



HAL
open science

Weighted approximations of the tail empirical expectile process

Abdelaati Daouia, Stéphane Girard, Gilles Stupfler

► **To cite this version:**

Abdelaati Daouia, Stéphane Girard, Gilles Stupfler. Weighted approximations of the tail empirical expectile process. 2018. hal-01744505v1

HAL Id: hal-01744505

<https://hal.science/hal-01744505v1>

Preprint submitted on 27 Mar 2018 (v1), last revised 28 Jun 2019 (v3)

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

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

WEIGHTED APPROXIMATIONS OF THE TAIL EMPIRICAL EXPECTILE PROCESS

BY ABDELAATI DAOUIA^{*}, STÉPHANE GIRARD[†] AND GILLES STUPFLER[‡]

Toulouse School of Economics, University of Toulouse Capitole^{}
Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK[†]
and School of Mathematical Sciences, University of Nottingham[‡]*

Expectiles define a least squares analogue of quantiles. They are determined by tail expectations rather than tail probabilities. For this reason and many other theoretical and practical merits, expectiles have recently received a lot of attention, especially in actuarial and financial risk management. Their estimation, however, typically requires to consider non-explicit asymmetric least squares estimates rather than the traditional order statistics used for quantile estimation. This makes the study of the tail expectile process a lot harder than that of the standard tail quantile process. Under the challenging model of heavy-tailed distributions, we derive joint weighted Gaussian approximations of the tail empirical expectile and quantile processes. We then use this powerful result to introduce and study new estimators of the tail index and extreme expectiles, as well as a novel expectile-based form of expected shortfall. Our estimators are built on general weighted combinations of both top order statistics and asymmetric least squares estimates. Some numerical simulations and an application to real data are provided.

1. Introduction. Least asymmetrically weighted squares estimation, borrowed from the econometrics literature, is one of the basic tools in statistical applications. This method often involves Newey and Powell's [32] concept of expectiles, a least squares analogue of traditional quantiles. Given an order $\tau \in (0, 1)$, Koenker and Bassett [27] elaborated an absolute error loss minimization to define the τ th quantile of the distribution of a random variable Y as the minimizer

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

with equality if the distribution function of Y is increasing, where $\rho_\tau(y) = |\tau - \mathbb{I}(y \leq 0)| |y|$ and $\mathbb{I}(\cdot)$ is the indicator function. This successfully extends the conventional definition of quantiles as left continuous inverse functions.

MSC 2010 subject classifications: Primary 62G20, 62G32; Secondary 62G05

Keywords and phrases: Asymmetric least squares, Coherent risk measures, Expected shortfall, Expