$$\frac{\sqrt{k}}{\log(np_n^{1/(1-\gamma)}/k)}\frac{\widetilde{\xi}^{\star}_{1-p_n} - \xi_{1-p_n}}{[a(n/k)(k/(np_n))^{\gamma}]^{1/(1-\gamma)}} \xrightarrow{d} -\gamma^{-1}(1-\gamma)^{-1/(1-\gamma)}(x^\star - \mathbb{E}(X))^{-\gamma/(1-\gamma)}\Gamma.$$

One may then compare the rates of convergence of $\widehat{\xi}^{\star}_{1-p_n}$ and $\widetilde{\xi}^{\star}_{1-p_n}$ by setting $\tau_n = 1 - k/n$. Using the convergence $a(s)/(x^\star - U(s)) \to -\gamma$ as $s \to \infty$ and Equation (2.5), one finds under the assumptions of Corollary 3 that

$$\frac{\sqrt{k}}{a(1/\overline{F}(\xi_{1-k/n}))} \Bigg/ \frac{\sqrt{k}}{\log(np_n^{1/(1-\gamma)}/k)[a(n/k)(k/(np_n))^{\gamma}]^{1/(1-\gamma)}}$$
$$\propto \log(n)\frac{[a(n/k)]^{1/(1-\gamma)}}{a(1/\overline{F}(\xi_{1-k/n}))}(k/(np_n))^{\gamma/(1-\gamma)} \propto n^{(\chi+\omega-1)\gamma/(1-\gamma)}\log(n) \to 0.$$

This means that $\widetilde{\xi}^{\star}_{1-p_n}$ converges to ξ_{1-p_n} faster than $\widehat{\xi}^{\star}_{1-p_n}$ when k (or $1-\tau_n$) is chosen sufficiently small. We shall illustrate this finding below in our simulation study. It is also interesting to note that the closer γ is to 0, or equivalently, the closer the data-generating distribution is to having a light tail, the stronger the constraint on χ through the condition $\chi < 2\omega\min(-\gamma,-\rho)/(1-\gamma)$. This is analogous to what happens in extreme expectile estimation for heavy-tailed distributions, where the condition $\sqrt{k}/q_{1-k/n} = \mathrm{O}(1)$ (see *e.g.* Daouia et al., 2020, Theorem 5) becomes a strong restriction as the tail gets less heavy, *i.e.* when γ approaches 0.

Condition $\chi < 1-\omega/(1-\gamma)$ may not hold in a given example, especially when ω is large enough, or equivalently, p_n is small enough. Yet, interestingly this condition can always be satisfied for sufficiently small χ in the standard setting $\omega = 1$ of extreme value analysis. If it is not satisfied, then $\widetilde{\xi}^{\star}_{1-p_n}$ tends to inherit the asymptotic behavior of $\widehat{x}^{\star}$, rather than $\widehat{\gamma}$, as established in the following result.

Corollary 4. *Under the assumptions of Theorem 3, if moreover* $\chi > 1-\omega/(1-\gamma)$, *then*

$$\widetilde{\xi}^{\star}_{1-p_n} - \xi_{1-p_n} = \frac{a(n/k)}{\sqrt{k}}\frac{1}{\gamma^2}\left(\Gamma + \gamma^2 B - \gamma\Lambda - \lambda\frac{\gamma}{\gamma+\rho} + \mathrm{o}_{\mathbb{P}}(1)\right)$$
$$+ \mathrm{O}(n^{\omega\gamma/(1-\gamma)}(n^{\omega\gamma/(1-\gamma)} + |A(n^{\omega/(1-\gamma)})|)).$$

It is important to note that the condition $\chi > 1-\omega/(1-\gamma)$ itself is not sufficient to ensure the convergence of $\widetilde{\xi}^{\star}_{1-p_n}$; in practice, depending on the choice of k, the bias term may dominate the asymptotics. This is most easily seen when $A(s)$ is asymptotically proportional to s^{ρ} and $\omega = 1$, corresponding to the standard extreme value situation where $p_n \approx c/n$. In this case:

- One automatically has $\chi + \omega - 1 = \chi > 0$,
- Condition $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$ essentially amounts to $\chi \le -2\rho/(1-2\rho)$,
- Condition $\chi > 1-\omega/(1-\gamma)$ becomes $\chi > -\gamma/(1-\gamma)$.

For the bias term in Corollary 4 to be negligible, one requires

$$\frac{\sqrt{k}}{a(n/k)} \times n^{\gamma/(1-\gamma)}(n^{\gamma/(1-\gamma)} + |A(n^{1/(1-\gamma)})|) \to 0.$$

Since $a(n/k)$ is asymptotically proportional to $(n/k)^{\gamma}$ by Lemma A.3(i), this is equivalent to assuming

$$\chi\left(\frac{1}{2}+\gamma\right) + \frac{\gamma^2 - \min(-\gamma,-\rho)}{1-\gamma} < 0.$$

When $\gamma > -1/2$, which is a case often encountered in practical applications, and $0 < -\rho < -\gamma$, representing situations where the bias due to the second-order framework is high, this condition becomes

$$\chi < -\frac{2(\gamma^2+\rho)}{(1-\gamma)(1+2\gamma)}.$$

Depending on the value of ρ, this final condition may not be compatible with $\chi > -\gamma/(1-\gamma)$: in fact, if ρ is close enough to 0, it may even be impossible to satisfy whatever the value of χ (since the right-hand side of the above displayed inequality tends to a negative constant as $\rho \to 0$, when $\gamma > -1/2$). In this case, with the choice $p_n = c/n$, the asymptotic behavior of $\widehat{x}^\star - x^\star$ can never dominate in $\widetilde{\xi}^\star_{1-p_n} - \xi_{1-p_n}$.

2.4 Selection of the expectile asymmetry level

In practical situations it is crucial to make an informed decision as to what the asymmetry level of the target expectile should be. In financial applications, where the dual interpretation of expectiles in terms of the gain-loss ratio is available (Bellini and Di Bernardino, 2017), it is sensible to set the expectile level so as to achieve a certain value of the gain-loss ratio. Otherwise, it has been proposed in the literature to select τ such that ξ_τ coincides with another pre-specified intuitive risk measure: Bellini and Di Bernardino (2017) suggest to choose the expectile level τ so that ξ_τ is identical to the Value-at-Risk (or quantile) q_π, where π is a high probability level specified by the statistician or the practitioner.

The proposal of Bellini and Di Bernardino (2017) is valid only when the underlying loss distribution is Gaussian. Daouia et al. (2018) later extended this idea to the heavy-tailed setup. We examine here the short-tailed situation, hitherto unexplored from this perspective. Fix a large quantile level $1-p_n$. Setting $\tau = \tau_n$ to be such that $\xi_\tau = q_{1-p_n}$, Equation (2.4) leads to

$$\frac{(x^\star - q_{1-p_n})p_n}{1-\tau_n} = \frac{(x^\star - \xi_\tau)\overline{F}(\xi_\tau)}{1-\tau} \approx (x^\star - \mathbb{E}(X))(1-\gamma^{-1}).$$

In other words,

$$1-\tau_n \approx \frac{x^\star - q_{1-p_n}}{(x^\star - \mathbb{E}(X))(1-\gamma^{-1})} p_n.$$

This approximation suggests to estimate the quantity τ_n by

$$\widehat{\tau}_n \equiv \widehat{\tau}_n(p_n) = 1 - \frac{\widehat{x}^\star - \widehat{q}^\star_{1-p_n}}{(\widehat{x}^\star - \overline{X}_n)(1-\widehat{\gamma}_n^{-1})} p_n$$

with the notation of (2.10). Our next main result shows that, under suitable bias conditions, this estimator converges at the rate $\log(k/(np_n))/\sqrt{k}$ in the framework of Section 2.3.

Proposition 2. *Under the assumptions of Theorem 3, if moreover* $\chi < \min(-2\omega\gamma, -2\rho/(1-2\rho))$*, then*

$$\frac{\sqrt{k}}{\log(k/(np_n))}\left(\frac{1-\widehat{\tau}_n}{1-\tau_n} - 1\right) \xrightarrow{d} \Gamma.$$

It is worth noting that in the standard setting when $p_n \approx c/n$, for a positive constant c, the bias condition $\chi < \min(-2\omega\gamma, -2\rho/(1-2\rho))$ will always be satisfied provided k is chosen small enough. In contrast to Proposition 2, in the heavy-tail setting and according to Section 5 in Daouia et al. (2018), it holds that $1-\tau_n \approx p_n/(\gamma^{-1}-1)$. An estimator of τ_n is then $\widehat{\tau}_n = 1 - p_n/(\widehat{\gamma}_n^{-1} - 1)$. In this setting, it is straightforward to obtain, under a suitable bias condition when $\sqrt{k}(\widehat{\gamma}_n - \gamma) \to \Gamma$, that

$$\sqrt{k}\left(\frac{1-\widehat{\tau}_n}{1-\tau_n} - 1\right) \xrightarrow{d} \frac{\Gamma}{\gamma(1-\gamma)}.$$

The estimator $1-\widehat{\tau}_n$ therefore converges at a slightly faster rate in the heavy-tailed model. The slower speed of convergence in the short-tailed framework is due to the presence of the quantity $\widehat{x}^\star - \widehat{q}^\star_{1-p_n} = -\widehat{a}(n/k)\widehat{\gamma}_n^{-1}(k/(np_n))^{\widehat{\gamma}_n}$ in the numerator of $1-\widehat{\tau}_n$, whose rate of convergence to $x^\star - q_{1-p_n}$ is precisely $\log(k/(np_n))/\sqrt{k}$.

2.5 Towards conditional and dynamic extreme expectile estimation

The tail behavior of X can often be better understood by estimating its conditional extremes given a relevant finite-dimensional covariate $\boldsymbol{Z} \in \mathbb{R}^d$. In financial applications, this covariate can contain, among others, global market information (through current global market index values, for instance) as well as lags of the target variable, in order to allow for dynamic prediction of future extreme risk levels given past financial information. Many popular models used in statistical practice for this purpose are particular examples of explicit regression models linking X_t to $\boldsymbol{Z}_t$ through the formula $X_t = m(\boldsymbol{Z}_t) + \varepsilon_t$, where $m(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is an unknown measurable function to be estimated and ε_t is an unobserved innovation. When the model is correctly specified, as we shall assume here, the $\varepsilon_t = X_t - m(\boldsymbol{Z}_t)$ are typically independent and identically distributed copies of a centered random variable ε with survival function $\overline{F}$, and for each t, ε_t is independent of $\boldsymbol{Z}_t$.

It follows from this model assumption that a conditional expectile $\xi_\tau(X_t|\boldsymbol{Z}_t = \boldsymbol{z})$ can be written as $\xi_\tau(X_t|\boldsymbol{Z}_t = \boldsymbol{z}) = m(\boldsymbol{z}) + \xi_\tau(\varepsilon)$. A reasonable idea in order to estimate the extreme conditional expectile $\xi_\tau(X_t|\boldsymbol{Z}_t = \boldsymbol{z})$, for $\tau = \tau_n \uparrow 1$, from data $(\boldsymbol{Z}_1, X_1), \ldots, (\boldsymbol{Z}_n, X_n)$ generated from this model, is to estimate first the regression function $m(\boldsymbol{z})$ and then the tail unconditional expectile $\xi_\tau(\varepsilon)$, using residuals $\widehat{\varepsilon}_t^{(n)}$ of the model instead of the unobserved ε_t. This eventually results in a two-step estimator of $\xi_{\tau_n}(X_t|\boldsymbol{Z}_t = \boldsymbol{z})$. The crucial difficulty, of course, is that these residuals, unlike the true unobserved innovations, will typically not be independent or even identically distributed, even in simple models such as those concerned with linear regression. However, since at least in parametric regression models one is typically able to estimate the function m in a straightforward fashion at the rate $1/\sqrt{n}$, which is faster than the rate of convergence of the extreme-value step, one should expect the estimator of $\xi_{\tau_n}(\varepsilon)$ based on residuals to behave asymptotically just like its unachievable true error-based counterpart.

An interesting question is therefore to consider whether this intuition is indeed correct under a reasonable condition. Our final asymptotic result goes in this direction.

Theorem 4. *Assume that the centered random variable ε satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$, with $\mathbb{E}(|\min(\varepsilon, 0)|^2) < \infty$ and right endpoint $e^\star > 0$, and let $\xi_\tau = \xi_\tau(\varepsilon)$ be its τth expectile. Let $\tau_n \uparrow 1$ be such that $n\overline{F}(\xi_{\tau_n}) \to \infty$. If the $\widehat{\varepsilon}_t^{(n)}$ satisfy*

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} \max_{1 \le t \le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| \xrightarrow{\mathbb{P}} 0, \tag{2.12}$$

then the LAWS residual-based estimator $\widehat{\xi}_{\tau_n} = \arg\min_{\theta \in \mathbb{R}} \sum_{t=1}^n \eta_{\tau_n}(\widehat{\varepsilon}_t^{(n)} - \theta)$ is such that

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} (\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \xrightarrow{d} \mathcal{N}\left(0, \frac{2}{(1-\gamma)(1-2\gamma)}\right).$$

The key condition to be checked as part of Theorem 4 is convergence (2.12), which essentially expresses that the model has to be estimated faster than the rate of the extreme-value procedure applied to the innovation term for the residual-based intermediate estimator to have the required asymptotic normality property. For example, in the linear model $X_t = \boldsymbol{Z}_t^\top \boldsymbol{\beta} + \varepsilon_t$, where $\boldsymbol{\beta} \in \mathbb{R}^d$ is estimated by $\widehat{\boldsymbol{\beta}}_n$, $|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| = |\boldsymbol{Z}_t^\top (\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})|$, so that

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} \max_{1 \le t \le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| \asymp \frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} \|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$$

when (for instance) the covariates $\boldsymbol{Z}_t$ have compact support. If moreover $\widehat{\boldsymbol{\beta}}_n$ is $\sqrt{n}-$consistent, as is for instance the case with the standard ordinary least squares estimator, then

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} \max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| \asymp \frac{\sqrt{\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}.$$

Besides, it follows from convergence $a(s)/(e^\star - U(s)) \to -\gamma$ as $s \to \infty$ and the fact that the function $s \mapsto e^\star - U(s)$ is regularly varying with index γ (see de Haan and Ferreira, 2006, Corollary 1.2.10.2 p.23) that one has

$$\frac{\sqrt{\overline{F}(x)}}{a(1/\overline{F}(x))} \asymp \frac{\sqrt{\overline{F}(x)}}{e^\star - U(1/\overline{F}(x))} = G(1/\overline{F}(x)) \quad \text{as} \quad x \uparrow e^\star,$$

where $G(s) = 1/(\sqrt{s}(e^\star - U(s)))$ is regularly varying with index $-(\gamma + 1/2)$. This means that convergence (2.12), and hence Theorem 4, will be satisfied for any choice of intermediate sequence (τ_n) when $\gamma > -1/2$. This discussion naturally leads to the following corollary.

Corollary 5. *Assume that the centered, nondegenerate and bounded random variable ε satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$, with $\gamma > -1/2$, and let $\xi_\tau = \xi_\tau(\varepsilon)$. Let $\tau_n \uparrow 1$ be such that $n\overline{F}(\xi_{\tau_n}) \to \infty$. If the linear model $X_t = \boldsymbol{Z}_t^\top \boldsymbol{\beta} + \varepsilon_t$ holds, where the covariates $\boldsymbol{Z}_t$ have compact support and $\boldsymbol{\beta} \in \mathbb{R}^d$ is estimated by a $\sqrt{n}-$consistent estimator $\widehat{\boldsymbol{\beta}}_n$, then the LAWS estimator $\widehat{\xi}_{\tau_n} = \arg\min_{\theta\in\mathbb{R}} \sum_{t=1}^n \eta_{\tau_n}(\widehat{\varepsilon}_t^{(n)} - \theta)$ based on the residuals $\widehat{\varepsilon}_t^{(n)} = X_t - \boldsymbol{Z}_t^\top \widehat{\boldsymbol{\beta}}_n$ is such that*

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \xrightarrow{d} \mathcal{N}\left(0, \frac{2}{(1-\gamma)(1-2\gamma)}\right),$$

and the tail conditional expectile estimator $\widehat{\xi}_{\tau_n}(X_t|\boldsymbol{Z}_t = \boldsymbol{z}) = \boldsymbol{z}^\top \widehat{\boldsymbol{\beta}}_n + \widehat{\xi}_{\tau_n}$ satisfies

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n}(X_t|\boldsymbol{Z}_t = \boldsymbol{z}) - \xi_{\tau_n}(X_t|\boldsymbol{Z}_t = \boldsymbol{z})) \xrightarrow{d} \mathcal{N}\left(0, \frac{2}{(1-\gamma)(1-2\gamma)}\right).$$

The condition $\gamma > -1/2$ is fairly natural in extreme value theory; it appears in particular in the asymptotic analysis of the semiparametric GPML estimators of the scale and shape extreme value parameters (see de Haan and Ferreira, 2006, Theorem 3.4.2 p.92). It is fortunately satisfied in most models and applications: for example, it is a straightforward consequence of Taylor's theorem with remainder in Lagrange form that this condition holds as soon as ε has a probability density function which is k times continuously differentiable in a neighborhood $[e^\star - \iota, e^\star + \iota]$ of the right endpoint $e^\star$ of ε (for a certain $k \ge 1$ and $\iota > 0$) and whose $(k+1)$th derivative exists in $[e^\star - \iota, e^\star)$ and has a finite positive left limit at $e^\star$. It is noteworthy that while the estimation of central regression parameters often requires assumptions about the smoothness of the probability density function of the errors near 0 (see for example Condition 3.2 in Chaudhuri, 1991, for regression quantile estimation), our framework of extremal regression naturally involves smoothness assumptions near the upper boundary of the support of ε instead.

In applied settings, one would of course require consistent and asymptotically normal estimators of properly extreme conditional expectiles, of the form $\xi_{1-p_n}(X_t|\boldsymbol{Z}_t = \boldsymbol{z})$, where $np_n \to c < \infty$. This in turn requires constructing extrapolated, residual-based estimators of $\xi_{1-p_n}(\varepsilon)$, which can only be obtained by constructing first residual-based estimators of $a(1/\overline{F}(\xi_{\tau_n}))$ and γ, as is already the case in the construction of extrapolated estimators of extreme unconditional expectiles in Section 2.2. Proving rigorously that these residual-based estimators of $a(1/\overline{F}(\xi_{\tau_n}))$ and γ are indeed consistent and asymptotically normal is a difficult mathematical task whose solution may require establishing asymptotic Gaussian theory for the tail empirical process of residuals, that is, $s \mapsto \widehat{\varepsilon}^{(n)}_{n-n(1-\tau_n)s}$, $0 \le s \le 1$. Current results on this empirical process are limited to the setting when the innovations are heavy-tailed, see Girard et al. (2021). The hard but interesting mathematical question of working out the asymptotic behavior of this stochastic process when the ε_t are short-tailed is left for future research.

3 Simulation study

The finite-sample performance of the estimators proposed in Sections 2.2 and 2.3 is illustrated here through a simulation study. The simulation setup first considers three models for i.i.d. observations:

(i) The X_t have a Beta distribution, whose density function is

$$f(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)},\ 0 \le x \le 1.$$

Here $\mathrm{B}(\alpha,\beta)$ is the Beta function and the shape parameters are set as $\alpha = 3$ and $\beta = 5/2$. The extreme value index and the upper endpoint of this model are $\gamma = -2/5$ and $x^\star = 1$, respectively.

(ii) The X_t have a short-tailed power-law distribution, whose distribution function is

$$F(x|x^\star, K, \alpha) = 1 - K(x^\star - x)^\alpha,\ x^\star - K^{-1/\alpha} \le x \le x^\star.$$

Here $x^\star$, K and α are the endpoint of the distribution, a positive constant and the shape parameter, respectively, which have been set as $x^\star = 5$, $K = 1/3$ and $\alpha = 3$, so that the extreme value index is $\gamma = -1/3$.

(iii) The X_t have a GEV distribution, whose distribution function is

$$F(x|\gamma) = \exp(-(1+\gamma x)^{-1/\gamma}),\ 1 + \gamma x > 0.$$

We set the extreme value index $\gamma = -1/3$, so that the upper endpoint is $x^\star = -1/\gamma = 3$.

We then consider the following three time series models, in which Φ denotes the standard normal distribution function and Y_t is the AR(1) process defined as $Y_{t+1} = \varrho Y_t + \sqrt{1-\varrho^2}\,\varepsilon_t$, with independent standard normal innovations ε_t, and where $\varrho \in (-1, 1)$:

(iv) $X_t = q_X(\Phi(Y_t))$, where q_X is the quantile function corresponding to the Beta distribution defined in (i), and where the correlation parameter is $\varrho = 0.95$.

(v) $X_t = q_X(\Phi(Y_t))$, where q_X is the quantile function corresponding to the short-tailed power-law distribution defined in (ii), and where the correlation parameter is $\varrho = 0.5$.

(vi) $X_t = q_X(\Phi(Y_t))$, where q_X is the quantile function corresponding to the GEV distribution defined in (iii), and where the correlation parameter is $\varrho = 0.8$.

The EVI and upper endpoints of models (iv), (v) and (vi) are those of models (i), (ii) and (iii), respectively, and the time series models (iv)-(vi) are geometrically β-mixing (and in particular geometrically $\alpha-$mixing) since the linear AR(1) process (Y_t) is so. We consider the sample sizes $n = 150, 300, 500$ and we aim to predict expectiles of extreme level $\tau_n' = 1 - p_n = 1 - 1/n = 0.9933, 0.9967, 0.9980$. The true expectile values cannot be given in closed form, but they have here been computed by intensive Monte Carlo simulations and are reported in Table 1.

Model	$\tau_n' = 0.9933$	$\tau_n' = 0.9967$	$\tau_n' = 0.9980$
(i), (iv)	0.8571	0.8814	0.8968
(ii), (v)	4.5284	4.5939	4.6372
(iii), (vi)	1.9523	2.1020	2.2000

Table 1: Values of the expectile $\xi_{\tau_n'}$ obtained through intensive Monte Carlo simulations for $\tau_n' = 1 - 1/n$, with $n = 150, 300, 500$.

We simulate $M = 10{,}000$ samples of n observations from each model and compare the purely empirical (LAWS) estimator $\widehat{\xi}_{\tau_n'}$ in (2.6), the extrapolating LAWS estimators $\widehat{\xi}^{\star}_{\tau_n'}$ in (2.8) obtained by setting $\widehat{\sigma}_n = \widehat{a}(1/\widehat{\overline{F}}_n(\widehat{\xi}_{\tau_n}))$, its alternative version $\overline{\xi}^{\star}_{\tau_n'}$ obtained with $\widehat{\sigma}_n = \widehat{a}((1-\tau_n)^{-1}) \times ((1-\tau_n)/\widehat{\overline{F}}_n(\widehat{\xi}_{\tau_n}))^{\widehat{\gamma}_n}$ in view of the approximation $a(1/\overline{F}(\xi_{\tau_n})) \approx ((1-\tau_n)/\overline{F}(\xi_{\tau_n}))^{\gamma} a((1-\tau_n)^{-1})$ that follows from the regular variation property of the scale function a, and the extrapolating quantile-based (QB) estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ in (2.11). In these last three estimators, $(\widehat{a}(n/k), \widehat{\gamma}_n)$ are either the pair of GPML estimators of $(a(n/k), \gamma)$ based on the top k observations in the sample, or their versions based on the Moment estimator. We set throughout $\tau_n = 1 - k/n$, let the effective sample size k range from 1% up to 25% of the total sample size n, and record Monte Carlo approximations of the relative bias, variance and Mean Squared Error (MSE) of the estimators as a function of k.

Results are reported in Figures B.1–B.6 in the Appendix. In each figure the relative bias, variance and MSE are displayed from left to right, and results related to sample sizes $n = 150, 300, 500$ are shown from top to bottom. For the sake of brevity we only report below in Figure 1 the results obtained with the Beta distribution, for the sample size $n = 300$ that we will also consider in our real data analysis of weekly loss returns, but we discuss the conclusions from the full set of models in Appendix B. The Beta model corresponds to a case in which the extreme value bias is present (unlike in the power-law setting, which is a transformation of a pure Pareto model) but not too disruptive in small samples (unlike in the case of the GEV distribution, which should be seen as difficult from that perspective). On the basis of the bias, the empirical estimator and extrapolating QB estimator tend to underestimate the true expectile along the entire range of the effective sample size, while the extrapolating LAWS estimator tends to overestimate the true expectile (at least when the scale and shape parameters are estimated via GPML). From the variance point of view, the extrapolating QB estimator is overall best among the estimators we consider, with the extrapolating LAWS estimators having large variance for small values of the effective sample size. Variability of the estimates seems to be highest when the data points come from time series. This conclusion carries over to the MSE: based on this criterion, the extrapolating QB estimator overall performs best, as expected from our discussion below Corollary 3, with the extrapolating LAWS estimator sometimes outperforming the extrapolating QB estimator for effective sample fractions larger than 20%. In general, both extrapolating QB and LAWS estimators seem to perform remarkably well relative to the purely empirical expectile estimator, especially when one takes into account the small sample size in this simulation study.

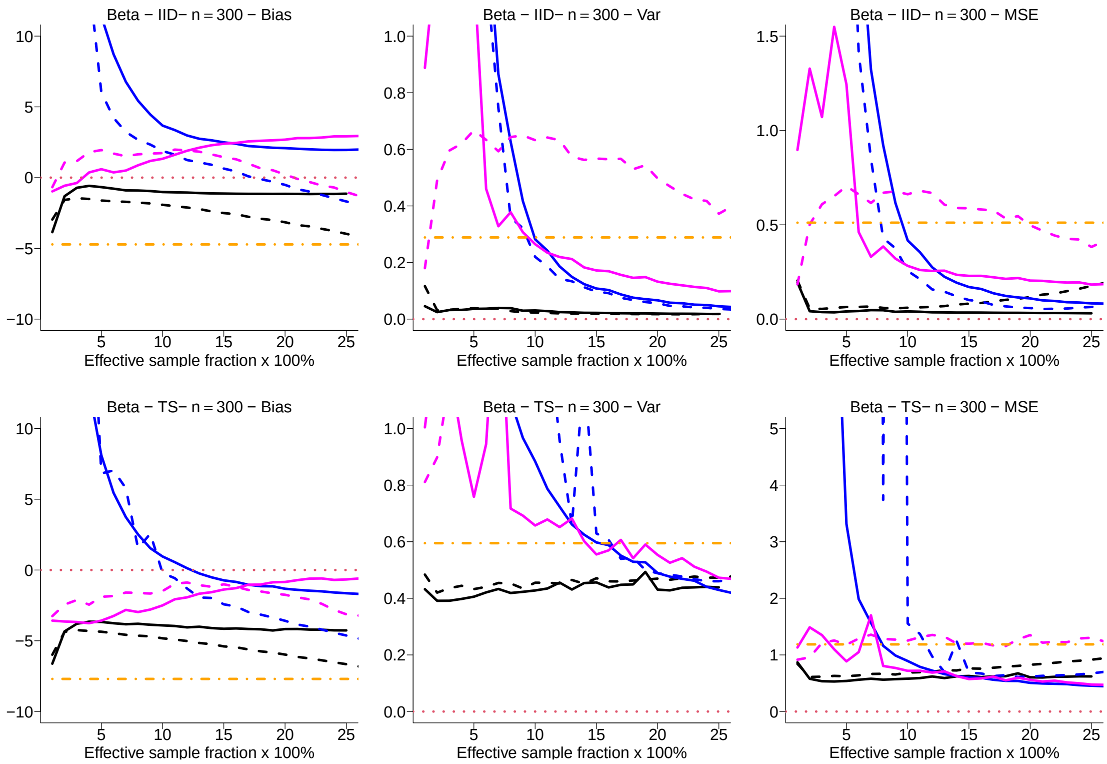

Figure 1: Empirical relative bias, variance and MSE (left, middle and right), multiplied by 100, for the estimators of $\xi_{\tau_n'}$ obtained with observations from a Beta distribution, $\tau_n' = 1 - 1/n$ and sample size $n = 300$. Empirical estimator $\widehat{\xi}_{\tau_n'}$ (orange line), extrapolating LAWS estimators $\widehat{\xi}^{\star}_{\tau_n'}$ (magenta lines) and $\overline{\xi}^{\star}_{\tau_n'}$ (blue lines), and extrapolating QB estimators $\widetilde{\xi}^{\star}_{\tau_n'}$ (black lines). The versions of the extrapolating estimators based on the GPML scale and shape parameter estimates are referred to using solid lines, and those based on the Moment estimators are referred to using dashed lines. Top: i.i.d. data, bottom: nonlinear AR(1) data.

4 Application to forecast verification and comparison

In this section, we apply our LAWS and QB estimation methods to estimate tail risk for Bitcoin (BTC-USD), a peer-to-peer digital decentralized cryptocurrency, and then for the Netflix stock. At the end of September 2014, Bitcoin had volatility seven times greater than gold, eight times greater than the S&P 500, and 18 times greater than the US dollar. Although the growth of Bitcoin prices has been often described as an economic bubble, the COVID-19 crisis has sparked substantial investment in this digital currency as an alternative to conventional asset classes. We construct a time series of weekly loss returns (*i.e.* negative log-returns) from averaged daily Bitcoin closing prices within the corresponding week, from September 28th, 2014, to June 12th, 2022. The time series of loss returns is represented in Figure 2 (A).

We consider risk assessment from a forecasting perspective. With our knowledge of this week, the goal is to give the best possible point estimate of the expectile risk measure $\xi_{\tau_n'}$ for the next week based on rolling windows of length $n = 300$. This window length results in 103 samples of size n over the observed timeframe. For each sample $(X_1, \ldots, X_n)$, the EVI of the underlying distribution was estimated by means of the ML method for peaks over a high threshold $X_{n-k,n}$. The plot of the estimates obtained over the successive 103 rolling windows is given in Figure 2 (B), where an appropriate k is chosen, for each sample, by regarding the path of the ML estimator of γ as a function of k and selecting the k value which corresponds to the median estimate over the most stable region of the path (this can be achieved by employing for instance the algorithm developed by El Methni and Stupfler (2017)). This selection is highlighted in Figure 2 (B) by a colour scheme, ranging from dark red (low) to dark violet (high). The final EVI estimates are found to be all negative in $[-0.147, -0.057]$. It should also be noted that we have comfortably concluded the stationarity of the time series samples across all $T = 103$ rolling windows of short-tailed data, from the Augmented Dickey-Fuller test in our exploratory analysis, at the three significance levels 0.10, 0.05 and 0.01. Similarly, the Kwiatkowski-Phillips-Schmidt-Shin test corroborates the stationarity hypothesis, see Figure 2 (C). A test specifically focused on the tail would of course be important, but the Quintos et al. (2001) test, which is, to the best of our knowledge, the only formally established test in order to detect structural breakpoints in extreme value analysis, does not apply here since it relies on the marginal distributions being heavy-tailed. We tested for heavy tails by implementing the test of Theorem 5.2.12 pp.172-173 in de Haan and Ferreira (2006) under the stringent condition of independent weekly loss returns: the plot of p-values displayed in Figure 2 (D) clearly rejects the assumption of heavy tails over each rolling window at the significance level 0.01. In any event, we only require a local form of stationarity to be valid, with model parameters being allowed to change as the rolling window changes; this is standard practice in the extreme value analysis of financial data, see *e.g.* McNeil and Frey (2000) and Drees (2003). As is to be expected from the resulting range of mildly negative EVI estimates, the 95% confidence intervals for γ in Figure 2 (B), derived from the asymptotic GPML theory under the independence condition (black curves), do not exclude the value 0. Likewise, the one-sided Wald test of $\gamma = 0$ versus $\gamma < 0$, induced by the asymptotic normality of the GPML estimator of γ under the independence condition (Theorem 3.4.2 p.92 in de Haan and Ferreira, 2006), does not reject the null hypothesis as indicated by the plot of p-values in Figure 2 (E). If the hypothesis $\gamma = 0$ were true, a natural implication would be to assume that the light-tailed Bitcoin loss returns have a normal distribution. However, both the Kolmogorov-Smirnov and Shapiro-Wilk tests reject the normality of the weekly data over all estimation windows at the significance level 0.01, as shown in Figure 2 (F). Therefore, the model assumption of a short tailed-distribution over each time period appears to be the only plausible choice.

Expectiles have recently received growing attention in quantitative risk management not only for their coherence as a tail risk measure, but also for their property of elicitability that corresponds to the existence of a natural methodology for backtesting and forecast verification. According to Gneiting (2011) and Ziegel (2016) among others, letting the random variable X model the future observation of interest, the expectile $\xi_{\tau_n'}$ equals the optimal point forecast for X given by the Bayes rule

$$\xi_{\tau_n'} = \underset{\xi \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left[L_{\tau_n'}(\xi, X)\right],$$

under the asymmetric quadratic scoring function

$$L_{\tau_n'} : \mathbb{R}^2 \longrightarrow [0, \infty), \quad (\xi, x) \mapsto \eta_{\tau_n'}(x - \xi),$$

where $L_{\tau_n'}(\xi, x)$ represents the loss or penalty when the point forecast ξ is issued and the realization x of X materializes. Following the ideas of Gneiting (2011) and Ziegel (2016), the competing estimation procedures for $\xi_{\tau_n'}$ can be compared by using the scoring function $L_{\tau_n'}$: Suppose that, in T forecast cases, we have point forecasts $\left(\xi_1^{(m)}, \ldots, \xi_T^{(m)}\right)$ and realizing observations $(x_1, \ldots, x_T)$, where the index m numbers the competing forecasters that are computed at each forecast case $t = 1, \ldots, T$. In the assessment, we compare the purely empirical expectile $\xi_t^{(1)} := \widehat{\xi}_{\tau_n'}$ in (2.6) with the direct extrapolating LAWS estimator $\xi_t^{(2)} := \widehat{\xi}^{\star}_{\tau_n'}$ in (2.8) and its alternative version $\xi_t^{(3)} := \overline{\xi}^{\star}_{\tau_n'}$ described in Section 3, and with the indirect QB extrapolating estimator $\xi_t^{(4)} := \widetilde{\xi}^{\star}_{\tau_n'}$ in (2.11), all of them being based on the GPML estimators $(\widehat{a}(n/k), \widehat{\gamma}_n)$ of $(a(n/k), \gamma)$. When the Moment estimators $(\widehat{a}(n/k), \widehat{\gamma}_n)$ are used instead of the ML estimators, the corresponding three extrapolating forecasters $\widehat{\xi}^{\star}_{\tau_n'}$, $\overline{\xi}^{\star}_{\tau_n'}$ and $\widetilde{\xi}^{\star}_{\tau_n'}$ will be denoted in the sequel by replacing "$\star$" with "$\clubsuit$" to define

$$\xi_t^{(5)} := \widehat{\xi}^{\clubsuit}_{\tau_n'}, \quad \xi_t^{(6)} := \overline{\xi}^{\clubsuit}_{\tau_n'}, \quad \xi_t^{(7)} := \widetilde{\xi}^{\clubsuit}_{\tau_n'}.$$

The seven competing point estimates can then be ranked in terms of their average scores (the lower the better):

$$\overline{L}^{(m)}_{\tau_n'} = \frac{1}{T} \sum_{t=1}^{T} L_{\tau_n'}\left(\xi_t^{(m)}, x_t\right), \quad m = 1, \ldots, 7.$$

The computation of the different extrapolated expectile estimators requires, like the EVI estimators, the determination of the optimal value of the effective sample size k. By balancing the potential estimation bias and variance, a usual practice in extreme value theory is to choose k from the first stable region of the plots [see, *e.g.*, Section 3 in de Haan and Ferreira (2006)]. This is achieved by using the path stability procedure for γ estimation. However, to achieve optimal point forecasts $\left(\xi_1^{(m)}, \ldots, \xi_T^{(m)}\right)$ for the future observation X, this requires the use of k values that minimize their associated realized loss $\overline{L}^{(m)}_{\tau_n'} \equiv \overline{L}^{(m)}_{\tau_n'}(k)$, for $m = 2, \ldots, 7$. Doing so, we obtain the final values of $\overline{L}^{(m)}_{\tau_n'}$ graphed in Figure 2 (G), as functions of the extreme level $\tau_n' \in [0.99, 1]$, for the seven competing estimators. It can be seen that the LAWS-Moment estimator $\widehat{\xi}^{\clubsuit}_{\tau_n'}$ (dashed magenta) is the best forecaster uniformly in τ_n', followed by the LAWS-ML estimator $\widehat{\xi}^{\star}_{\tau_n'}$ (solid magenta) and then the QB-ML estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ (solid black). The remaining three extrapolating estimators do not seem, for this particular choice of $T = 103$ rolling windows of length $n = 300$, to outperform the naive sample expectile $\widehat{\xi}_{\tau_n'}$ (dashed orange). The values of the top-ranked forecaster $\widehat{\xi}^{\clubsuit}_{\tau_n'}$, computed on the 103 successive rolling windows for the extreme levels $\tau_n' \in \{0.99, 0.9933, 0.9966\}$, are displayed in Figure 2 (H), along with the realizing observation at each forecast case. The point forecasts seem to smoothly increase with τ_n' approaching the worst expected (finite) losses at $\tau_n' = 1$. From the perspective of pessimistic decision making, the forecasts obtained at the lower level $\tau_n' = 0.99$ (orange curve) are already cautious since they do lie almost overall beyond the range of the data: This is mainly due to the short-tailed nature of Bitcoin data that is closer to light-tailedness.

Extreme expectiles can also serve as a useful tool for estimating the conventional Value at Risk (VaR) itself. Stated differently, if the statistician or the practitioner wishes to forecast a coherent expectile $\xi_{\tau_n'}$ that has the same probabilistic interpretation as an extreme quantile q_{π_n}, for a pre-specified tail probability level π_n, a natural way of doing so is to select the asymmetry level τ_n' so that $\xi_{\tau_n'} \equiv q_{\pi_n}$. As justified in Section 2.4, such a τ_n'

can be estimated by $\widehat{\tau}'_n = 1 - \frac{(\widehat{x}^\star - \widehat{q}^\star_{\pi_n})(1-\pi_n)}{(\widehat{x}^\star - \overline{X}_n)(1-\widehat{\gamma}_n^{-1})}$. When substituting this estimated value in place of τ'_n in our $\xi_{\tau'_n}$ extrapolated estimators, the latter estimate the VaR q_{π_n} itself and can then be compared with the popular GP fit $\widehat{q}^\star_{\pi_n}$ defined in (2.10). Here also, forecast verification and comparison is possible thanks to the elicitability property of quantiles (see *e.g.* Gneiting (2011)). Given that it is the quantile level π_n which is fixed in advance, the accuracy of the associated VaR forecasts is to be assessed by means of the realized loss

$$\overline{L}^{(m)}_{\pi_n} = \frac{1}{T}\sum_{t=1}^{T} L_{\pi_n}\left(q_t^{(m)}, x_t\right), \quad m = 1, \ldots, 8,$$

under the asymmetric piecewise linear scoring function

$$L_{\pi_n} : \mathbb{R}^2 \longrightarrow [0, \infty), \quad (q, x) \mapsto \varrho_{\pi_n}(x - q),$$

for the competing ML-based forecasters $q_t^{(1)} := \widehat{q}^\star_{\pi_n}$, $q_t^{(2)} := \widehat{\xi}^\star_{\widehat{\tau}'_n}$, $q_t^{(3)} := \overline{\xi}^\star_{\widehat{\tau}'_n}$, $q_t^{(4)} := \widetilde{\xi}^\star_{\widehat{\tau}'_n}$, and their Moment-based versions $q_t^{(5)} := \widehat{q}^\clubsuit_{\pi_n}$, $q_t^{(6)} := \widehat{\xi}^\clubsuit_{\widehat{\tau}'_n}$, $q_t^{(7)} := \overline{\xi}^\clubsuit_{\widehat{\tau}'_n}$, and $q_t^{(8)} := \widetilde{\xi}^\clubsuit_{\widehat{\tau}'_n}$. The resulting realized losses $\overline{L}^{(m)}_{\pi_n}$ are graphed in Figure 2 (I), as functions of the quantile level $\pi_n \in [0.99, 1]$, for the eight competing estimators of $q_{\pi_n} \equiv \xi_{\tau'_n}$. It is remarkable that the best forecaster is still the LAWS-Moment estimator $\widehat{\xi}^\clubsuit_{\widehat{\tau}'_n}$ (dashed magenta), followed by the LAWS-ML estimator $\widehat{\xi}^\star_{\widehat{\tau}'_n}$ (solid magenta). Most importantly, these expectile-based forecasters clearly outperform the usual GP-ML fit $\widehat{q}^\star_{\pi_n}$ (solid orange) and GP-Moment fit $\widehat{q}^\clubsuit_{\pi_n}$ (dashed orange), which is good news to practitioners whose concern is to assess the accuracy of forecasts. Figure 2 (J) contrasts the evolution of the optimal point forecasts $\widehat{\xi}^\clubsuit_{\widehat{\tau}'_n}$, for the risk measure q_{π_n} at the extreme levels $\pi_n \in \{0.99, 0.9933, 0.9966\}$, with the realizing observation at each forecast time. By comparing these π_nth quantile estimates with their expectile analogs from Figure 2 (H) at the same asymmetry levels ($\pi_n = \tau'_n$), it may be seen that expectiles are ultimately less conservative than quantiles, which empirically corroborates the theoretical result for short-tailed data in Proposition 2.2 by Bellini and Di Bernardino (2017). This more liberal expectile assessment of tail risk is indeed a consequence of the diversification principle satisfied by expectiles. Interestingly, the conservative LAWS-Moment (expectile-based) forecasts $\widehat{\xi}^\clubsuit_{\widehat{\tau}'_n}$, for q_{π_n} in Figure 2 (J), seem also to be more sensitive to the variability of weekly losses compared with their analog forecasts $\widehat{\xi}^\clubsuit_{\pi_n}$ for ξ_{π_n} in Figure 2 (H).

We repeated the same exercise based on rolling windows of length $n = 150$ and found negative EVI estimates in $[-0.324, -0.014]$ over 160 successive rolling windows. However, the dominant forecasters become the alternative LAWS-Moment estimator $\overline{\xi}^\clubsuit_{\tau'_n}$ and the LAWS-ML estimator $\widehat{\xi}^\star_{\tau'_n}$ for the expectile risk measure $\xi_{\tau'_n}$, and their composite versions $\overline{\xi}^\clubsuit_{\widehat{\tau}'_n}$ and $\widehat{\xi}^\star_{\widehat{\tau}'_n}$ for the VaR q_{π_n}.

Now, we consider the time series of weekly loss returns of the Netflix stock observed from September 26th, 2014, to April 22nd, 2022 (Figure 3 (A)), and estimate its unconditional EVI over successive rolling windows of length $n = 150$, before forecasting the tail risk based on the pre-identified estimation windows of short-tailed data. First, we conclude the stationarity of the time series samples across all the resulting 246 estimation windows from the Augmented Dickey-Fuller test (Figure 3 (B)). The ML estimates of the EVI obtained over the successive windows are superimposed (as rainbow curve) in Figure 3 (C) with their associated asymptotic 95% confidence intervals derived from the asymptotic theory of the ML estimator under either an independent data assumption, see Theorem 3.4.2 p.92 in de Haan and Ferreira (2006) (black curves), or the condition of

β-mixing data, see Corollary 3.2 p.1283 (see also p.1288) of Drees (2000) (gray curves). It should be noted that, while inference via GPML theory is practically feasible under the assumption of independent data (using low frequency data is a practical solution to reduce the potential serial dependence substantially), it is so far only theoretically possible in a $\beta-$mixing model. The major difficulty in exploiting the asymptotic result in Corollary 3.2 of Drees (2000) lies in the unknown asymptotic variance of the ML γ estimator, which crucially depends on the cumulative serial extremal dependence coefficient $\sum_{t\geq 1} R_t(1,1)$, whose estimation is notoriously difficult. Instead, we used the intuitive average squared estimator $(\log \frac{k}{j_n})^{-1}\sum_{i=j_n}^{k}(\widehat{\gamma}_n^{(i)} - \widehat{\gamma}_n^{(k)})^2$ of the asymptotic variance, based on the ML estimator $\widehat{\gamma}_n^{(i)}$ that uses the $i+1$ largest order statistics of the sample, with a choice of the tuning parameter $j_n = \mathrm{o}(k)$ being eyeballed between the thresholds 5 and 65 over each estimation window. The consistency of this asymptotic variance estimator can be shown by adapting the arguments of the proof of Theorem 2.3 p.630 of Drees (2003) using the expression of the asymptotic variance term in convergence (50) p.652 therein combined with the asymptotic Gaussian representation of $\widehat{\gamma}_n^{(i)}$ given on p.654. As expected from the asymptotic theory in p.1288 of Drees (2000) where the asymptotic variance of the ML estimator is higher under β-mixing serial dependence, the asymptotic Gaussian 95% confidence intervals are wider in this case than in the i.i.d. case. Although they are fairly wide, both confidence intervals comfortably indicate negative EVI values over the first successive 234 estimation windows. This is, however, no longer valid starting from the 235th rolling window due to the appearance of a severe loss return (corresponding to the week of 2022-01-28, as indicated by a vertical red line in Figure 3 (A)) into this estimation window. These results are corroborated by the p-values related to the one-sided Wald test of $\gamma = 0$ versus $\gamma < 0$ obtained in Figure 3 (D), induced by the asymptotic normality of the GPML estimator of γ under the independence condition from Theorem 3.4.2 of de Haan and Ferreira (2006) and under the $\beta-$mixing condition from Corollary 3.2 of Drees (2000).

For the comparison and validation of our competing estimation procedures on historical short-tailed data, we restrict the forecast assessment to the $T = 234$ rolling windows that result in negative EVI estimates in $[-0.6, -0.25]$. The realizations of the future observation to be forecast in an optimal way under both the expectile and the quantile scoring functions on each rolling window are delimited by the vertical red lines in Figure 3 (A). When using the expectile $\xi_{\tau_n'}$ as an optimal point forecast for the future observation, we obtain in Figure 3 (E) the realized losses $\overline{L}_{\tau_n'}^{(m)}$ for its seven competing estimators $\widehat{\xi}_{\tau_n'}$, $\widehat{\xi}^{\star}_{\tau_n'}$, $\overline{\xi}^{\star}_{\tau_n'}$, $\widetilde{\xi}^{\star}_{\tau_n'}$, $\widehat{\xi}^{\clubsuit}_{\tau_n'}$, $\overline{\xi}^{\clubsuit}_{\tau_n'}$ and $\widetilde{\xi}^{\clubsuit}_{\tau_n'}$, as functions of the extreme level $\tau_n' \in [0.99, 1]$. The best forecasters are the two LAWS-ML estimators $\widehat{\xi}^{\star}_{\tau_n'}$ (solid magenta) and $\overline{\xi}^{\star}_{\tau_n'}$ (solid blue), followed by the LAWS-Moment estimator $\overline{\xi}^{\clubsuit}_{\tau_n'}$ (dashed blue). The optimal point forecasts $\widehat{\xi}^{\star}_{\tau_n'}$ obtained at the extreme levels $\tau_n' \in \{0.99, 0.9933, 0.9966\}$ are displayed in Figure 3 (F), along with the realizing observation at each forecast case. When using the quantile q_{π_n} as an optimal point forecast for the future observation, we obtain in Figure 3 (G) the realized losses $\overline{L}_{\pi_n}^{(m)}$ for its eight competing estimators $\widehat{q}^{\star}_{\pi_n}$, $\widehat{\xi}^{\star}_{\widehat{\tau}_n'}$, $\overline{\xi}^{\star}_{\widehat{\tau}_n'}$, $\widetilde{\xi}^{\star}_{\widehat{\tau}_n'}$, $\widehat{q}^{\clubsuit}_{\pi_n}$, $\widehat{\xi}^{\clubsuit}_{\widehat{\tau}_n'}$, $\overline{\xi}^{\clubsuit}_{\widehat{\tau}_n'}$ and $\widetilde{\xi}^{\clubsuit}_{\widehat{\tau}_n'}$, as functions of $\pi_n \in [0.99, 1]$. Globally, the top forecaster is still the LAWS-ML estimator with both its variants $\widehat{\xi}^{\star}_{\widehat{\tau}_n'}$ (solid magenta) and $\overline{\xi}^{\star}_{\widehat{\tau}_n'}$ (solid blue) that clearly outperform the traditional GP-ML fit $\widehat{q}^{\star}_{\pi_n}$ (solid orange) and GP-Moment fit $\widehat{q}^{\clubsuit}_{\pi_n}$ (dashed orange). Finally, Figure 3 (H) contrasts the evolution of the resulting optimal point forecasts $\widehat{\xi}^{\star}_{\widehat{\tau}_n'}$ at $\pi_n \in \{0.99, 0.9933, 0.9966\}$ with the realizing observation at each forecast time.

5 Discussion

We concentrated on the estimation of, and inference about, extreme expectiles of short-tailed distributions in a general setting of weakly dependent and strictly stationary time series. Our assumptions require that the strong mixing coefficients of the data-generating process decay algebraically fast and, in particular, that they form a summable series. This does not cover the interesting frameworks of long memory processes or high-frequency data, both of which have been extensively analyzed in heavy-tailed models (see *e.g.* Kulik and Soulier (2020) and Mao and Zhang (2018)), but whose theory remains untouched in our short-tailed setup. In contrast to high-frequency data analysis, we explored and provided tools that may be used for long-run market risk assessment, which explains the focus on expectiles of the underlying stationary distribution in our theory and weekly loss data in our application. While we did discuss the extension of our results about intermediate expectile estimation to the altogether different problem of conditional/dynamic intermediate expectile estimation for short-tailed models, we could not provide asymptotic theory about the estimation of properly extreme conditional expectiles where the expectile level can grow arbitrarily fast to 1. This is a very difficult problem requiring to obtain asymptotic theory about residual-based estimators of extreme value parameters, which is well beyond the scope of the current paper. Theoretical results along these lines are left for future research.

Acknowledgments

The authors acknowledge the Editor Viktor Todorov, an anonymous Associate Editor and two anonymous reviewers for their helpful comments that led to an improved version of this article. This research was supported by the French National Research Agency under the grants ANR-19-CE40-0013 (ExtremReg) and ANR-17-EURE-0010 (EUR CHESS). A. Daouia and G. Stupfler acknowledge financial support from the TSE-HEC ACPR Chair "Regulation and systemic risks" and from an AXA Research Fund Award on "Mitigating risk in the wake of the COVID-19 pandemic". S.A. Padoan is supported by the Bocconi Institute for Data Science and Analytics (BIDSA), Italy and project MUR - Prin 2022 - Prot. 20227YZ9JK. G. Stupfler acknowledges further support from the Centre Henri Lebesgue (ANR-11-LABX-0020-01) and the Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique.

References

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

Bellini, F. and Di Bernardino, E. (2017). Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506.

Bellini, F., Klar, B., Müller, A., and Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54:41–48.

Billingsley, P. (1995). *Probability and Measure (Third Edition)*. John Wiley and Sons.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144.

Chaudhuri, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, 39(2):246–269.

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. (2020). Tail expectile process and risk assessment. *Bernoulli*, 26(1):531–556.

Davison, A. C., Padoan, S. A., and Stupfler, G. (2023). Tail risk inference via expectiles in heavy-tailed time series. *Journal of Business and Economic Statistics*, 41(3):876–889.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag, New York.

Dekkers, A., Einmahl, J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4):1833–1855.

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *The Annals of Applied Probability*, 10(4):1274–1301.

Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, 9(4):617–657.

El Methni, J. and Stupfler, G. (2017). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. *Statistica Sinica*, 27(2):907–930.

Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *Annals of Statistics*, 49(6):3358–3382.

Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2022). On automatic bias reduction for extreme expectile estimation. *Statistics & Computing*, 32(4):64.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, pages 127–145. Springer Berlin Heidelberg.

Jones, M. C. (1994). Expectiles and M-quantiles are quantiles. *Statistics* & *Probability Letters*, 20(2):149–153.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Kokic, P., Chambers, R., Breckling, J., and Beare, S. (1997). A measure of production performance. *Journal of Business & Economic Statistics*, 15(4):445–451.

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.

Kulik, R. and Soulier, P. (2020). *Heavy-Tailed Time Series*. Springer, New York.

Mao, G. and Zhang, Z. (2018). Stochastic tail index model for high frequency financial data with bayesian analysis. *Journal of Econometrics*, 205(2):470–487.

Mao, T., Ng, K., and Hu, T. (2015). Asymptotic expansions of generalized quantiles and expectiles for extreme risks. *Probability in the Engineering and Informational Sciences*, 29(3):309–327.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3–4):271–300.

Nelsen, R. B. (2006). *An Introduction to Copulas (second edition)*. Springer Science+Business Media, New York.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847.

Padoan, S. A. and Stupfler, G. (2022). Joint inference on extreme expectiles for multivariate heavy-tailed distributions. *Bernoulli*, 28(2):1021–1048.

Pratt, J. W. (1960). On interchanging limits and integrals. *The Annals of Mathematical Statistics*, 31(1):74–77.

Quintos, C., Fan, Z. H., and Phillips, P. C. B. (2001). Structural change tests in tail behavior and the Asian crisis. *Review of Economic Studies*, 68(3):633–663.

Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes. In Probability Theory and Stochastic Modelling*, volume 80, Springer-Verlag, Berlin-Heidelberg.

Rootzén, H., Leadbetter, M. R., and de Haan, L. (1998). On the distribution of tail array sums for strongly mixing stationary sequences. *Annals of Applied Probability*, 8(3):868–885.

Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33(2):307–335.

Taylor, J. (2008). Estimating Value at Risk and Expected Shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252.

Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*, 26(4):901–918.

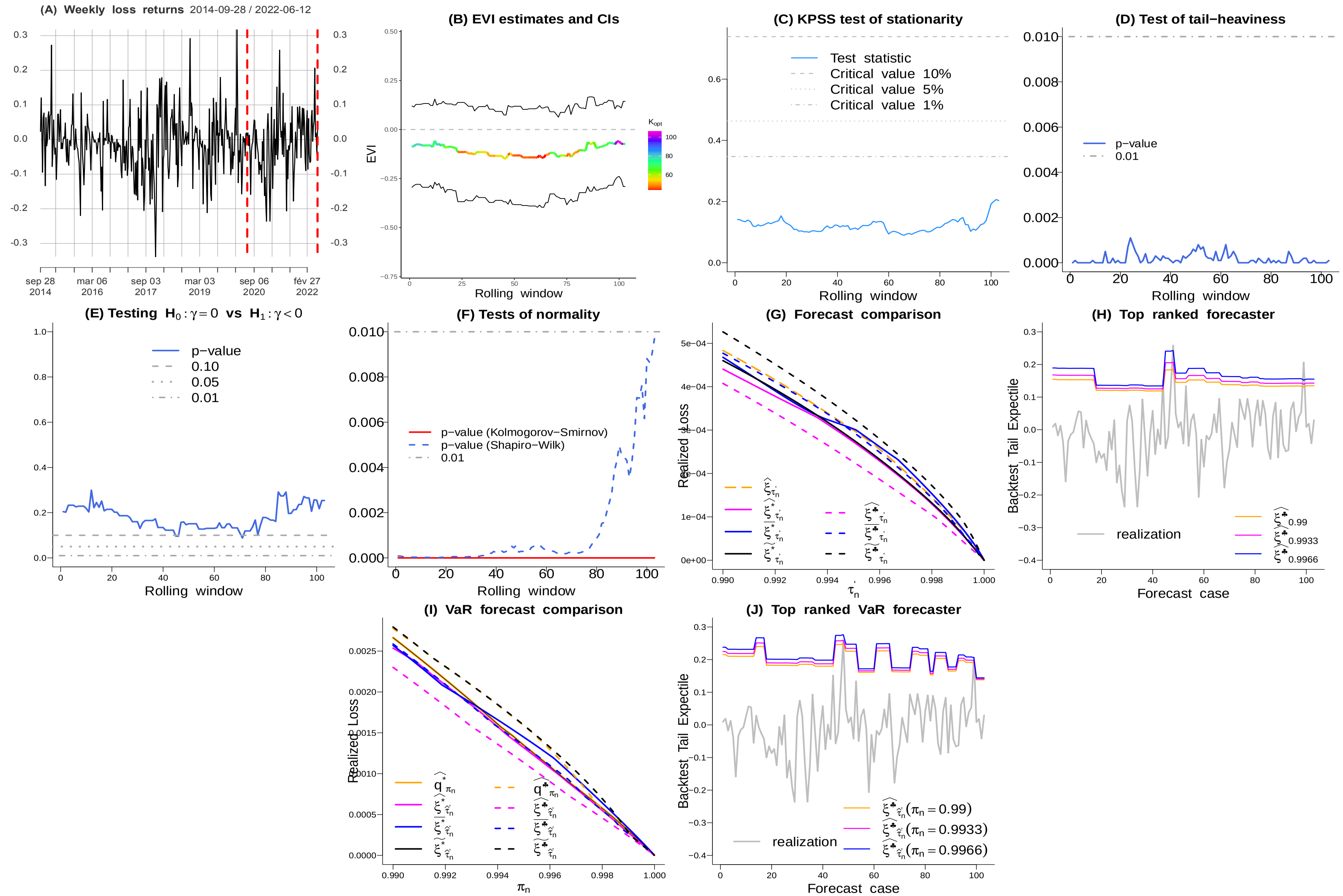

Figure 2: (A) Bitcoin weekly loss returns from September 28th, 2014, to June 12th, 2022. The vertical red lines delimit the realizations of the future observation to be forecast. (B) ML estimates of γ over the 103 rolling windows (rainbow curve), and the associated 95% confidence intervals (black curves). (C) Kwiatkowski-Phillips-Schmidt-Shin test. (D) Test of tail-heaviness. (E) Test of $\gamma = 0$ versus $\gamma < 0$. (F) Kolmogorov-Smirnov and Shapiro-Wilk tests. (G) Realized loss functions $\tau_n' \mapsto \overline{L}^{(m)}_{\tau_n'}$. (H) Top-ranked forecaster of $\xi_{\tau_n'}$, for $\tau_n' = 0.99, 0.9933, 0.9966$. (I) Realized loss functions $\pi_n \mapsto \overline{L}^{(m)}_{\pi_n}$. (J) Top-ranked forecaster of q_{π_n}, for $\pi_n = 0.99, 0.9933, 0.9966$.

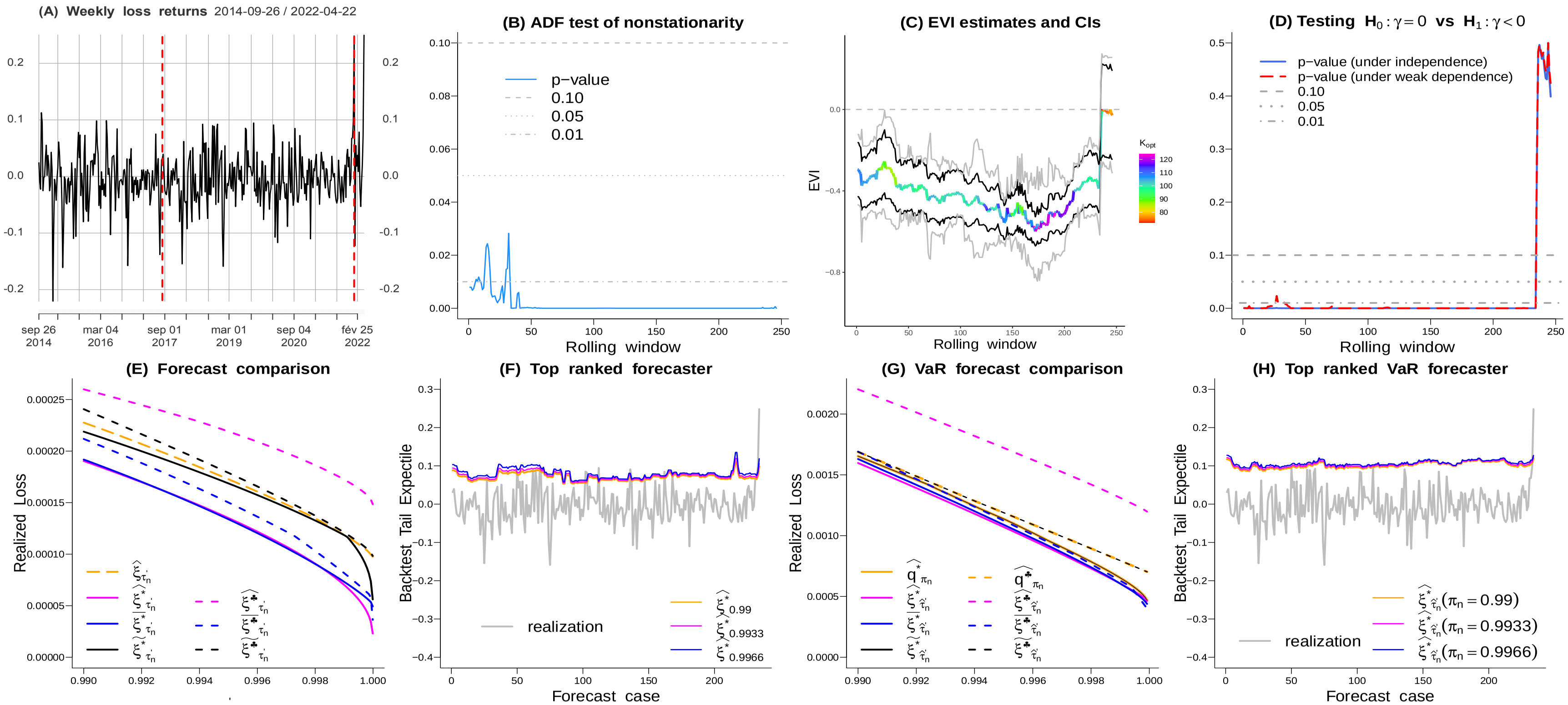

Figure 3: (A) Netflix weekly loss returns from September 26th, 2014, to April 22nd, 2022. The vertical red lines delimit the realizations of the future observation to be forecast. (B) Augmented Dickey-Fuller test. (C) ML estimates of γ over the 246 rolling windows (rainbow curve), along with the associated 95% confidence intervals derived under an independent data assumption (black curves) and a β-mixing data condition (gray curves). (D) Test of $\gamma = 0$ versus $\gamma < 0$ with the p-values under the independence condition (blue curve) and the $\beta-$mixing condition (red curve). (E) Realized loss functions $\tau_n' \mapsto \overline{L}^{(m)}_{\tau_n'}$ for the seven competing $\xi_{\tau_n'}$ forecasters. (F) Top-ranked forecaster of $\xi_{\tau_n'}$, for $\tau_n' = 0.99, 0.9933, 0.9966$, along with the realizing observations. (G) Realized loss functions $\pi_n \mapsto \overline{L}^{(m)}_{\pi_n}$ for the eight competing q_{π_n} forecasters. (H) Top-ranked forecaster of q_{π_n}, for $\pi_n = 0.99, 0.9933, 0.9966$.

Appendix to the paper "Extreme expectile estimation for short-tailed data"

Abdelaati Daouia, Simone A. Padoan and Gilles Stupfler

This appendix contains all necessary proofs and provides extra finite-sample results about our simulation study.

A Proofs of the main results

A.1 Auxiliary results

We first of all list a number of facts that will be used numerous times in our proofs: if condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds, then

- Condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds locally uniformly in z, see Remark B.3.8.1 on relationship (B.3.3) in de Haan and Ferreira (2006).
- The right endpoint of F is finite and will be denoted in the sequel by $x^\star$, see Theorem 1.2.1 on p.19 of de Haan and Ferreira (2006).
- One has $a(s)/(x^\star - U(s)) \to -\gamma$ as $s \to \infty$, see Lemma 1.2.9 on p.22 of de Haan and Ferreira (2006).
- The functions $x \mapsto \overline{F}(x^\star - 1/x)$ and $s \mapsto x^\star - U(s)$ are regularly varying with indices $1/\gamma$ and γ, respectively, see Theorem 1.2.1.2 on p.19 and Corollary 1.2.10.2 on p.23 of de Haan and Ferreira (2006).

Our first auxiliary result is a useful asymptotic inversion lemma that will be used several times.

Lemma A.1. *Suppose that condition $\mathcal{C}_2(\gamma, a, \rho, A)$ holds.*

(i) One has

$$\lim_{x \uparrow x^\star} \frac{U(1/\overline{F}(x)) - x}{a(1/\overline{F}(x))A(1/\overline{F}(x))} = 0.$$

In particular

$$\lim_{x \uparrow x^\star} \frac{1}{A(1/\overline{F}(x))} \left(\frac{x^\star - U(1/\overline{F}(x))}{x^\star - x} - 1 \right) = 0.$$

(ii) One has

$$\lim_{\tau \uparrow 1} \frac{1}{A((1-\tau)^{-1})} \left(\frac{\overline{F}(q_\tau)}{1-\tau} - 1 \right) = 0.$$

Proof of Lemma A.1. We only show (i); the proof of (ii) is similarly written by using an equivalent second-order condition on $\overline{F}$ (see de Haan and Ferreira, 2006, Theorem 2.3.8

p.48). Assume that A is positive; the proof for a negative A is similar. Condition $\mathcal{C}_2(\gamma,a,\rho,A)$ holds locally uniformly in z, so pick $\varepsilon\neq 0$ and apply this condition to obtain

$$\lim_{x\uparrow x^\star}\left|\frac{1}{A(1/\overline{F}(x))}\left(\frac{U([1+\varepsilon A(1/\overline{F}(x))]/\overline{F}(x))-U(1/\overline{F}(x))}{a(1/\overline{F}(x))}-\frac{(1+\varepsilon A(1/\overline{F}(x)))^\gamma-1}{\gamma}\right)\right.$$
$$\left.-\int_1^{1+\varepsilon A(1/\overline{F}(x))} v^{\gamma-1}\left(\int_1^v u^{\rho-1}\mathrm{d}u\right)\mathrm{d}v\right|=0$$

and therefore

$$\lim_{x\uparrow x^\star}\frac{1}{a(1/\overline{F}(x))A(1/\overline{F}(x))}\left|U([1+\varepsilon A(1/\overline{F}(x))]/\overline{F}(x))-U(1/\overline{F}(x))\right.$$
$$\left.-a(1/\overline{F}(x))\frac{(1+\varepsilon A(1/\overline{F}(x)))^\gamma-1}{\gamma}\right|=0.$$

Conclude that

$$\lim_{x\uparrow x^\star}\frac{U([1+\varepsilon A(1/\overline{F}(x))]/\overline{F}(x))-U(1/\overline{F}(x))}{a(1/\overline{F}(x))A(1/\overline{F}(x))}=\varepsilon.$$

By definition of U as the left-continuous inverse of $1/\overline{F}$, one has $U([1+\varepsilon A(1/\overline{F}(x))]/\overline{F}(x))\geq x$ when $\varepsilon>0$ (resp. $\leq x$ when $\varepsilon<0$), so

$$\limsup_{x\uparrow x^\star}\frac{x-U(1/\overline{F}(x))}{a(1/\overline{F}(x))A(1/\overline{F}(x))}\leq\lim_{x\uparrow x^\star}\frac{U([1+\varepsilon A(1/\overline{F}(x))]/\overline{F}(x))-U(1/\overline{F}(x))}{a(1/\overline{F}(x))A(1/\overline{F}(x))}=\varepsilon$$

and a similar lower bound applies. Let $\varepsilon\to 0$ to complete the proof of the first convergence. The second convergence follows because $a(s)/(x^\star-U(s))\to-\gamma$ as $s\to\infty$. □

A fine understanding of the asymptotic behavior of extreme expectiles requires an asymptotic expansion of $\varphi^{(1)}(x)=\mathbb{E}((X-x)\mathbb{1}\{X>x\})$ for x close to the right endpoint $x^\star$. This is the focus of the below lemma, where we recall that more generally $\varphi^{(\kappa)}(x)=\mathbb{E}((X-x)^\kappa\mathbb{1}\{X>x\})$.

Lemma A.2. *Suppose that condition $\mathcal{C}_2(\gamma,a,\rho,A)$ holds.*

(i) Then, for any $\kappa\geq 1$, and as $x\uparrow x^\star$,

$$\frac{\varphi^{(\kappa)}(x)}{\overline{F}(x)[a(1/\overline{F}(x))]^\kappa}=\mathrm{O}(1).$$

(ii) As $x\uparrow x^\star$,

$$\frac{\varphi^{(1)}(x)}{\overline{F}(x)a(1/\overline{F}(x))}=\frac{1}{1-\gamma}\left(1+\frac{1}{1-\gamma-\rho}A(1/\overline{F}(x))+\mathrm{o}(|A(1/\overline{F}(x))|)\right).$$

(iii) As $x\uparrow x^\star$,

$$\frac{\varphi^{(2)}(x)}{\overline{F}(x)[a(1/\overline{F}(x))]^2}$$
$$=\frac{2}{(1-\gamma)(1-2\gamma)}\left(1+\frac{3-4\gamma-2\rho}{(1-\gamma-\rho)(1-2\gamma-\rho)}A(1/\overline{F}(x))+\mathrm{o}(|A(1/\overline{F}(x))|)\right).$$

Proof of Lemma A.2. Statement (i) is shown by writing

$$\frac{\varphi^{(\kappa)}(x)}{\overline{F}(x)[a(1/\overline{F}(x))]^{\kappa}} \le \left(\frac{x^\star - x}{a(1/\overline{F}(x))}\right)^{\kappa} = \left(-\frac{1}{\gamma} \times \frac{x^\star - x}{x^\star - U(1/\overline{F}(x))}\right)^{\kappa} (1 + \mathrm{o}(1)) = \mathrm{O}(1)$$

by Lemma A.1(i) and because $a(s)/(x^\star - U(s)) \to -\gamma$ as $s \to \infty$.

To show (ii) and (iii), recall that $X \stackrel{d}{=} U(Y)$, where Y is a unit Pareto random variable. Write $\varphi^{(1)}(x) = \mathbb{E}((U(Y) - x)\mathbb{1}\{Y > 1/\overline{F}(x)\})$. By Lemma A.1(i),

$$\begin{aligned}\varphi^{(1)}(x) &= \mathbb{E}((U(Y) - U(1/\overline{F}(x)))\mathbb{1}\{Y > 1/\overline{F}(x)\}) + \mathrm{o}(\overline{F}(x)a(1/\overline{F}(x))|A(1/\overline{F}(x))|) \\ &= \int_{1/\overline{F}(x)}^{\infty} (U(y) - U(1/\overline{F}(x)))\frac{\mathrm{d}y}{y^2} + \mathrm{o}(\overline{F}(x)a(1/\overline{F}(x))|A(1/\overline{F}(x))|) \\ &= \frac{1}{s}\int_1^{\infty} (U(sz) - U(s))\frac{\mathrm{d}z}{z^2} + \mathrm{o}(\overline{F}(x)a(1/\overline{F}(x))|A(1/\overline{F}(x))|) \end{aligned} \tag{A.1}$$

with $s = s(x) = 1/\overline{F}(x) \to \infty$ (as $x \uparrow x^\star$). Define now $a_\star$ and $A_\star$ as

$$a_\star(s) = \begin{cases} a(s)\left(1 - \dfrac{1}{\rho}A(s)\right), & \rho < 0, \\ a(s)\left(1 - \dfrac{1}{\gamma}A(s)\right), & \rho = 0, \end{cases} \quad \text{and } A_\star(s) = \begin{cases} \dfrac{1}{\rho}A(s), & \rho < 0, \\ A(s), & \rho = 0. \end{cases} \tag{A.2}$$

By the set of uniform inequalities in Theorem 2.3.6 of de Haan and Ferreira (2006), there exist functions a_0 and A_0 such that A_0 is asymptotically equivalent to $A_\star$ and $a_0(s)/a_\star(s) = 1 + \mathrm{o}(|A_\star(s)|)$ as $s \to \infty$, and, for any $\varepsilon > 0$, the following inequality holds for s large enough:

$$\forall z \ge 1, \ \left|\frac{1}{A_0(s)}\left(\frac{U(sz) - U(s)}{a_0(s)} - \frac{z^\gamma - 1}{\gamma}\right) - \Psi_{\gamma,\rho}(z)\right| \le \varepsilon z^{\gamma+\rho+\varepsilon},$$
$$\text{where } \Psi_{\gamma,\rho}(z) = \begin{cases} \dfrac{z^{\gamma+\rho} - 1}{\gamma + \rho}, & \rho < 0, \\ \dfrac{1}{\gamma} z^\gamma \log(z), & \rho = 0. \end{cases} \tag{A.3}$$

Write then

$$\frac{1}{s}\int_1^{\infty} (U(sz) - U(s))\frac{\mathrm{d}z}{z^2} = \frac{a_0(s)}{s}\int_1^{\infty} \frac{z^\gamma - 1}{\gamma}\frac{\mathrm{d}z}{z^2} + \frac{a_0(s)A_0(s)}{s}\int_1^{\infty} \frac{1}{A_0(s)}\left(\frac{U(sz) - U(s)}{a_0(s)} - \frac{z^\gamma - 1}{\gamma}\right)\frac{\mathrm{d}z}{z^2}$$

and use (A.3) in conjunction with the dominated convergence theorem together with straightforward calculations to get

$$\frac{1}{s}\int_1^{\infty} (U(sz) - U(s))\frac{\mathrm{d}z}{z^2} = \frac{a_0(s)}{s}\left(\frac{1}{1-\gamma} + A_0(s)\left[\frac{1}{1-\gamma-\rho}\mathbb{1}\{\rho < 0\} + \frac{1}{\gamma(1-\gamma)^2}\mathbb{1}\{\rho = 0\} + \mathrm{o}(1)\right]\right).$$

Combine this last identity with (A.1), (A.2) and a straightforward calculation to conclude the proof of (ii). We turn to showing (iii). Start by writing

$$(X - x)^2 \stackrel{d}{=} (U(Y) - U(1/\overline{F}(x)))^2 + (U(1/\overline{F}(x)) - x) \times (2(U(Y) - U(1/\overline{F}(x))) + (U(1/\overline{F}(x)) - x))$$

and use the results of the proof of (ii) along with Lemma A.1(i) to obtain

$$\begin{aligned}
\varphi^{(2)}(x) &= \mathbb{E}((U(Y)-U(1/\overline{F}(x)))^2 \mathbb{1}\{Y>1/\overline{F}(x)\}) + \mathrm{o}(\overline{F}(x)[a(1/\overline{F}(x))]^2|A(1/\overline{F}(x))|) \\
&= \int_{1/\overline{F}(x)}^{\infty} (U(y)-U(1/\overline{F}(x)))^2 \frac{\mathrm{d}y}{y^2} + \mathrm{o}(\overline{F}(x)[a(1/\overline{F}(x))]^2|A(1/\overline{F}(x))|) \\
&= \frac{1}{s}\int_1^{\infty} (U(sz)-U(s))^2 \frac{\mathrm{d}z}{z^2} + \mathrm{o}(\overline{F}(x)[a(1/\overline{F}(x))]^2|A(1/\overline{F}(x))|) \qquad \text{(A.4)}
\end{aligned}$$

where again $s=s(x)=1/\overline{F}(x)\to\infty$ as $x\uparrow x^\star$. Now

$$\begin{aligned}
\frac{1}{s}\int_1^{\infty}(U(sz)-U(s))^2\frac{\mathrm{d}z}{z^2} = \frac{[a_0(s)]^2}{s}&\int_1^{\infty}\left(\frac{z^\gamma-1}{\gamma}\right)^2\frac{\mathrm{d}z}{z^2} \\
+\,2\frac{[a_0(s)]^2}{s}&\int_1^{\infty}\left(\frac{U(sz)-U(s)}{a_0(s)}-\frac{z^\gamma-1}{\gamma}\right)\frac{z^\gamma-1}{\gamma}\frac{\mathrm{d}z}{z^2} \\
+\,\frac{[a_0(s)]^2}{s}&\int_1^{\infty}\left(\frac{U(sz)-U(s)}{a_0(s)}-\frac{z^\gamma-1}{\gamma}\right)^2\frac{\mathrm{d}z}{z^2}.
\end{aligned}$$

Combine (A.3) with the dominated convergence theorem and straightforward calculations to find

$$\begin{aligned}
&\frac{1}{s}\int_1^{\infty}(U(sz)-U(s))^2\frac{\mathrm{d}z}{z^2} = \frac{[a_0(s)]^2}{s}\Bigg(\frac{2}{(1-\gamma)(1-2\gamma)} \\
&+A_0(s)\left[\frac{2(2-2\gamma-\rho)}{(1-\gamma)(1-\gamma-\rho)(1-2\gamma-\rho)}\mathbb{1}\{\rho<0\}+\frac{2(2-3\gamma)}{\gamma(1-\gamma)^2(1-2\gamma)^2}\mathbb{1}\{\rho=0\}+\mathrm{o}(1)\right]\Bigg).
\end{aligned}$$

Conclude the proof by combining (A.2) with (A.4) and further straightforward calculations. □

Inverting the limiting relationship (2.4), and providing an asymptotic expansion that strengthens (2.5), requires in particular an asymptotic expansion of $x^\star-\xi_\tau$. We do so in the following lemma.

Lemma A.3. *Suppose that condition $\mathcal{C}_2(\gamma,a,\rho,A)$ holds with $\rho<0$.*

(i) The limit $C=\lim_{s\to\infty}s^{-\gamma}(x^\star-U(s))$ exists, is positive and finite, with

$$x^\star-U(s)=Cs^\gamma\left(1+\frac{\gamma}{\rho(\gamma+\rho)}A(s)+\mathrm{o}(|A(s)|)\right)$$

$$\text{and}\quad a(s)=-\gamma Cs^\gamma\left(1+\frac{A(s)}{\rho}+\mathrm{o}(|A(s)|)\right)\quad\text{as } s\to\infty.$$

(ii) With the notation of (i), as $x\uparrow x^\star$,

$$\overline{F}(x)=C^{1/\gamma}(x^\star-x)^{-1/\gamma}\left(1+\frac{1}{\rho(\gamma+\rho)}A(1/\overline{F}(x))+\mathrm{o}(|A(1/\overline{F}(x))|)\right)$$

$$\text{and } \overline{F}(x)a(1/\overline{F}(x))=-\gamma C^{1/\gamma}(x^\star-x)^{1-1/\gamma}\left(1+\frac{\rho+1}{\rho(\gamma+\rho)}A(1/\overline{F}(x))+\mathrm{o}(|A(1/\overline{F}(x))|)\right).$$

Proof of Lemma A.3. The main difficulty is to show (i). We start by the assertion on U. By Remark B.3.7 in de Haan and Ferreira (2006) with $(c_1, c_2) = (1, 0)$, the limit $c = \lim_{s\to\infty} s^{-\gamma} a(s) \in (0, \infty)$ exists and the function g defined by

$$g(s) = U(s) - c\frac{s^\gamma - 1}{\gamma}$$

satisfies

$$\lim_{s\to\infty} \frac{g(sz) - g(s)}{a(s)A(s)} = \frac{1}{\rho} \times \frac{z^{\gamma+\rho} - 1}{\gamma + \rho}, \text{ for all } z > 0.$$

By Theorem B.2.2 on p.373 of de Haan and Ferreira (2006) applied to $-\operatorname{sign}(A)g$, $c' = \lim_{s\to\infty} g(s)$ exists and

$$\lim_{s\to\infty} \frac{c' - g(s)}{a(s)A(s)} = -\frac{1}{\rho(\gamma + \rho)}.$$

The identity $c' = \lim_{s\to\infty} g(s)$ yields

$$c' = x^\star + \frac{c}{\gamma} \text{ and thus } c' - g(s) = x^\star - U(s) + c\frac{s^\gamma}{\gamma}.$$

The above convergence and the convergence $a(s)/(x^\star - U(s)) \to -\gamma$, as $s \to \infty$, then provide

$$\begin{aligned} x^\star - U(s) &= -c\frac{s^\gamma}{\gamma} - \frac{1}{\rho(\gamma+\rho)} a(s)A(s)(1 + \mathrm{o}(1)) \\ &= -c\frac{s^\gamma}{\gamma} + \frac{\gamma}{\rho(\gamma+\rho)}(x^\star - U(s))A(s)(1 + \mathrm{o}(1)). \end{aligned}$$

Set finally $C = -c/\gamma$ to find

$$x^\star - U(s) = Cs^\gamma \left(1 + \frac{\gamma}{\rho(\gamma+\rho)} A(s) + \mathrm{o}(|A(s)|)\right), \ s \to \infty,$$

as required. To show the assertion on a, set $h(s) = s^{-\gamma}a(s)$ and rewrite Equation (2.3.7) on p.44 of de Haan and Ferreira (2006) as

$$\lim_{s\to\infty} \frac{h(sz) - h(s)}{s^{-\gamma}a(s)A(s)} = \frac{z^\rho - 1}{\rho}, \text{ for all } z > 0.$$

By Theorem B.2.2 on p.373 of de Haan and Ferreira (2006) again,

$$\lim_{s\to\infty} \frac{c - h(s)}{s^{-\gamma}a(s)A(s)} = -\frac{1}{\rho}.$$

The conclusion in (i) is now immediate since $h(s) = s^{-\gamma}a(s)$ and $c/C = -\gamma$. The expansions in (ii) are obtained by taking $s = 1/\overline{F}(x)$ in the expansion of $x^\star - U(s)$ and then using Lemma A.1(i). □

The following lemma is the essential element in obtaining the joint asymptotic normality of the LAWS estimator and empirical quantile when the data generating process is $\alpha-$mixing. In the proof of this lemma and later on we shall use the following result: under condition $\mathcal{C}_2(\gamma, a, \rho, A)$, if $x_n, u_n \uparrow x^\star$ satisfy $(x^\star - x_n)/(x^\star - u_n) \to 1$, then

$$\frac{\overline{F}(x_n)}{\overline{F}(u_n)} \to 1 \text{ and } \frac{a(1/\overline{F}(x_n))}{a(1/\overline{F}(u_n))} \to 1. \tag{A.5}$$

The first convergence is found by using the regular variation property of $x \mapsto \overline{F}(x^\star - 1/x)$. The second one is then obtained by combining the convergence $a(s)/(x^\star - U(s)) \to -\gamma$, as $s \to \infty$, with the regular variation property of $s \mapsto x^\star - U(s)$.

Lemma A.4. *Assume that X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$. Suppose that F is continuous and that $(X_t)_{t\geq 1}$ is a strictly stationary sequence of copies of X satisfying conditions $\mathcal{M}$ and $\mathcal{D}$. Let finally $u_n \uparrow x^\star$ be such that $n\overline{F}(u_n) \to \infty$, $r_n\overline{F}(u_n) \to 0$, and $x_n, x_n' \uparrow x^\star$ be such that $(x^\star - x_n)/(x^\star - u_n) \to 1$ and $(x^\star - x_n')/(x^\star - u_n) \to 1$.*

(i) If $s_n \to \infty$ is such that $s_n = \mathrm{O}(r_n)$, one has

$$\frac{n}{s_n}\operatorname{Var}\left(\sum_{t=1}^{s_n}\sqrt{\frac{\overline{F}(u_n)}{n}}\left(\frac{\mathbb{1}\{X_t > x_n\}}{\mathbb{P}(X > x_n)} - 1\right)\right) \to 1 + 2\sum_{t=1}^{\infty} R_t(1,1),$$

$$\frac{n}{s_n}\operatorname{Cov}\left(\sum_{t=1}^{s_n}\sqrt{\frac{\overline{F}(u_n)}{n}}\left(\frac{(X_t - x_n)\mathbb{1}\{X_t > x_n\}}{\mathbb{E}((X - x_n)\mathbb{1}\{X > x_n\})} - 1\right), \sum_{t=1}^{s_n}\sqrt{\frac{\overline{F}(u_n)}{n}}\left(\frac{\mathbb{1}\{X_t > x_n'\}}{\mathbb{P}(X > x_n')} - 1\right)\right)$$

$$\to 1 + (1-\gamma^{-1})\int_0^1 \sum_{t=1}^{\infty}[R_t(x^{-1/\gamma}, 1) + R_t(1, x^{-1/\gamma})]\,\mathrm{d}x, \ \textit{and}$$

$$\frac{n}{s_n}\operatorname{Var}\left(\sum_{t=1}^{s_n}\sqrt{\frac{\overline{F}(u_n)}{n}}\left(\frac{(X_t - x_n)\mathbb{1}\{X_t > x_n\}}{\mathbb{E}((X - x_n)\mathbb{1}\{X > x_n\})} - 1\right)\right)$$

$$\to \frac{2(1-\gamma)}{1-2\gamma} + 2(1-\gamma^{-1})^2 \iint_{(0,1]^2}\sum_{t=1}^{\infty} R_t(x^{-1/\gamma}, y^{-1/\gamma})\,\mathrm{d}x\,\mathrm{d}y.$$

(ii) If the assumption $s_n \to \infty$ is dropped, then each of the three sequences in (i) stays bounded.

Proof of Lemma A.4. We prove both statements for the third sequence because the proofs for the first two sequences are simpler, and we start by preliminary calculations. By Lemma A.2(ii),

$$\frac{n}{s_n}\operatorname{Var}\left(\sum_{t=1}^{s_n}\sqrt{\frac{\overline{F}(u_n)}{n}}\left(\frac{(X_t - x_n)\mathbb{1}\{X_t > x_n\}}{\mathbb{E}((X - x_n)\mathbb{1}\{X > x_n\})} - 1\right)\right)$$

$$= (1-\gamma)^2 \times \frac{1+\mathrm{o}(1)}{\overline{F}(u_n)[a(1/\overline{F}(u_n))]^2} \times \frac{1}{s_n}\operatorname{Var}\left(\sum_{t=1}^{s_n}(X_t - x_n)\mathbb{1}\{X_t > x_n\}\right). \tag{A.6}$$

Then, combining Lemma A.2(ii) and (iii) with (A.5) and $s_n\overline{F}(u_n) = \mathrm{O}(r_n\overline{F}(u_n)) \to 0$,

$$\begin{aligned}
&\frac{1}{s_n}\operatorname{Var}\left(\sum_{t=1}^{s_n}(X_t - x_n)\mathbb{1}\{X_t > x_n\}\right) \\
&= \mathbb{E}((X - x_n)^2\mathbb{1}\{X > x_n\}) - s_n\,[\mathbb{E}((X - x_n)\mathbb{1}\{X > x_n\})]^2 \\
&+ \frac{2}{s_n}\sum_{t=1}^{s_n - 1}(s_n - t)\mathbb{E}((X_1 - x_n)(X_{t+1} - x_n)\mathbb{1}\{X_1 > x_n, X_{t+1} > x_n\}) \\
&= \frac{2}{(1-\gamma)(1-2\gamma)}\overline{F}(u_n)[a(1/\overline{F}(u_n))]^2(1+\mathrm{o}(1)) \\
&+ 2\sum_{t=1}^{\infty}\left(1 - \frac{t}{s_n}\right)\iint_{[x_n, x^\star)^2}\mathbb{P}(X_1 > v, X_{t+1} > v')\,\mathrm{d}v\,\mathrm{d}v'\,\mathbb{1}\{t < s_n\}.
\end{aligned} \tag{A.7}$$

It remains to control the integral in (A.7). Taking into account the continuity of F, the change of variables $(v, v') = (x_n - \gamma(x^\star - x_n)w, x_n - \gamma(x^\star - x_n)w') = (x^\star - (1 + \gamma w)(x^\star -$

$x_n), x^\star - (1+\gamma w')(x^\star - x_n))$, convergence $a(s)/(x^\star - U(s)) \to -\gamma$ as $s \to \infty$, Lemma A.1(i) and convergence (A.5) yield

$$\frac{1}{\overline{F}(u_n)[a(1/\overline{F}(u_n))]^2} \iint_{[x_n, x^\star)^2} \mathbb{P}(X_1 > v, X_{t+1} > v')\,\mathrm{d}v\,\mathrm{d}v'$$
$$= \iint_{[0,-1/\gamma)^2} \frac{1}{\overline{F}(u_n)} \mathbb{P}(\overline{F}(X_1) \leq \overline{F}(x^\star - (1+\gamma w)(x^\star - x_n)),$$
$$\overline{F}(X_{t+1}) \leq \overline{F}(x^\star - (1+\gamma w')(x^\star - x_n)))\,\mathrm{d}w\,\mathrm{d}w'(1+\mathrm{o}(1)). \quad (A.8)$$

Since $x \mapsto \overline{F}(x^\star - 1/x)$ is regularly varying with index $1/\gamma$, one has

$$\forall w \in [0, -1/\gamma), \ \lim_{n\to\infty} \frac{\overline{F}(x^\star - (1+\gamma w)(x^\star - x_n))}{\overline{F}(x_n)} = (1+\gamma w)^{-1/\gamma}.$$

Then, we find, using Potter bounds (see Proposition B.1.9.5 on p.367 of de Haan and Ferreira, 2006) and the 1-homogeneity of the function R_t in condition $\mathcal{D}$ (as a direct consequence of its definition) along with (A.5) that, for any $w, w' \in [0, -1/\gamma)$ and $t \geq 1$,

$$\frac{1}{\overline{F}(u_n)} \mathbb{P}(\overline{F}(X_1) \leq \overline{F}(x^\star - (1+\gamma w)(x^\star - x_n)), \overline{F}(X_{t+1}) \leq \overline{F}(x^\star - (1+\gamma w')(x^\star - x_n)))$$
$$\to R_t((1+\gamma w)^{-1/\gamma}, (1+\gamma w')^{-1/\gamma}) \text{ as } n \to \infty. \quad (A.9)$$

We now assume that $s_n \to \infty$ and we prove (i). Fix $\varepsilon \in (0, -1/\gamma)$. From condition $\mathcal{D}$, Potter bounds and (A.5) again, we have, for n large enough,

$$\frac{1}{\overline{F}(u_n)} \mathbb{P}(\overline{F}(X_1) \leq \overline{F}(x^\star - (1+\gamma w)(x^\star - x_n)), \overline{F}(X_{t+1}) \leq \overline{F}(x^\star - (1+\gamma w')(x^\star - x_n)))$$
$$\leq C\left(\rho(t)\sqrt{(1+\gamma w)^{-1/\gamma-\varepsilon}(1+\gamma w')^{-1/\gamma-\varepsilon}} + \overline{F}(u_n)(1+\gamma w)^{-1/\gamma-\varepsilon}(1+\gamma w')^{-1/\gamma-\varepsilon}\right) \quad (A.10)$$

for any $t \geq 1$ and any $w, w' \in [0, -1/\gamma)$, where C is a positive constant. Notice that for any $t < s_n$,

$$\rho(t)\sqrt{(1+\gamma w)^{-1/\gamma-\varepsilon}(1+\gamma w')^{-1/\gamma-\varepsilon}} + \overline{F}(u_n)(1+\gamma w)^{-1/\gamma-\varepsilon}(1+\gamma w')^{-1/\gamma-\varepsilon}$$
$$\to \rho(t)\sqrt{(1+\gamma w)^{-1/\gamma-\varepsilon}(1+\gamma w')^{-1/\gamma-\varepsilon}} \quad (A.11)$$

as $n \to \infty$, and

$$\sum_{t=1}^{\infty} \iint_{[0,-1/\gamma)^2} \left(\rho(t)\sqrt{(1+\gamma w)^{-1/\gamma-\varepsilon}(1+\gamma w')^{-1/\gamma-\varepsilon}}\right.$$
$$\left. + \overline{F}(u_n)(1+\gamma w)^{-1/\gamma-\varepsilon}(1+\gamma w')^{-1/\gamma-\varepsilon}\right)\mathrm{d}w\,\mathrm{d}w'\,\mathbb{1}\{t < s_n\}$$
$$\to \left(\int_{[0,-1/\gamma)^2} \sqrt{(1+\gamma w)^{-1/\gamma-\varepsilon}}\,\mathrm{d}w\right)^2 \sum_{t=1}^{\infty} \rho(t) < \infty \quad (A.12)$$

by splitting the sum and using the assumption that $s_n\overline{F}(u_n) = \mathrm{O}(r_n\overline{F}(u_n)) \to 0$. Combine Theorem 1 in Pratt (1960) with (A.8), (A.9), (A.10), (A.11) and (A.12) to get

$$\frac{1}{\overline{F}(u_n)[a(1/\overline{F}(u_n))]^2} \sum_{t=1}^{\infty} \left(1 - \frac{t}{s_n}\right) \iint_{[x_n, x^\star)^2} \mathbb{P}(X_1 > v, X_{t+1} > v')\,\mathrm{d}v\,\mathrm{d}v'\mathbb{1}\{t < s_n\}$$
$$\to \iint_{[0,-1/\gamma)^2} \sum_{t=1}^{\infty} R_t((1+\gamma w)^{-1/\gamma}, (1+\gamma w')^{-1/\gamma})\,\mathrm{d}w\,\mathrm{d}w'.$$

Plug this into (A.7) and use a change of variables to complete the proof of (i). To show (ii), write $s_n \le C' r_n$ where C' is a positive constant, and note that, by (A.7),

$$\begin{aligned}
&\frac{1}{s_n} \operatorname{Var}\left(\sum_{t=1}^{s_n}(X_t - x_n)\mathbb{1}\{X_t > x_n\}\right) \\
&\le \frac{2}{(1-\gamma)(1-2\gamma)} \overline{F}(u_n)[a(1/\overline{F}(u_n))]^2(1+\mathrm{o}(1)) \\
&+ 2\sum_{t=1}^{\infty}\left(1 - \frac{t}{C' r_n}\right) \iint_{[x_n, x^\star)^2} \mathbb{P}(X_1 > v, X_{t+1} > v')\,\mathrm{d}v\,\mathrm{d}v'\,\mathbb{1}\{t < C' r_n\}.
\end{aligned}$$

Follow then the proof of (i) and use (A.6) to obtain that the upper bound converges as $n \to \infty$. The desired conclusion is now immediate. □

Lemma A.5 below provides the asymptotic normality of the empirical survival function $\widehat{\overline{E}}_n$ at intermediate levels, when the data generating process is at least $\alpha-$mixing. The asymptotic normality of the intermediate LAWS estimator will follow from that result.

Lemma A.5. *Assume that X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$. Suppose that F is continuous and that $(X_t)_{t\ge 1}$ is a strictly stationary sequence of copies of X satisfying conditions $\mathcal{M}$ and $\mathcal{D}$. Let finally $u_n \uparrow x^\star$ be such that $n\overline{F}(u_n) \to \infty$, $r_n\overline{F}(u_n) \to 0$, and $x_n, x_n' \uparrow x^\star$ be such that $(x^\star - x_n)/(x^\star - u_n) \to 1$ and $(x^\star - x_n')/(x^\star - u_n) \to 1$.*

(i) If there is $\delta > 0$ such that $r_n(r_n/\sqrt{n\overline{F}(u_n)})^\delta \to 0$, then one has

$$\sqrt{n\overline{F}(u_n)}\left(\frac{\widehat{\varphi}_n^{(1)}(x_n)}{\varphi^{(1)}(x_n)} - 1, \frac{\widehat{\overline{F}}_n(x_n')}{\overline{F}(x_n')} - 1\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}(\gamma) + 2\,\boldsymbol{D}(\gamma, R))$$

where the 2×2 symmetric matrices $\boldsymbol{\Sigma}(\gamma)$ and $\boldsymbol{D}(\gamma, R)$ are defined elementwise as $\Sigma_{11}(\gamma) = 2(1-\gamma)/(1-2\gamma)$, $\Sigma_{12}(\gamma) = \Sigma_{22}(\gamma) = 1$,

$$D_{11}(\gamma, R) = (1-\gamma^{-1})^2 \iint_{(0,1]^2} \sum_{t=1}^{\infty} R_t(x^{-1/\gamma}, y^{-1/\gamma})\,\mathrm{d}x\,\mathrm{d}y$$

$$D_{12}(\gamma, R) = \frac{1}{2}(1-\gamma^{-1}) \int_0^1 \sum_{t=1}^{\infty}[R_t(x^{-1/\gamma}, 1) + R_t(1, x^{-1/\gamma})]\,\mathrm{d}x$$

and $D_{22}(\gamma, R) = \sum_{t=1}^{\infty} R_t(1,1).$

(ii) If, choosing $\delta > 0$ as in (i), one has $\mathbb{E}(|\min(X,0)|^{2+\delta}) < \infty$ and $\sum_{l\ge 1} l^{2/\delta}\alpha(l) < \infty$, then

$$\sqrt{n\overline{F}(u_n)}\left(\frac{\widehat{\overline{E}}_n(x_n)}{\overline{E}(x_n)} - 1, \frac{\widehat{\overline{F}}_n(x_n')}{\overline{F}(x_n')} - 1\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}(\gamma) + 2\,\boldsymbol{D}(\gamma, R)).$$

If X is bounded, then in (ii) assumption $\sum_{l\ge 1} l^{2/\delta}\alpha(l) < \infty$ can be weakened to $\sum_{l\ge 1}\alpha(l) < \infty$ and no integrability assumption on X is necessary.

If the X_i are in fact i.i.d. then both results hold with $\boldsymbol{D}(\gamma, R) = 0$ under the sole assumptions that X satisfies condition $\mathcal{C}_2(\gamma, a, \rho, A)$, $u_n \uparrow x^\star$ is such that $n\overline{F}(u_n) \to \infty$, and $x_n, x_n' \uparrow x^\star$ are such that $(x^\star - x_n)/(x^\star - u_n) \to 1$ and $(x^\star - x_n')/(x^\star - u_n) \to 1$, with the extra requirement that $\mathbb{E}(|\min(X,0)|^2) < \infty$ for (ii) only.

Proof of Lemma A.5. (i) Pick $(\lambda, \mu) \in \mathbb{R}^2 \setminus \{(0,0)\}$. Clearly

$$\sqrt{n\overline{F}(u_n)} \left\{ \lambda \left(\frac{\widehat{\varphi}_n^{(1)}(x_n)}{\varphi^{(1)}(x_n)} - 1 \right) + \mu \left(\frac{\widehat{\overline{F}}_n(x_n')}{\overline{F}(x_n')} - 1 \right) \right\}$$
$$= \sum_{t=1}^{n} \lambda \times \sqrt{\frac{\overline{F}(u_n)}{n}} \left(\frac{(X_t - x_n)\mathbb{1}\{X_t > x_n\}}{\mathbb{E}((X - x_n)\mathbb{1}\{X > x_n\})} - 1 \right) + \mu \times \sqrt{\frac{\overline{F}(u_n)}{n}} \left(\frac{\mathbb{1}\{X_t > x_n'\}}{\mathbb{P}(X > x_n')} - 1 \right)$$
$$= \sum_{t=1}^{n} \mathcal{X}_{n,t}(\lambda, \mu)$$

is a mean of identically distributed and centered random variables for every n. We start by the case when $(X_t)_{t\geq 1}$ is an $\alpha-$mixing sequence. We aim to apply Lemma C.7(ii) in Davison et al. (2023), of which we check each condition. By Lemma A.4, and using condition $\mathcal{M}$,

$$\frac{n}{r_n} \operatorname{Var}\left(\sum_{t=1}^{l_n} \mathcal{X}_{n,t}(\lambda, \mu) \right) = \mathrm{O}(l_n / r_n) \to 0,$$
$$\operatorname{Var}\left(\sum_{t=1}^{n - r_n \lfloor n/r_n \rfloor} \mathcal{X}_{n,t}(\lambda, \mu) \right) = \mathrm{O}((n - r_n \lfloor n/r_n \rfloor)/n) = \mathrm{O}(r_n/n) \to 0,$$
$$\lim_{n\to\infty} \frac{n}{r_n} \operatorname{Var}\left(\sum_{t=1}^{r_n} \mathcal{X}_{n,t}(\lambda, \mu) \right) = \lambda^2 (\Sigma_{11}(\gamma) + 2D_{11}(\gamma, R)) + 2\lambda\mu(\Sigma_{12}(\gamma) + 2D_{12}(\gamma, R)) + \mu^2(\Sigma_{22}(\gamma) + 2D_{22}(\gamma, R)).$$

Besides, for any $\varepsilon > 0$,

$$\frac{n}{r_n} \mathbb{E}\left(\left| \sum_{t=1}^{r_n} \mathcal{X}_{n,t}(\lambda, \mu) \right|^2 \mathbb{1}\left\{ \left| \sum_{t=1}^{r_n} \mathcal{X}_{n,t}(\lambda, \mu) \right| > \varepsilon \right\} \right) \leq \varepsilon^{-\delta} \times \frac{n}{r_n} \mathbb{E}\left(\left| \sum_{t=1}^{r_n} \mathcal{X}_{n,t}(\lambda, \mu) \right|^{2+\delta} \right)$$
$$= \mathrm{O}\left(n r_n^{1+\delta} \mathbb{E}(|\mathcal{X}_{n,1}(\lambda, \mu)|^{2+\delta}) \right)$$
$$= \mathrm{O}\left(r_n \left[\frac{r_n}{\sqrt{n\overline{F}(u_n)}} \right]^{\delta} \right)$$

by the Hölder inequality and Lemma A.2(i) and (ii). This converges to 0 by assumption, so Lemma C.7(ii) in Davison et al. (2023) applies and yields the desired conclusion in the $\alpha-$mixing framework thanks to the Cramér-Wold device. When the X_i are i.i.d., one may apply the standard Lyapunov central limit theorem (Billingsley, 1995, Theorem 27.3 p.362) instead: first of all

$$n \operatorname{Var}(\mathcal{X}_{n,1}(\lambda, \mu)) \to \lambda^2 \frac{2(1-\gamma)}{1-2\gamma} + 2\lambda\mu + \mu^2$$

by Lemma A.2, because of (A.5). Then, by the Hölder inequality and Lemma A.2(i) and (ii),

$$n\, \mathbb{E}|\mathcal{X}_{n,1}(\lambda, \mu)|^4 = \mathrm{O}\left(n \left(\frac{\overline{F}(u_n)}{n} \right)^2 \left(\frac{\varphi^{(4)}(x_n)}{[\varphi^{(1)}(x_n)]^4} + \frac{1}{[\overline{F}(u_n)]^3} \right) \right) = \mathrm{O}\left(\frac{1}{n\overline{F}(u_n)} \right).$$

This converges to 0, so the Lyapunov central limit theorem applies and the proof of (i) is complete.

To show (ii), write

$$\log \frac{\widehat{\overline{E}}_n(x_n)}{\overline{E}(x_n)} = \log \frac{\widehat{\varphi}_n^{(1)}(x_n)}{\varphi^{(1)}(x_n)} - \log \left(\frac{2\widehat{\varphi}_n^{(1)}(x_n) + x_n - \overline{X}_n}{2\varphi^{(1)}(x_n) + x_n - \mathbb{E}(X)} \right).$$

Since

$$\frac{2\widehat{\varphi}_n^{(1)}(x_n) + x_n - \overline{X}_n}{2\varphi^{(1)}(x_n) + x_n - \mathbb{E}(X)} - 1 = \frac{2(\widehat{\varphi}_n^{(1)}(x_n) - \varphi^{(1)}(x_n)) - (\overline{X}_n - \mathbb{E}(X))}{2\varphi^{(1)}(x_n) + x_n - \mathbb{E}(X)}$$
$$= \mathrm{O}_{\mathbb{P}} \left(\frac{\varphi^{(1)}(x_n)}{\sqrt{n\overline{F}(u_n)}} \times \sqrt{n\overline{F}(u_n)} \left(\frac{\widehat{\varphi}_n^{(1)}(x_n)}{\varphi^{(1)}(x_n)} - 1 \right) \right) + \mathrm{O}_{\mathbb{P}} \left(\frac{1}{\sqrt{n}} \times \sqrt{n}(\overline{X}_n - \mathbb{E}(X)) \right),$$

it follows that

$$\frac{2\widehat{\varphi}_n^{(1)}(x_n) + x_n - \overline{X}_n}{2\varphi^{(1)}(x_n) + x_n - \mathbb{E}(X)} - 1 = \mathrm{o}_{\mathbb{P}} \left(\frac{1}{\sqrt{n\overline{F}(u_n)}} \right)$$

by Lemma A.2(i) and Corollary 1.2 on p.10 of Rio (2017) along with (1.25a) and (1.25b) on p.12 therein, when the X_i are $\alpha-$mixing and under the assumptions $\mathbb{E}(|\min(X,0)|^{2+\delta}) < \infty$ and $\sum_{l\geq 1} l^{2/\delta}\alpha(l) < \infty$. [When X is also bounded, condition $\sum_{l\geq 1}\alpha(l) < \infty$ is sufficient, see (1.24) on p.11 of Rio (2017).] In the case when the X_i are i.i.d., the usual central limit theorem can be applied instead in order to control $\overline{X}_n - \mathbb{E}(X)$, under the condition $\mathbb{E}(|\min(X,0)|^2) < \infty$. Hence, in both cases, the equality

$$\log \frac{\widehat{\overline{E}}_n(x_n)}{\overline{E}(x_n)} = \log \frac{\widehat{\varphi}_n^{(1)}(x_n)}{\varphi^{(1)}(x_n)} + \mathrm{o}_{\mathbb{P}} \left(\frac{1}{\sqrt{n\overline{F}(u_n)}} \right)$$

from which (ii) follows by applying (i). □

A.2 Proofs of the main results

Proof of Proposition 1. The starting point is to combine Equation (2.2) and Lemma A.2, in order to obtain

$$x^\star - \mathbb{E}(X) - (x^\star - \xi_\tau)$$
$$= \frac{2\tau - 1}{1-\tau} \times \frac{\overline{F}(\xi_\tau)a(1/\overline{F}(\xi_\tau))}{1-\gamma} \left(1 + \frac{1}{1-\gamma-\rho} A(1/\overline{F}(\xi_\tau)) + \mathrm{o}(|A(1/\overline{F}(\xi_\tau))|) \right)$$

as $\tau \uparrow 1$. In other words,

$$\frac{\overline{F}(\xi_\tau)a(1/\overline{F}(\xi_\tau))}{1-\tau} = (1-\gamma)[(x^\star - \mathbb{E}(X)) - (x^\star - \xi_\tau)] \times (1 - 2(1-\tau))^{-1}$$
$$\times \left(1 - \frac{1}{1-\gamma-\rho} A(1/\overline{F}(\xi_\tau)) + \mathrm{o}(|A(1/\overline{F}(\xi_\tau))|) \right) \quad \text{as } \tau \uparrow 1. \qquad \text{(A.13)}$$

Then, by Lemma A.3(i) with $s = 1/\overline{F}(q_\tau)$ and Lemma A.3(ii) with $s = 1/\overline{F}(\xi_\tau)$ combined with Lemma A.1, one has the alternative expansion

$$\frac{\overline{F}(\xi_\tau)a(1/\overline{F}(\xi_\tau))}{1-\tau} = -\gamma\frac{(x^\star-\xi_\tau)^{1-1/\gamma}}{(x^\star-q_\tau)^{-1/\gamma}}\left(1+\frac{\rho+1}{\rho(\gamma+\rho)}A(1/\overline{F}(\xi_\tau))+\mathrm{o}(|A(1/\overline{F}(\xi_\tau))|)\right)$$
$$\times\left(1-\frac{1}{\rho(\gamma+\rho)}A((1-\tau)^{-1})+\mathrm{o}(|A((1-\tau)^{-1})|)\right) \quad \text{as } \tau\uparrow 1.$$

A consequence of Equation (2.3) is that $1/\overline{F}(\xi_\tau) = \mathrm{o}((1-\tau)^{-1})$. Therefore, since $|A|$ is regularly varying with index $\rho<0$, one has $A((1-\tau)^{-1}) = \mathrm{o}(|A(1/\overline{F}(\xi_\tau))|)$, from which it follows that

$$\frac{\overline{F}(\xi_\tau)a(1/\overline{F}(\xi_\tau))}{1-\tau} = -\gamma\frac{(x^\star-\xi_\tau)^{1-1/\gamma}}{(x^\star-q_\tau)^{-1/\gamma}}\left(1+\frac{\rho+1}{\rho(\gamma+\rho)}A(1/\overline{F}(\xi_\tau))+\mathrm{o}(|A(1/\overline{F}(\xi_\tau))|)\right) \tag{A.14}$$

as $\tau\uparrow 1$. Combine (A.13) and (A.14) to find

$$\frac{(x^\star-\xi_\tau)^{1-1/\gamma}}{(x^\star-q_\tau)^{-1/\gamma}} = (x^\star-\mathbb{E}(X))(1-\gamma^{-1})\left[1-\frac{1}{x^\star-\mathbb{E}(X)}(x^\star-\xi_\tau)\right]\times(1-2(1-\tau))^{-1}$$
$$\times\left(1-\frac{1-\gamma}{\rho(\gamma+\rho)(1-\gamma-\rho)}A(1/\overline{F}(\xi_\tau))+\mathrm{o}(|A(1/\overline{F}(\xi_\tau))|)\right) \quad \text{as } \tau\uparrow 1.$$

A consequence of Equation (2.4) is that $1-\tau = \mathrm{o}(x^\star-\xi_\tau)$. Hence

$$\frac{(x^\star-\xi_\tau)^{1-1/\gamma}}{(x^\star-q_\tau)^{-1/\gamma}} = (x^\star-\mathbb{E}(X))(1-\gamma^{-1})\left[1-\frac{1}{x^\star-\mathbb{E}(X)}(x^\star-\xi_\tau)+\mathrm{o}(x^\star-\xi_\tau)\right]$$
$$\times\left(1-\frac{1-\gamma}{\rho(\gamma+\rho)(1-\gamma-\rho)}A(1/\overline{F}(\xi_\tau))+\mathrm{o}(|A(1/\overline{F}(\xi_\tau))|)\right) \quad \text{as } \tau\uparrow 1. \tag{A.15}$$

Combine the above expansion with Equations (2.4) and (2.5) and the regular variation property of $|A|$ to get

$$\frac{(x^\star-\xi_\tau)^{1-1/\gamma}}{(x^\star-q_\tau)^{-1/\gamma}}$$
$$= (x^\star-\mathbb{E}(X))(1-\gamma^{-1})$$
$$\times\left[1-(x^\star-\mathbb{E}(X))^{-1/(1-\gamma)}(1-\gamma^{-1})^{-\gamma/(1-\gamma)}(x^\star-q_\tau)^{1/(1-\gamma)}(1+\mathrm{o}(1))\right]$$
$$\times\left(1-\frac{(1-\gamma)[(x^\star-\mathbb{E}(X))(1-\gamma^{-1})]^{-\rho/(1-\gamma)}}{\rho(\gamma+\rho)(1-\gamma-\rho)}A((1-\tau)^{-1}(x^\star-q_\tau)^{1/(1-\gamma)})(1+\mathrm{o}(1))\right)$$

as $\tau\uparrow 1$. Then clearly

$$x^\star-\xi_\tau = [(x^\star-\mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}(x^\star-q_\tau)^{1/(1-\gamma)}$$
$$\times\left[1-[(x^\star-\mathbb{E}(X))(1-\gamma^{-1})]^{-1/(1-\gamma)}(x^\star-q_\tau)^{1/(1-\gamma)}(1+\mathrm{o}(1))\right]$$
$$\times\left(1+\frac{\gamma[(x^\star-\mathbb{E}(X))(1-\gamma^{-1})]^{-\rho/(1-\gamma)}}{\rho(\gamma+\rho)(1-\gamma-\rho)}A((1-\tau)^{-1}(x^\star-q_\tau)^{1/(1-\gamma)})(1+\mathrm{o}(1))\right)$$

as $\tau\uparrow 1$, which is the first desired asymptotic expansion. The second statement is then a direct consequence of a combination of Lemma A.3(i), for $s = (1-\tau)^{-1}$, with this asymptotic expansion. □

Proof of Theorem 1. Fix $u, v \in \mathbb{R}$. To prove the desired joint convergence, it is sufficient to examine the convergence of the sequence

$$\mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \leq u, \ \frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{q}_{\pi_n} - q_{\pi_n}) \leq v\right).$$

First of all, since $a(s)/(x^\star - U(s)) \to -\gamma$ as $s \to \infty$, and $s \mapsto x^\star - U(s)$ is regularly varying, the assumption $\overline{F}(\xi_{\tau_n})/(1-\pi_n) \to 1$ yields

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} = \frac{\sqrt{n(1-\pi_n)}}{a((1-\pi_n)^{-1})}(1 + \mathrm{o}(1)).$$

As a result, it is equivalent to analyze the asymptotic behavior of

$$\mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \leq u, \ \frac{\sqrt{n(1-\pi_n)}}{a((1-\pi_n)^{-1})}(\widehat{q}_{\pi_n} - q_{\pi_n}) \leq v\right).$$

The key observation in order to do so is that, for fixed $u, v \in \mathbb{R}$, if

$$x_n = x_n(u) = \xi_{\tau_n} + u\frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}} \text{ and } x'_n = x'_n(v) = q_{\pi_n} + v\frac{a((1-\pi_n)^{-1})}{\sqrt{n(1-\pi_n)}}$$

then, following a simple calculation,

$$\begin{aligned}
&\mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \leq u, \ \frac{\sqrt{n(1-\pi_n)}}{a((1-\pi_n)^{-1})}(\widehat{q}_{\pi_n} - q_{\pi_n}) \leq v\right) \\
&= \mathbb{P}\left(\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\widehat{\overline{E}}_n(x_n)}{\overline{E}(x_n)} - 1\right) \leq \sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\overline{E}(\xi_{\tau_n})}{\overline{E}(x_n)} - 1\right),\right. \\
&\qquad \left.\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\widehat{\overline{F}}_n(x'_n)}{\overline{F}(x'_n)} - 1\right) \leq \sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{1-\pi_n}{\overline{F}(x'_n)} - 1\right)\right)
\end{aligned} \tag{A.16}$$

because ξ_{τ_n} (resp. $\widehat{\xi}_{\tau_n}$) is the τ_nth quantile of the continuous distribution function E (resp. the distribution function $\widehat{E}_n$), and likewise q_{π_n} (resp. $\widehat{q}_{\pi_n}$) is the π_nth quantile of the distribution function F (resp. the distribution function $\widehat{F}_n$). We first handle the right-hand sides of both of the inequalities in (A.16). Note that

$$\frac{x^\star - x_n}{x^\star - \xi_{\tau_n}} - 1 = -u\frac{a(1/\overline{F}(\xi_{\tau_n}))}{x^\star - \xi_{\tau_n}} \times \frac{1}{\sqrt{n\overline{F}(\xi_{\tau_n})}} = \mathrm{O}\left(\frac{1}{\sqrt{n\overline{F}(\xi_{\tau_n})}}\right) \to 0 \tag{A.17}$$

using the convergence $a(s)/(x^\star - U(s)) \to -\gamma$ as $s \to \infty$, and Lemma A.1(i). Note further that the function $\overline{E}$ is absolutely continuous on any compact interval, because

- The function $x \mapsto \varphi^{(1)}(x) = \int_x^\infty \overline{F}(y)\,\mathrm{d}y$ is Lipschitz continuous,
- The denominator $x \mapsto \mathbb{E}(|X - x|) = 2\varphi^{(1)}(x) + x - \mathbb{E}(X)$ of $\overline{E}$ defines a Lipschitz continuous function that is bounded away from zero.

A straightforward calculation shows that $\overline{E}$ has Lebesgue derivative

$$\overline{E}'(x) = -\frac{\varphi^{(1)}(x) + \overline{F}(x)(x - \mathbb{E}(X))}{(2\varphi^{(1)}(x) + x - \mathbb{E}(X))^2}.$$

In particular, it comes as a consequence of Lemma A.2(ii) that $-\overline{E}'(x)/\overline{F}(x) \to 1/(x^\star - \mathbb{E}(X))$ as $x \uparrow x^\star$. Then

$$\begin{aligned}
\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\overline{E}(\xi_{\tau_n})}{\overline{E}(x_n)} - 1\right) &= \sqrt{n\overline{F}(\xi_{\tau_n})}\int_{x_n}^{\xi_{\tau_n}} \frac{\overline{E}'(y)}{\overline{E}(x_n)}\mathrm{d}y \\
&= \sqrt{n\overline{F}(\xi_{\tau_n})}\int_{\xi_{\tau_n}}^{x_n} \frac{(1-\gamma)\overline{F}(y)}{\overline{F}(x_n)a(1/\overline{F}(x_n))}\mathrm{d}y(1 + \mathrm{o}(1)) \\
&= (1-\gamma) \times \sqrt{n\overline{F}(\xi_{\tau_n})}\frac{x_n - \xi_{\tau_n}}{a(1/\overline{F}(\xi_{\tau_n}))}(1 + \mathrm{o}(1)) \\
&\to (1-\gamma)u
\end{aligned} \tag{A.18}$$

as $n \to \infty$, by combining (A.5) with Lemma A.2(ii) and (A.17). Besides

$$\frac{x^\star - x_n'}{x^\star - \xi_{\tau_n}} - 1 = \left(\frac{x^\star - q_{\pi_n}}{x^\star - \xi_{\tau_n}} - 1\right) - v\frac{a((1-\pi_n)^{-1})}{x^\star - q_{\pi_n}} \times \frac{1 + \mathrm{o}(1)}{\sqrt{n\overline{F}(\xi_{\tau_n})}} \to 0 \tag{A.19}$$

because of the convergence $(x^\star - q_{\pi_n})/(x^\star - \xi_{\tau_n}) \to 1$, granted by the assumption $\overline{F}(\xi_{\tau_n})/(1-\pi_n) \to 1$, the regular variation property of $s \mapsto x^\star - U(s)$ and convergence $a(s)/(x^\star - U(s)) \to -\gamma$ as $s \to \infty$, and Lemma A.1(i). Then

$$\begin{aligned}
\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{1-\pi_n}{\overline{F}(x_n')} - 1\right) &= \sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\overline{F}(q_{\pi_n})}{\overline{F}(x_n')} - 1\right) \\
&\quad + \sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{1-\pi_n}{\overline{F}(q_{\pi_n})} - 1\right)\frac{\overline{F}(q_{\pi_n})}{\overline{F}(x_n')} \\
&= \sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\overline{F}(q_{\pi_n})}{\overline{F}(x_n')} - 1\right) + \mathrm{o}(1)
\end{aligned}$$

because of Lemma A.1(ii), the asymptotic equivalence between $\overline{F}(q_{\pi_n})$, $\overline{F}(\xi_{\tau_n})$ and $\overline{F}(x_n')$ due (in part) to (A.19), assumption $\sqrt{n\overline{F}(\xi_{\tau_n})}A(1/\overline{F}(\xi_{\tau_n})) = \mathrm{O}(1)$ and the regular variation property of $|A|$. Note now that the convergence in Theorem 2.3.8 on p.48 of de Haan and Ferreira (2006), which is equivalent to the convergence granted by condition $\mathcal{C}_2(\gamma, a, \rho, A)$, is actually locally uniform by Theorem B.3.19 on p.401 of de Haan and Ferreira (2006), and therefore

$$\begin{aligned}
\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\overline{F}(q_{\pi_n})}{\overline{F}(x_n')} - 1\right) &= \sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\overline{F}(q_{\pi_n})}{\overline{F}(q_{\pi_n} + v\,a((1-\pi_n)^{-1})/\sqrt{n(1-\pi_n)})} - 1\right) \\
&\to v
\end{aligned}$$

using the asymptotic equivalence between $1 - \pi_n$, $\overline{F}(q_{\pi_n})$ and $\overline{F}(\xi_{\tau_n})$. Conclude that

$$\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{1-\pi_n}{\overline{F}(x_n')} - 1\right) \to v. \tag{A.20}$$

Combine (A.16), (A.18) and (A.20) to get

$$\mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n}-\xi_{\tau_n})\le u,\ \frac{\sqrt{n(1-\pi_n)}}{a((1-\pi_n)^{-1})}(\widehat{q}_{\pi_n}-q_{\pi_n})\le v\right)$$
$$=\mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{1-\gamma}\left(\frac{\widehat{\overline{E}}_n(x_n)}{\overline{E}(x_n)}-1\right)+\mathrm{o}(1)\le u,\ \sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\widehat{\overline{F}}_n(x_n')}{\overline{F}(x_n')}-1\right)+\mathrm{o}(1)\le v\right).$$

Use Lemma A.5(ii) (this is allowed because of (A.17) and (A.19)) to conclude the proof. □

Proof of Corollary 2. Fix $v\in\mathbb{R}$. The proof relies on the identity

$$\left\{\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\widehat{\overline{F}}_n(\widehat{\xi}_{\tau_n})}{\overline{F}(\xi_{\tau_n})}-1\right)\le v\right\}=\{\widehat{q}_{\pi_n}\le\widehat{\xi}_{\tau_n}\}$$

where

$$\pi_n=\pi_n(v)=1-\overline{F}(\xi_{\tau_n})\left(1+\frac{v}{\sqrt{n\overline{F}(\xi_{\tau_n})}}\right).$$

We therefore investigate the asymptotic behavior of $\mathbb{P}(\widehat{q}_{\pi_n}\le\widehat{\xi}_{\tau_n})$. To this end we write

$$\{\widehat{q}_{\pi_n}\le\widehat{\xi}_{\tau_n}\}$$
$$=\left\{\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{q}_{\pi_n}-q_{\pi_n})-\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n}-\xi_{\tau_n})\le\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\xi_{\tau_n}-q_{\pi_n})\right\}.$$

Now

$$\xi_{\tau_n}-q_{\pi_n}=-\left[U\left(\frac{1}{\overline{F}(\xi_{\tau_n})}\left(1+\frac{v}{\sqrt{n\overline{F}(\xi_{\tau_n})}}\right)^{-1}\right)-U\left(\frac{1}{\overline{F}(\xi_{\tau_n})}\right)\right]+\mathrm{o}\left(\frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}}\right)$$

by Lemma A.1(i) and condition $\sqrt{n\overline{F}(\xi_{\tau_n})}A(1/\overline{F}(\xi_{\tau_n}))=\mathrm{O}(1)$. Since condition $\mathcal{C}_2(\gamma,a,\rho,A)$ holds locally uniformly in z, a Taylor expansion and condition $\sqrt{n\overline{F}(\xi_{\tau_n})}A(1/\overline{F}(\xi_{\tau_n}))=\mathrm{O}(1)$ yield

$$\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\xi_{\tau_n}-q_{\pi_n})\to v.$$

As a consequence

$$\{\widehat{q}_{\pi_n}\le\widehat{\xi}_{\tau_n}\}=\left\{\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{q}_{\pi_n}-q_{\pi_n})-\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n}-\xi_{\tau_n})+\mathrm{o}(1)\le v\right\}.$$

The conclusion follows by applying Theorem 1. □

Proof of Theorem 2. Recall from Equation (2.4) that $1-\tau_n = \mathrm{o}(\overline{F}(\xi_{\tau_n}))$ and write

$$\begin{aligned}
&\frac{\sqrt{n(1-\tau_n)}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}^{\star}_{1-p_n} - \xi_{1-p_n}) \\
&= \frac{\sqrt{n(1-\tau_n)}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \\
&+ \sqrt{n(1-\tau_n)}\left(\frac{\widehat{\sigma}_n}{a(1/\overline{F}(\xi_{\tau_n}))} - 1\right)\frac{((1-\tau_n)/p_n)^{\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - 1}{\widehat{\gamma}_n} \\
&+ \sqrt{n(1-\tau_n)}\left(\frac{((1-\tau_n)/p_n)^{\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - 1}{\widehat{\gamma}_n} - \frac{((1-\tau_n)/p_n)^{\gamma/(1-\gamma)} - 1}{\gamma}\right) \\
&+ \sqrt{n(1-\tau_n)}\left(\frac{((1-\tau_n)/p_n)^{\gamma/(1-\gamma)} - (\overline{F}(\xi_{\tau_n})/\overline{F}(\xi_{1-p_n}))^{\gamma}}{\gamma}\right) \\
&- \sqrt{n(1-\tau_n)}\left(\frac{\xi_{1-p_n} - \xi_{\tau_n}}{a(1/\overline{F}(\xi_{\tau_n}))} - \frac{(\overline{F}(\xi_{\tau_n})/\overline{F}(\xi_{1-p_n}))^{\gamma} - 1}{\gamma}\right) \qquad (A.21)\\
&= \sqrt{n(1-\tau_n)}\left(\frac{((1-\tau_n)/p_n)^{\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - 1}{\widehat{\gamma}_n} - \frac{((1-\tau_n)/p_n)^{\gamma/(1-\gamma)} - 1}{\gamma}\right) \\
&+ \sqrt{n(1-\tau_n)}\left(\frac{((1-\tau_n)/p_n)^{\gamma/(1-\gamma)} - (\overline{F}(\xi_{\tau_n})/\overline{F}(\xi_{1-p_n}))^{\gamma}}{\gamma}\right) + \mathrm{o}_{\mathbb{P}}(1) \qquad (A.22)
\end{aligned}$$

by Theorem 1 (for the control of the first term in (A.21)), the assumption on $\widehat{\sigma}_n$, the convergence of $\widehat{\gamma}_n$ to $\gamma < 0$, the assumption $(1-\tau_n)/p_n \to \infty$ (for the second term in (A.21)), Lemma A.1 and the arguments leading to the control of the nonrandom bias term IV in the proof of Theorem 4.3.1 on p.134 of de Haan and Ferreira (2006) (for the fifth term in (A.21)). Now

$$\begin{aligned}
&\frac{((1-\tau_n)/p_n)^{\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - 1}{\widehat{\gamma}_n} - \frac{((1-\tau_n)/p_n)^{\gamma/(1-\gamma)} - 1}{\gamma} \\
&= \frac{((1-\tau_n)/p_n)^{\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - ((1-\tau_n)/p_n)^{\gamma/(1-\gamma)}}{\widehat{\gamma}_n} + [((1-\tau_n)/p_n)^{\gamma/(1-\gamma)} - 1](\widehat{\gamma}_n^{-1} - \gamma^{-1}) \\
&= \frac{((1-\tau_n)/p_n)^{\gamma/(1-\gamma)}}{\gamma}(((1-\tau_n)/p_n)^{\widehat{\gamma}_n/(1-\widehat{\gamma}_n)-\gamma/(1-\gamma)} - 1)(1+\mathrm{o}_{\mathbb{P}}(1)) + \frac{\widehat{\gamma}_n - \gamma}{\gamma^2}(1+\mathrm{o}_{\mathbb{P}}(1)) \\
&= \frac{\widehat{\gamma}_n - \gamma}{\gamma^2}(1+\mathrm{o}_{\mathbb{P}}(1)) + \mathrm{O}_{\mathbb{P}}\left(\frac{((1-\tau_n)/p_n)^{\gamma/(1-\gamma)}\log((1-\tau_n)/p_n)}{\sqrt{n(1-\tau_n)}}\right) \\
&= \frac{\widehat{\gamma}_n - \gamma}{\gamma^2} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right) \qquad (A.23)
\end{aligned}$$

because $\widehat{\gamma}_n$ is $\sqrt{n(1-\tau_n)}$–consistent and $x^{-c}\log x \to 0$ as $x \to \infty$ for any $c > 0$. Finally, combining Lemma A.3 with Corollary 1 results in

$$\frac{\overline{F}(\xi_{\tau_n})}{\overline{F}(\xi_{1-p_n})} = \left(\frac{1-\tau_n}{p_n}\right)^{1/(1-\gamma)}\Big(1 + \mathrm{O}((x^\star - q_{\tau_n})^{1/(1-\gamma)}) + \mathrm{O}(|A((1-\tau_n)^{-1})|)\Big) \qquad (A.24)$$

because $1/\overline{F}(\xi_\tau) = \mathrm{o}((1-\tau)^{-1})$ as $\tau \uparrow 1$, and therefore $A((1-\tau_n)^{-1}) = \mathrm{o}(|A(1/\overline{F}(\xi_{\tau_n}))|)$. Combine (A.22), (A.23) and (A.24) to complete the proof. □

Proof of Theorem 3. We decompose $\widetilde{\xi}^{\star}_{1-p_n} - \xi_{1-p_n}$ in the following way:

$$\begin{aligned}
&\widetilde{\xi}^{\star}_{1-p_n} - \xi_{1-p_n} \\
&= \widehat{x}^{\star} - x^{\star} - (\widehat{x}^{\star} - \widetilde{\xi}^{\star}_{1-p_n} - (x^{\star} - \xi_{1-p_n})) \\
&= \widehat{x}^{\star} - x^{\star} \\
&- \{[(\widehat{x}^{\star} - \overline{X}_n)(1-\widehat{\gamma}_n^{-1})]^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - [(x^{\star} - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}]\}(\widehat{x}^{\star} - \widehat{q}^{\star}_{1-p_n})^{1/(1-\widehat{\gamma}_n)} \\
&- [(x^{\star} - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}[(\widehat{x}^{\star} - \widehat{q}^{\star}_{1-p_n})^{1/(1-\widehat{\gamma}_n)} - (x^{\star} - q_{1-p_n})^{1/(1-\gamma)}] \\
&+ (x^{\star} - \xi_{1-p_n}) - [(x^{\star} - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}(x^{\star} - q_{1-p_n})^{1/(1-\gamma)}.
\end{aligned}$$

By Proposition 1 and Lemma A.3(i), it follows that

$$\begin{aligned}
&\widetilde{\xi}^{\star}_{1-p_n} - \xi_{1-p_n} \\
&= \widehat{x}^{\star} - x^{\star} \\
&- \{[(\widehat{x}^{\star} - \overline{X}_n)(1-\widehat{\gamma}_n^{-1})]^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - [(x^{\star} - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}]\}(\widehat{x}^{\star} - \widehat{q}^{\star}_{1-p_n})^{1/(1-\widehat{\gamma}_n)} \\
&- [(x^{\star} - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}[(\widehat{x}^{\star} - \widehat{q}^{\star}_{1-p_n})^{1/(1-\widehat{\gamma}_n)} - (x^{\star} - q_{1-p_n})^{1/(1-\gamma)}] \\
&+ \mathrm{O}(p_n^{-\gamma/(1-\gamma)}(p_n^{-\gamma/(1-\gamma)} + |A(p_n^{-1/(1-\gamma)})|)). \qquad \text{(A.25)}
\end{aligned}$$

We control each of the three terms in (A.25). By the Skorokhod lemma, up to changing the probability space and with appropriate versions of the estimators involved, Theorem 4.5.1 on p.146 of de Haan and Ferreira (2006) provides

$$\widehat{x}^{\star} - x^{\star} = \frac{a(n/k)}{\sqrt{k}} \times \frac{1}{\gamma^2}\left(\Gamma + \gamma^2 B - \gamma\Lambda - \lambda\frac{\gamma}{\gamma+\rho} + \mathrm{o}_{\mathbb{P}}(1)\right). \qquad \text{(A.26)}$$

Writing

$$\begin{aligned}
&[(\widehat{x}^{\star} - \overline{X}_n)(1-\widehat{\gamma}_n^{-1})]^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - [(x^{\star} - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)} \\
&= \left[\left(\frac{\widehat{x}^{\star} - \overline{X}_n}{x^{\star} - \mathbb{E}(X)}\right)^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - 1\right][(x^{\star} - \mathbb{E}(X))(1-\widehat{\gamma}_n^{-1})]^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} \\
&+ [(x^{\star} - \mathbb{E}(X))(1-\widehat{\gamma}_n^{-1})]^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - [(x^{\star} - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)},
\end{aligned}$$

and combining (A.26) with the assumption that $\sqrt{k}(\overline{X}_n - \mathbb{E}(X)) \xrightarrow{d} 0$ and the delta-method, we find

$$[(\widehat{x}^{\star} - \overline{X}_n)(1-\widehat{\gamma}_n^{-1})]^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - [(x^{\star} - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)} = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right). \qquad \text{(A.27)}$$

It follows from our assumptions that $\log(k/(np_n))/\sqrt{k} = \mathrm{O}(\log(n)/\sqrt{k}) \to 0$, and therefore

$$\begin{aligned}
\widehat{x}^{\star} - \widehat{q}^{\star}_{1-p_n} &= -\widehat{a}(n/k)\frac{(k/(np_n))^{\widehat{\gamma}_n}}{\widehat{\gamma}_n} \\
&= -a(n/k)\frac{(k/(np_n))^{\gamma}}{\gamma}\left(1 + \frac{\log(k/(np_n))}{\sqrt{k}}\Gamma + \mathrm{o}_{\mathbb{P}}\left(\frac{\log n}{\sqrt{k}}\right)\right). \qquad \text{(A.28)}
\end{aligned}$$

Recalling that, by Lemma A.3(i), $(k/n)^{\gamma}a(n/k) \to -\gamma C < \infty$ with $C = \lim_{s\to\infty} s^{-\gamma}(x^{\star} - U(s))$, it follows that

$$\begin{aligned}
&(\widehat{x}^{\star} - \widehat{q}^{\star}_{1-p_n})^{1/(1-\widehat{\gamma}_n) - 1/(1-\gamma)} \\
&= \left(-a(n/k)\frac{(k/(np_n))^{\gamma}}{\gamma}\right)^{1/(1-\widehat{\gamma}_n) - 1/(1-\gamma)}\left(1 + \mathrm{o}_{\mathbb{P}}\left(\frac{\log n}{\sqrt{k}}\right)\right) \\
&= 1 - \frac{\log(p_n)}{\sqrt{k}} \times \frac{\gamma}{(1-\gamma)^2}(\Gamma + \mathrm{o}_{\mathbb{P}}(1)) + \mathrm{o}_{\mathbb{P}}\left(\frac{\log n}{\sqrt{k}}\right).
\end{aligned}$$

Hence the asymptotic expansion

$$\begin{aligned}
&(\widehat{x}^\star - \widehat{q}^\star_{1-p_n})^{1/(1-\widehat{\gamma}_n)} \\
&= \left(-a(n/k)\frac{(k/(np_n))^\gamma}{\gamma}\right)^{1/(1-\gamma)} \\
&\times \left(1 + \frac{\log(k/(np_n))}{\sqrt{k}} \times \frac{1}{1-\gamma}\Gamma - \frac{\log(p_n)}{\sqrt{k}} \times \frac{\gamma}{(1-\gamma)^2}\Gamma + \mathrm{o}_{\mathbb{P}}\left(\frac{\log n}{\sqrt{k}}\right)\right) \\
&= \left(-a(n/k)\frac{(k/(np_n))^\gamma}{\gamma}\right)^{1/(1-\gamma)} \left(1 + \frac{\log(k/(np_n^{1/(1-\gamma)}))}{\sqrt{k}} \times \frac{1}{1-\gamma}\Gamma + \mathrm{o}_{\mathbb{P}}\left(\frac{\log n}{\sqrt{k}}\right)\right) \quad \text{(A.29)} \\
&= \mathrm{O}_{\mathbb{P}}\left(\left(-a(n/k)\frac{(k/(np_n))^\gamma}{\gamma}\right)^{1/(1-\gamma)}\right) \quad \text{(A.30)}
\end{aligned}$$

where the convergence $(k/n)^\gamma a(n/k) \to -\gamma C < \infty$ was used. Combining (A.27) and (A.30) results in particular in

$$\begin{aligned}
&\{[(\widehat{x}^\star - \overline{X}_n)(1-\widehat{\gamma}_n^{-1})]^{-\widehat{\gamma}_n/(1-\widehat{\gamma}_n)} - [(x^\star - \mathbb{E}(X))(1-\gamma^{-1})]^{-\gamma/(1-\gamma)}]\}(\widehat{x}^\star - \widehat{q}^\star_{1-p_n})^{1/(1-\widehat{\gamma}_n)} \\
&= \mathrm{o}_{\mathbb{P}}\left(\left(-a(n/k)\frac{(k/(np_n))^\gamma}{\gamma}\right)^{1/(1-\gamma)} \frac{\log n}{\sqrt{k}}\right). \quad \text{(A.31)}
\end{aligned}$$

Using Lemma A.3(i) and the convergence $(k/n)^\gamma a(n/k) \to -\gamma C < \infty$ again in conjunction with the regular variation property of $|A|$,

$$\begin{aligned}
x^\star - q_{1-p_n} &= \frac{x^\star - q_{1-p_n}}{a(1/p_n)}\frac{a(1/p_n)}{a(n/k)}a(n/k) \\
&= -a(n/k)\frac{(k/(np_n))^\gamma}{\gamma}\left(1 - \frac{1}{\sqrt{k}} \times \frac{\lambda}{\rho} + \mathrm{o}\left(\frac{1}{\sqrt{k}}\right)\right) \quad \text{(A.32)}
\end{aligned}$$

and so

$$(x^\star - q_{1-p_n})^{1/(1-\gamma)} = \left(-a(n/k)\frac{(k/(np_n))^\gamma}{\gamma}\right)^{1/(1-\gamma)} \left(1 - \frac{1}{\sqrt{k}} \times \frac{\lambda}{\rho(1-\gamma)} + \mathrm{o}\left(\frac{1}{\sqrt{k}}\right)\right). \quad \text{(A.33)}$$

Combine (A.29) and (A.33) to obtain

$$\begin{aligned}
&(\widehat{x}^\star - \widehat{q}^\star_{1-p_n})^{1/(1-\widehat{\gamma}_n)} - (x^\star - q_{1-p_n})^{1/(1-\gamma)} \\
&= \left(-a(n/k)\frac{(k/(np_n))^\gamma}{\gamma}\right)^{1/(1-\gamma)} \left(\frac{\log(k/(np_n^{1/(1-\gamma)}))}{\sqrt{k}} \frac{\Gamma}{1-\gamma} + \mathrm{o}_{\mathbb{P}}\left(\frac{\log n}{\sqrt{k}}\right)\right). \quad \text{(A.34)}
\end{aligned}$$

Finally, combine (A.26), (A.31) and (A.34) to complete the proof. □

Proof of Corollary 3. Under the assumptions of the result, and by Lemma A.3(i), $a(n/k)$ is asymptotically proportional to $n^{(1-\chi)\gamma}$, and $[a(n/k)(k/(np_n))^\gamma]^{1/(1-\gamma)} \log(np_n^{1/(1-\gamma)}/k)$ is asymptotically proportional to $n^{\omega\gamma/(1-\gamma)} \log n$. The assumption $\chi < 1 - \omega/(1-\gamma)$ therefore ensures

$$\frac{a(n/k)}{\sqrt{k}} = \mathrm{o}\left(\frac{[a(n/k)(k/(np_n))^\gamma]^{1/(1-\gamma)} \log(np_n^{1/(1-\gamma)}/k)}{\sqrt{k}}\right).$$

Moreover, it is a consequence of Proposition B.1.9.1 on p.366 of de Haan and Ferreira (2006) that $|A(s)| = \mathrm{o}(s^{\rho+\varepsilon})$ as $s \to \infty$ for any $\varepsilon > 0$, so $A(n^{\omega/(1-\gamma)}) = \mathrm{o}(n^{\omega\rho/(1-\gamma)+\delta})$ for any $\delta > 0$. Hence, using the assumption $\chi < 2\omega \min(-\gamma, -\rho)/(1-\gamma)$, the convergence

$$\frac{\sqrt{k}}{[a(n/k)(k/(np_n))^\gamma]^{1/(1-\gamma)} \log(np_n^{1/(1-\gamma)}/k)} \times n^{\omega\gamma/(1-\gamma)}(n^{\omega\gamma/(1-\gamma)} + |A(n^{\omega/(1-\gamma)})|) \to 0.$$

Apply Theorem 3 to complete the proof. □

Proof of Corollary 4. It was shown in the proof of Corollary 3 that $a(n/k)$ is asymptotically proportional to $n^{(1-\chi)\gamma}$, and $[a(n/k)(k/(np_n))^\gamma]^{1/(1-\gamma)} \log(np_n^{1/(1-\gamma)}/k)$ is asymptotically proportional to $n^{\omega\gamma/(1-\gamma)} \log n$. Under the assumptions of the result,

$$\frac{[a(n/k)(k/(np_n))^\gamma]^{1/(1-\gamma)} \log(np_n^{1/(1-\gamma)}/k)}{\sqrt{k}} = \mathrm{o}\left(\frac{a(n/k)}{\sqrt{k}}\right).$$

Apply Theorem 3 to complete the proof. □

Proof of Proposition 2. Write

$$\begin{aligned}\log \frac{1-\widehat{\tau}_n}{1-\tau_n} &= \log\left(\frac{\widehat{x}^\star - \widehat{q}^\star_{1-p_n}}{x^\star - q_{1-p_n}}\right) - \log\left(\frac{\widehat{x}^\star - \overline{X}_n}{x^\star - \mathbb{E}(X)}\right) - \log\left(\frac{1-\widehat{\gamma}_n^{-1}}{1-\gamma^{-1}}\right) \\ &\quad + \log\left(\frac{x^\star - q_{1-p_n}}{(x^\star - \mathbb{E}(X))(1-\gamma^{-1})} \frac{p_n}{1-\tau_n}\right). \end{aligned} \tag{A.35}$$

Note that condition $\chi < \min(-2\omega\gamma, -2\rho/(1-2\rho))$ ensures in particular that $\sqrt{k}A(n/k) \to 0$. Combine (A.26), (A.28), (A.32), convergence $\sqrt{k}(\overline{X}_n - \mathbb{E}(X)) \xrightarrow{\mathbb{P}} 0$ and the delta-method to get

$$\begin{aligned}&\frac{\sqrt{k}}{\log(k/(np_n))}\left(\log\left(\frac{\widehat{x}^\star - \widehat{q}^\star_{1-p_n}}{x^\star - q_{1-p_n}}\right) - \log\left(\frac{\widehat{x}^\star - \overline{X}_n}{x^\star - \mathbb{E}(X)}\right) - \log\left(\frac{1-\widehat{\gamma}_n^{-1}}{1-\gamma^{-1}}\right)\right) \\ &= \frac{\sqrt{k}}{\log(k/(np_n))} \log\left(\frac{\widehat{x}^\star - \widehat{q}^\star_{1-p_n}}{x^\star - q_{1-p_n}}\right) + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{d} \Gamma. \end{aligned} \tag{A.36}$$

Combine now Proposition 1 and Lemma A.3 with the relationship $1-\tau = \mathrm{o}(x^\star - \xi_\tau)$ as $\tau \uparrow 1$ (coming as a consequence of Equation (2.4)) to get

$$\frac{(x^\star - \xi_\tau)\overline{F}(\xi_\tau)}{1-\tau} = (x^\star - \mathbb{E}(X))(1-\gamma^{-1})(1 + \mathrm{O}(x^\star - \xi_\tau) + \mathrm{O}(|A(1/\overline{F}(\xi_\tau))|)) \text{ as } \tau \uparrow 1.$$

With $\tau = \tau_n$ such that $\xi_\tau = \xi_{\tau_n} = q_{1-p_n}$ and using Lemma A.1(ii), we find

$$\frac{(x^\star - q_{1-p_n})}{(x^\star - \mathbb{E}(X))(1-\gamma^{-1})} \frac{p_n}{1-\tau_n} = 1 + \mathrm{O}(x^\star - q_{1-p_n}) + \mathrm{O}(|A(1/p_n)|) \text{ as } n \to \infty. \tag{A.37}$$

Assumptions $\chi + \omega - 1 > 0$ and $\chi < \min(-2\omega\gamma, -2\rho/(1-2\rho))$ ensure that $A(1/p_n) = \mathrm{o}(|A(n/k)|)$ and $\sqrt{k}(x^\star - q_{1-p_n}) \to 0$. Plug (A.36) and (A.37) into (A.35) to complete the proof. □

Proof of Theorem 4. We know that $\tau \mapsto \widehat{\xi}_\tau$ is the inverse of the distribution function $\widehat{E}_n = 1 - \widehat{\overline{E}}_n$ defined by

$$\widehat{\overline{E}}_n(x) = \frac{\widehat{\varphi}_n(x)}{2\widehat{\varphi}_n(x) + x - \frac{1}{n}\sum_{t=1}^n \widehat{\varepsilon}_t^{(n)}}, \text{ where } \widehat{\varphi}_n(x) = \widehat{\varphi}_n^{(1)}(x) = \frac{1}{n}\sum_{t=1}^n (\widehat{\varepsilon}_t^{(n)} - x)\mathbb{1}\{\widehat{\varepsilon}_t^{(n)} > x\}.$$

The conclusion of the proof of Theorem 1 contains the fact that for any fixed $u \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \le u\right) \\
&= \mathbb{P}\left(\sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\widehat{\overline{E}}_n(x_n)}{\overline{E}(x_n)} - 1\right) \le \sqrt{n\overline{F}(\xi_{\tau_n})}\left(\frac{\overline{E}(\xi_{\tau_n})}{\overline{E}(x_n)} - 1\right)\right) \\
&= \mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{1-\gamma}\left(\frac{\widehat{\overline{E}}_n(x_n)}{\overline{E}(x_n)} - 1\right) + \mathrm{o}(1) \le u\right),
\end{aligned}$$

where

$$x_n = x_n(u) = \xi_{\tau_n} + u\frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}}.$$

Define also the unfeasible, innovation-based LAWS estimator of $\xi_{\tau_n}(\varepsilon)$ by

$$\widetilde{\xi}_{\tau_n} = \arg\min_{\theta \in \mathbb{R}} \sum_{t=1}^{n} \eta_{\tau_n}(\varepsilon_t - \theta).$$

Then $\tau \mapsto \widetilde{\xi}_\tau$ is the inverse of the distribution function $\widetilde{E}_n = 1 - \widetilde{\overline{E}}_n$ given by

$$\widetilde{\overline{E}}_n(x) = \frac{\widetilde{\varphi}_n(x)}{2\widetilde{\varphi}_n(x) + x - \overline{\varepsilon}_n}, \text{ where } \widetilde{\varphi}_n(x) = \widetilde{\varphi}_n^{(1)}(x) = \frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t - x)\mathbb{1}\{\varepsilon_t > x\}$$

and $\overline{\varepsilon}_n$ is the sample mean of the ε_t, $1 \le t \le n$. We are going to prove that if $x_n = x_n(u)$ as above, then

$$\sqrt{n\overline{F}(\xi_{\tau_n})}\left|\frac{\widehat{\overline{E}}_n(x_n) - \widetilde{\overline{E}}_n(x_n)}{\overline{E}(x_n)}\right| \xrightarrow{\mathbb{P}} 0. \tag{A.38}$$

This will result in the fact that, for any fixed $u \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))}(\widehat{\xi}_{\tau_n} - \xi_{\tau_n}) \le u\right) = \mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{1-\gamma}\left(\frac{\widetilde{\overline{E}}_n(x_n)}{\overline{E}(x_n)} - 1\right) + \mathrm{o}_{\mathbb{P}}(1) \le u\right),$$

from which the conclusion will immediately follow by applying Lemma A.5(ii) to the i.i.d. sequence (ε_t).

Clearly

$$\begin{aligned}
\left|\widehat{\overline{E}}_n(x_n) - \widetilde{\overline{E}}_n(x_n)\right| &= \left|\frac{\widehat{\varphi}_n(x_n)}{2\widehat{\varphi}_n(x_n) + x_n - \frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_t^{(n)}} - \frac{\widetilde{\varphi}_n(x_n)}{2\widetilde{\varphi}_n(x_n) + x_n - \overline{\varepsilon}_n}\right| \\
&\le \widehat{\varphi}_n(x_n)\frac{2|\widehat{\varphi}_n(x_n) - \widetilde{\varphi}_n(x_n)| + |\frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_t^{(n)} - \overline{\varepsilon}_n|}{(2\widehat{\varphi}_n(x_n) + x_n - \frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_t^{(n)})(2\widetilde{\varphi}_n(x_n) + x_n - \overline{\varepsilon}_n)} \\
&\quad + \frac{|\widehat{\varphi}_n(x_n) - \widetilde{\varphi}_n(x_n)|}{2\widetilde{\varphi}_n(x_n) + x_n - \overline{\varepsilon}_n}.
\end{aligned} \tag{A.39}$$

Our assumption on $|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|$ immediately entails

$$\left| \frac{1}{n} \sum_{t=1}^{n} \widehat{\varepsilon}_t^{(n)} - \overline{\varepsilon}_n \right| \le \max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| = \mathrm{o}_{\mathbb{P}}\left(\frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}} \right). \tag{A.40}$$

In particular,

$$\frac{1}{n} \sum_{t=1}^{n} \widehat{\varepsilon}_t^{(n)} \xrightarrow{\mathbb{P}} 0 \tag{A.41}$$

from the law of large numbers and the fact that $a(x) \to 0$ as $x \uparrow e^\star$. Besides

$$\begin{aligned} |\widehat{\varphi}_n(x_n) - \widetilde{\varphi}_n(x_n)| &\le \frac{1}{n} \sum_{t=1}^{n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| \mathbb{1}\{\varepsilon_t > x_n\} \\ &+ \frac{1}{n} \sum_{t=1}^{n} |\widehat{\varepsilon}_t^{(n)} - x_n| \left| \mathbb{1}\{\widehat{\varepsilon}_t^{(n)} > x_n\} - \mathbb{1}\{\varepsilon_t > x_n\} \right|. \end{aligned} \tag{A.42}$$

Denoting the empirical survival function of the ε_t by $\widetilde{\overline{F}}_n$, we obviously have

$$\frac{1}{n} \sum_{t=1}^{n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| \mathbb{1}\{\varepsilon_t > x_n\} \le \widetilde{\overline{F}}_n(x_n) \max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| = \mathrm{o}_{\mathbb{P}}\left(\frac{\overline{F}(\xi_{\tau_n}) a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}} \right) \tag{A.43}$$

using our assumption on $|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|$, along with the Chebyshev inequality showing that $\widetilde{\overline{F}}_n(x_n)/\overline{F}(x_n) = 1 + \mathrm{o}_{\mathbb{P}}(1)$ and the asymptotic equivalence between $\overline{F}(x_n)$ and $\overline{F}(\xi_{\tau_n})$ due to a combination of (A.5) with (A.17) applied to the distribution of ε. Note then that, using again our assumption on $|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|$, one may define by induction a sequence of increasing integers N_k, for $k \ge 1$, such that for any $n > N_k$,

$$\mathbb{P}\left(\frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} \max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| > \frac{1}{k} \right) \le \frac{1}{2^k}.$$

Setting $\delta_n = 1/k$ when $n \in \{N_k + 1, \ldots, N_{k+1}\}$ results in a nonrandom positive sequence (δ_n) converging to 0 and such that the event

$$A_n = \left\{ \frac{\sqrt{n\overline{F}(\xi_{\tau_n})}}{a(1/\overline{F}(\xi_{\tau_n}))} \max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| \le \delta_n \right\}$$

has probability arbitrarily close to 1 as $n \to \infty$. On A_n,

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^{n} |\widehat{\varepsilon}_t^{(n)} - x_n| \left| \mathbb{1}\{\widehat{\varepsilon}_t^{(n)} > x_n\} - \mathbb{1}\{\varepsilon_t > x_n\} \right| \\ &\le \frac{1}{n} \sum_{t=1}^{n} |\widehat{\varepsilon}_t^{(n)} - x_n| \left(\mathbb{1}\{\varepsilon_t > x_{n,-}\} - \mathbb{1}\{\varepsilon_t > x_{n,+}\} \right) \\ &\le \frac{1}{n} \sum_{t=1}^{n} (|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| + |\varepsilon_t - x_n|) \left(\mathbb{1}\{\varepsilon_t > x_{n,-}\} - \mathbb{1}\{\varepsilon_t > x_{n,+}\} \right) \\ &\le 2\delta_n \frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}} \times \frac{1}{n} \sum_{t=1}^{n} \left(\mathbb{1}\{\varepsilon_t > x_{n,-}\} - \mathbb{1}\{\varepsilon_t > x_{n,+}\} \right) \end{aligned} \tag{A.44}$$

where

$$x_{n,+} = x_n + \delta_n \frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}} = \xi_{\tau_n} + (u+\delta_n)\frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}}$$
$$\text{and } x_{n,-} = x_n - \delta_n \frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}} = \xi_{\tau_n} + (u-\delta_n)\frac{a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}}.$$

The upper bound in (A.44) is positive, so it is stochastically bounded from above by its expectation in view of the Markov inequality. This means that

$$\frac{1}{n}\sum_{t=1}^{n} |\widehat{\varepsilon}_t^{(n)} - x_n| \left| \mathbb{1}\{\widehat{\varepsilon}_t^{(n)} > x_n\} - \mathbb{1}\{\varepsilon_t > x_n\} \right| = o_{\mathbb{P}}\left(\frac{\overline{F}(\xi_{\tau_n}) a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}} \right).$$

Combining this with (A.42) and (A.43) yields

$$|\widehat{\varphi}_n(x_n) - \widetilde{\varphi}_n(x_n)| = o_{\mathbb{P}}\left(\frac{\overline{F}(\xi_{\tau_n}) a(1/\overline{F}(\xi_{\tau_n}))}{\sqrt{n\overline{F}(\xi_{\tau_n})}} \right). \tag{A.45}$$

In particular, if $\varphi(x_n) = \mathbb{E}((\varepsilon - x_n)\mathbb{1}\{\varepsilon > x_n\})$,

$$\frac{\widetilde{\varphi}_n(x_n)}{\varphi(x_n)} \xrightarrow{\mathbb{P}} 1, \text{ so that } \frac{\widehat{\varphi}_n(x_n)}{\varphi(x_n)} \xrightarrow{\mathbb{P}} 1 \text{ and then } \widehat{\varphi}_n(x_n) \xrightarrow{\mathbb{P}} 0 \tag{A.46}$$

from Lemma A.5(i) in the i.i.d. setting. Finally, recalling that, from Lemma A.2(ii),

$$\overline{E}(x_n) = \frac{\varphi(x_n)}{2\varphi(x_n) + x_n - \mathbb{E}(\varepsilon)} = \frac{\varphi(x_n)}{2\varphi(x_n) + x_n}$$
$$\sim \frac{\varphi(x_n)}{e^\star} \sim \frac{\overline{F}(x_n) a(1/\overline{F}(x_n))}{(1-\gamma)e^\star} \sim \frac{\overline{F}(\xi_{\tau_n}) a(1/\overline{F}(\xi_{\tau_n}))}{(1-\gamma)e^\star}$$

as $n \to \infty$, (A.38) follows from combining (A.39), (A.40), (A.41), (A.45) and (A.46). □

B Further finite-sample results

We enclose here the full set of graphs we obtained in our numerical results for the six models we discuss in Section 3 and for the three sample sizes $n = 150, 300, 500$.

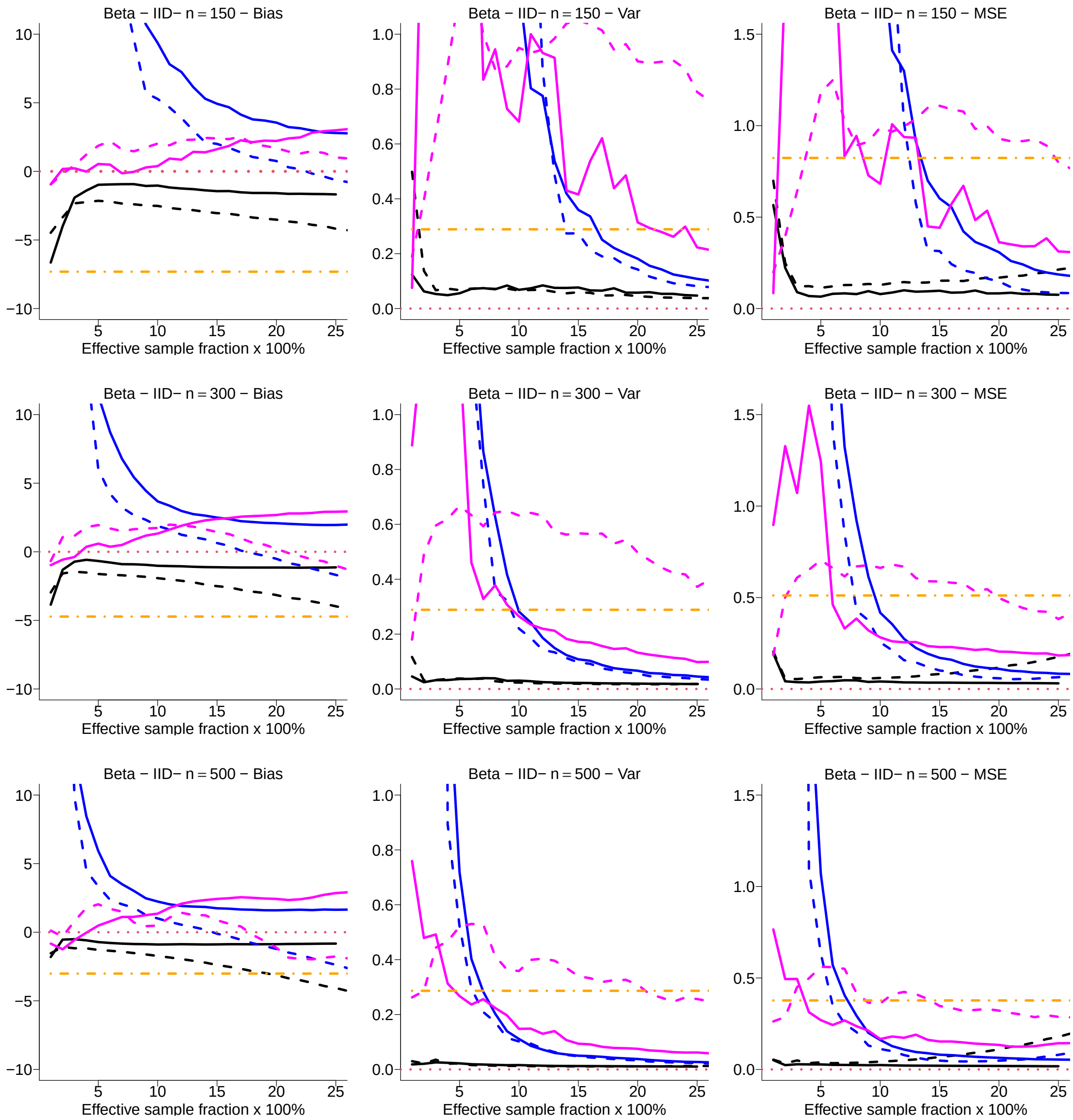

Figure B.1: Empirical relative bias, variance and MSE (left, middle and right), multiplied by 100, for the estimators of $\xi_{\tau_n'}$ obtained with i.i.d. observations from a Beta distribution (simulation setup (i)), $\tau_n' = 1 - 1/n$ and sample size $n = 150, 300, 500$ (top, middle, bottom). Purely empirical estimator $\widehat{\xi}_{\tau_n'}$ (orange line), extrapolating LAWS estimators $\widehat{\xi}^{\star}_{\tau_n'}$ (magenta lines) and $\overline{\xi}^{\star}_{\tau_n'}$ (blue lines), and extrapolating QB estimators $\widetilde{\xi}^{\star}_{\tau_n'}$ (black lines). The versions of the extrapolating estimators based on the GPML scale and shape parameter estimates are referred to using solid lines, and those based on the Moment estimators are referred to using dashed lines.

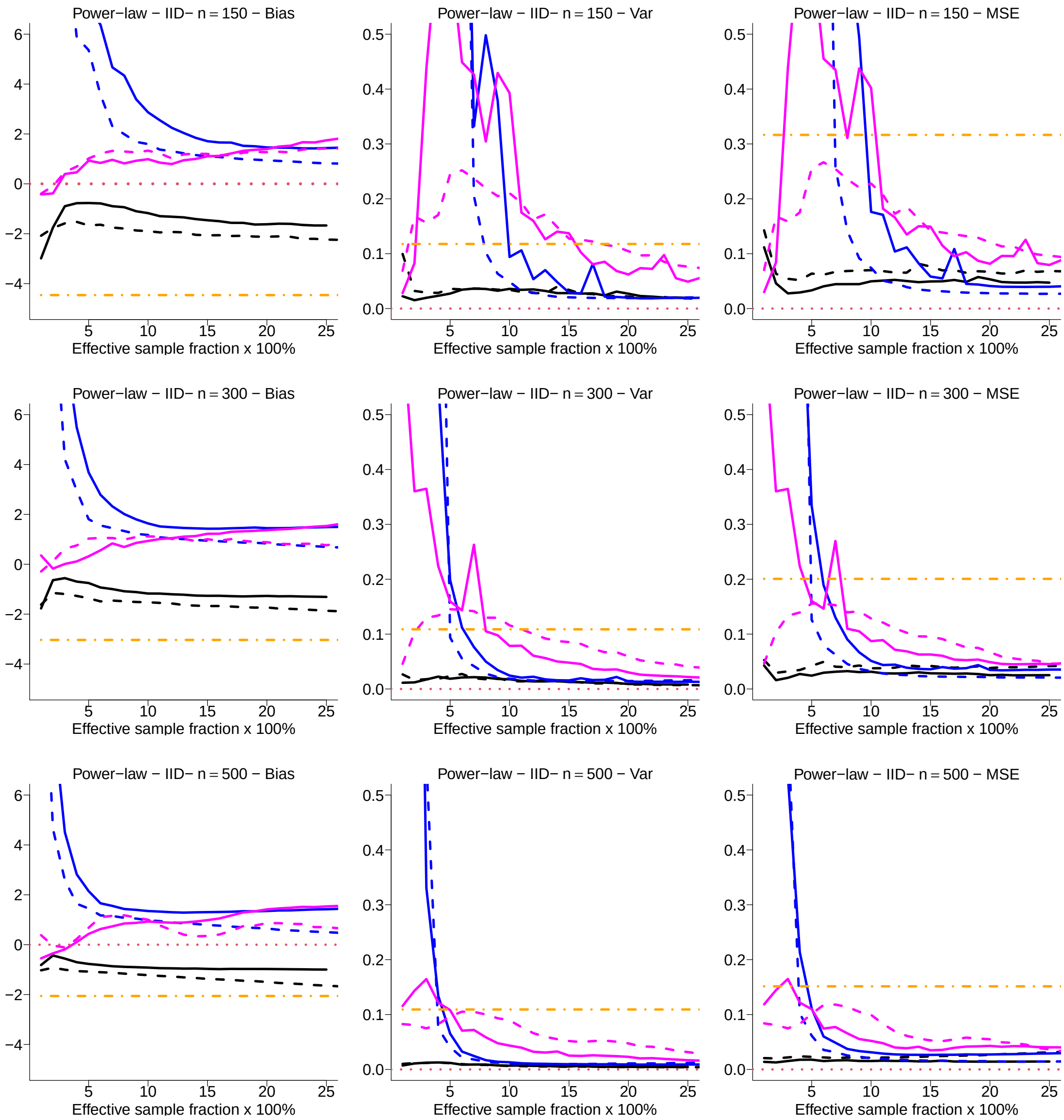

Figure B.2: Empirical relative bias, variance and MSE (left, middle and right), multiplied by 100, for the estimators of $\xi_{\tau'_n}$ obtained with i.i.d. observations from a power-law distribution (simulation setup (ii)), $\tau'_n = 1 - 1/n$ and sample size $n = 150, 300, 500$ (top, middle, bottom). Purely empirical estimator $\widehat{\xi}_{\tau'_n}$ (orange line), extrapolating LAWS estimators $\widehat{\xi}^{\star}_{\tau'_n}$ (magenta lines) and $\overline{\xi}^{\star}_{\tau'_n}$ (blue lines), and extrapolating QB estimators $\widetilde{\xi}^{\star}_{\tau'_n}$ (black lines). The versions of the extrapolating estimators based on the GPML scale and shape parameter estimates are referred to using solid lines, and those based on the Moment estimators are referred to using dashed lines.

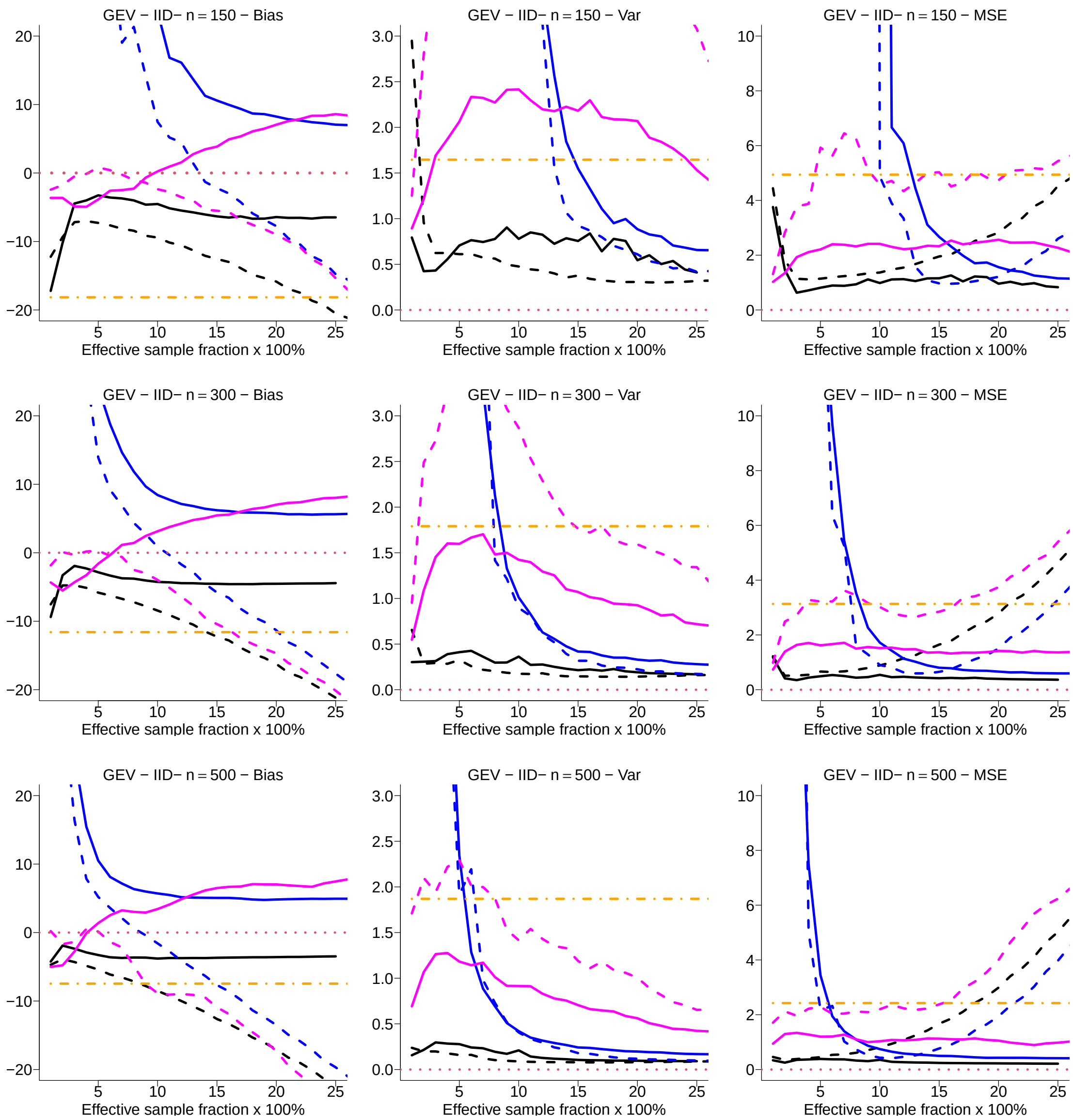

Figure B.3: Empirical relative bias, variance and MSE (left, middle and right), multiplied by 100, for the estimators of $\xi_{\tau_n'}$ obtained with i.i.d. observations from a GEV distribution (simulation setup (iii)), $\tau_n' = 1 - 1/n$ and sample size $n = 150, 300, 500$ (top, middle, bottom). Purely empirical estimator $\widehat{\xi}_{\tau_n'}$ (orange line), extrapolating LAWS estimators $\widehat{\xi}^{\star}_{\tau_n'}$ (magenta lines) and $\overline{\xi}^{\star}_{\tau_n'}$ (blue lines), and extrapolating QB estimators $\widetilde{\xi}^{\star}_{\tau_n'}$ (black lines). The versions of the extrapolating estimators based on the GPML scale and shape parameter estimates are referred to using solid lines, and those based on the Moment estimators are referred to using dashed lines.

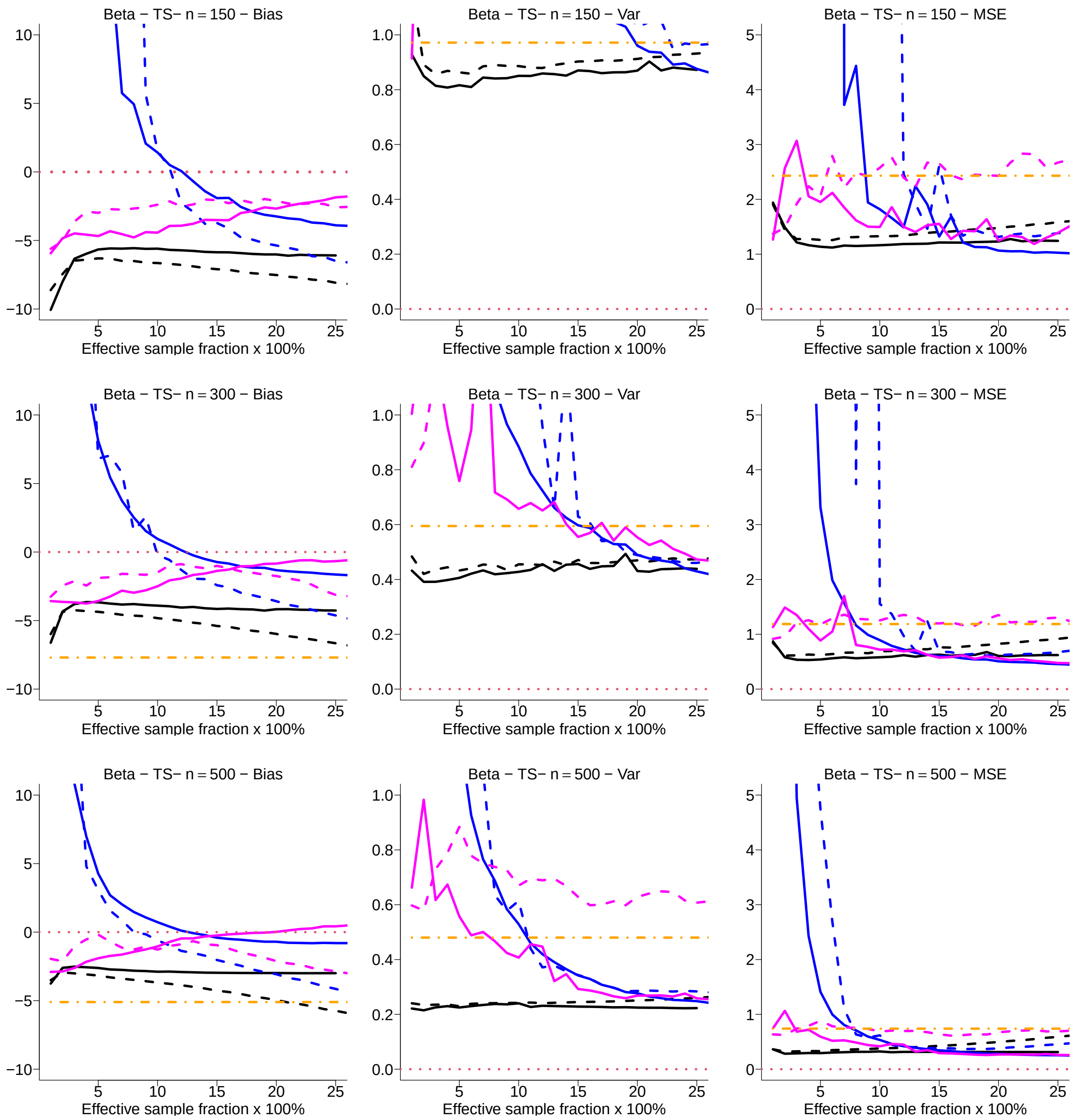

Figure B.4: Empirical relative bias, variance and MSE (left, middle and right), multiplied by 100, for the estimators of $\xi_{\tau_n'}$ obtained with nonlinear AR(1) observations from a Beta distribution (simulation setup (iv)), $\tau_n' = 1 - 1/n$ and sample size $n = 150, 300, 500$ (top, middle, bottom). Purely empirical estimator $\widehat{\xi}_{\tau_n'}$ (orange line), extrapolating LAWS estimators $\widehat{\xi}^{\star}_{\tau_n'}$ (magenta lines) and $\overline{\xi}^{\star}_{\tau_n'}$ (blue lines), and extrapolating QB estimators $\widetilde{\xi}^{\star}_{\tau_n'}$ (black lines). The versions of the extrapolating estimators based on the GPML scale and shape parameter estimates are referred to using solid lines, and those based on the Moment estimators are referred to using dashed lines.

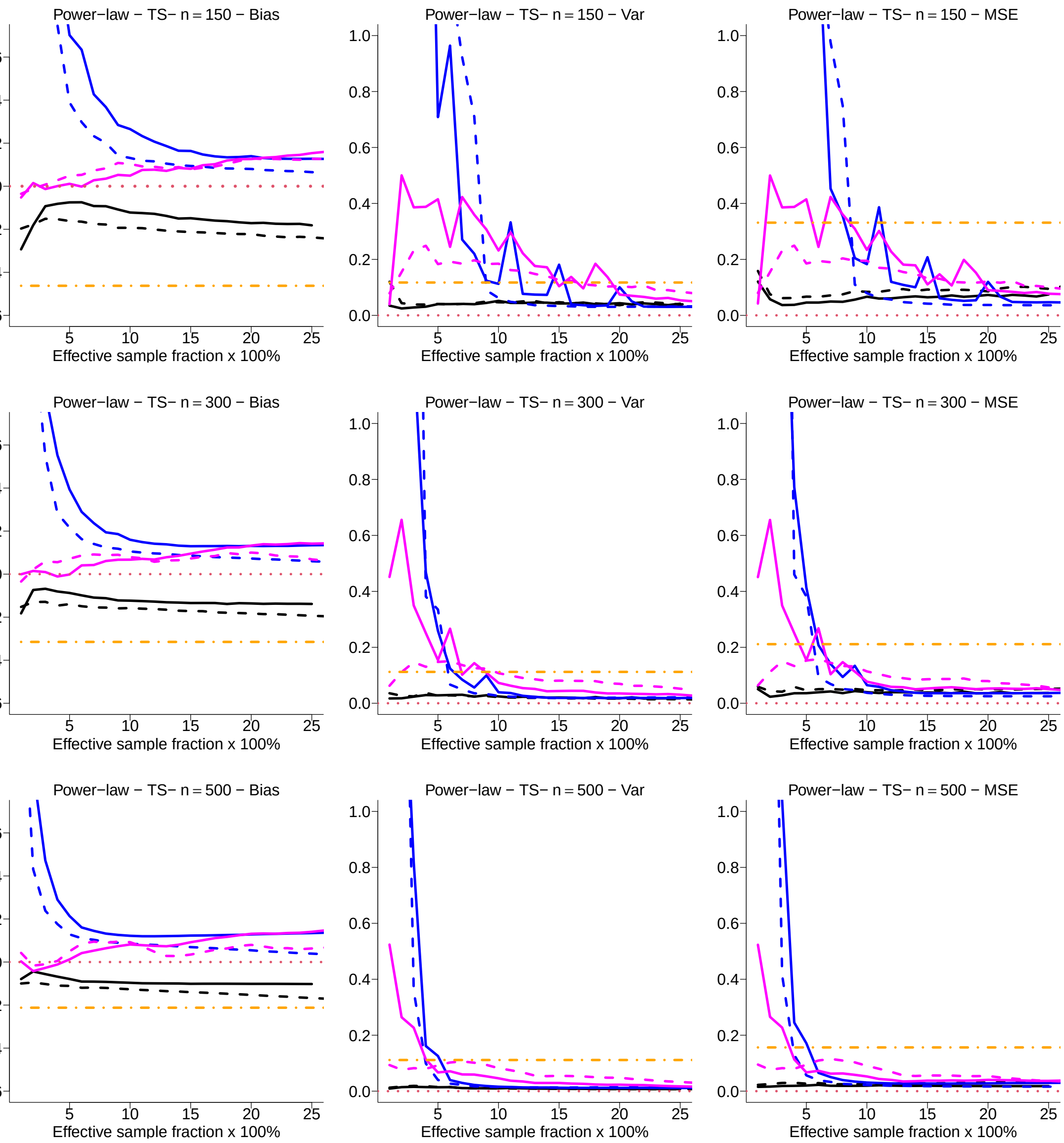

Figure B.5: Empirical relative bias, variance and MSE (left, middle and right), multiplied by 100, for the estimators of $\xi_{\tau_n'}$ obtained with nonlinear AR(1) observations from a power-law distribution (simulation setup (iv)), $\tau_n' = 1 - 1/n$ and sample size $n = 150, 300, 500$ (top, middle, bottom). Purely empirical estimator $\widehat{\xi}_{\tau_n'}$ (orange line), extrapolating LAWS estimators $\widehat{\xi}^{\star}_{\tau_n'}$ (magenta lines) and $\overline{\xi}^{\star}_{\tau_n'}$ (blue lines), and extrapolating QB estimators $\widetilde{\xi}^{\star}_{\tau_n'}$ (black lines). The versions of the extrapolating estimators based on the GPML scale and shape parameter estimates are referred to using solid lines, and those based on the Moment estimators are referred to using dashed lines.

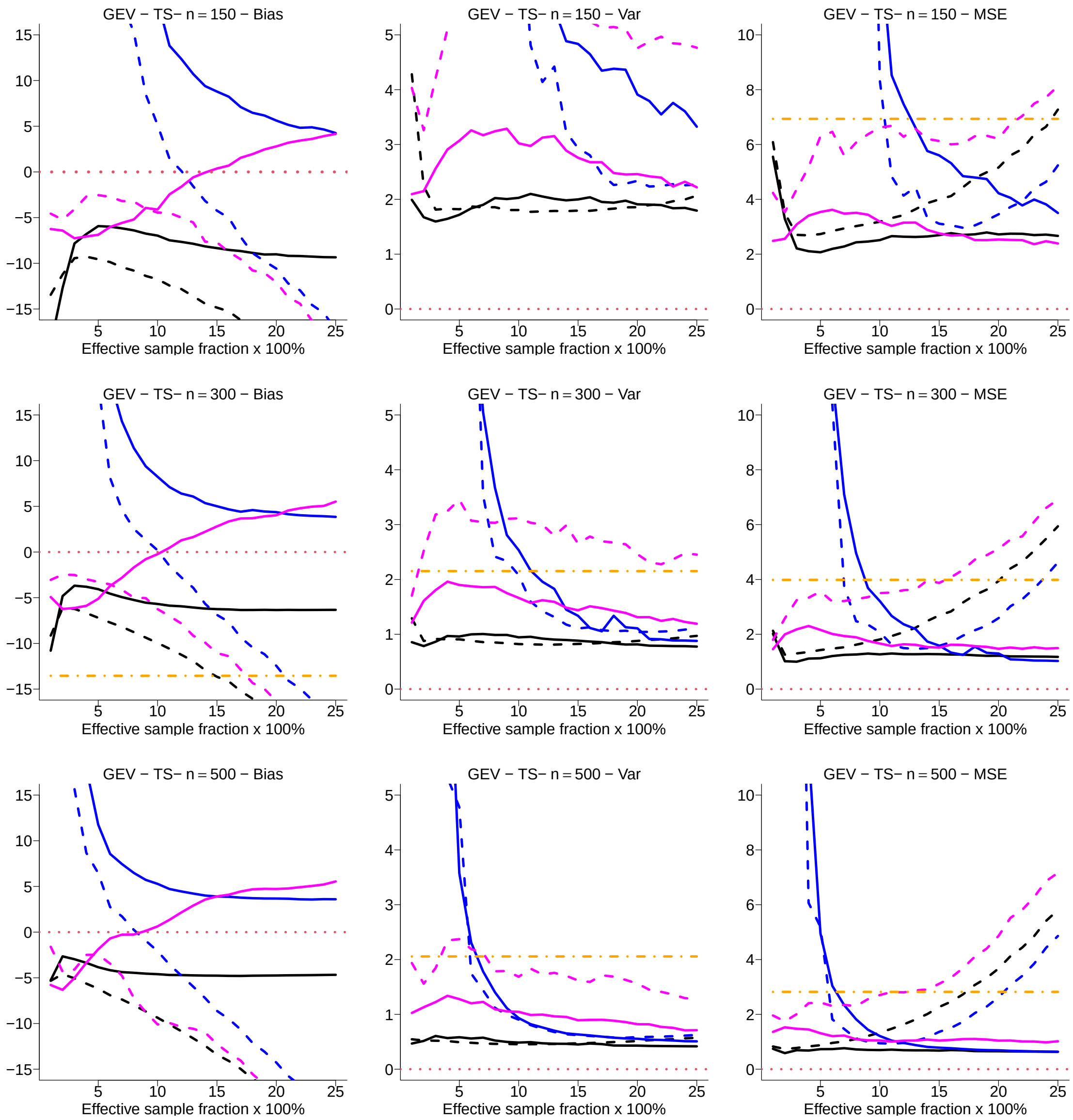

Figure B.6: Empirical relative bias, variance and MSE (left, middle and right), multiplied by 100, for the estimators of $\xi_{\tau_n'}$ obtained with nonlinear AR(1) observations from a GEV distribution (simulation setup (iv)), $\tau_n' = 1 - 1/n$ and sample size $n = 150, 300, 500$ (top, middle, bottom). Purely empirical estimator $\widehat{\xi}_{\tau_n'}$ (orange line), extrapolating LAWS estimators $\widehat{\xi}^{\star}_{\tau_n'}$ (magenta lines) and $\overline{\xi}^{\star}_{\tau_n'}$ (blue lines), and extrapolating QB estimators $\widetilde{\xi}^{\star}_{\tau_n'}$ (black lines). The versions of the extrapolating estimators based on the GPML scale and shape parameter estimates are referred to using solid lines, and those based on the Moment estimators are referred to using dashed lines.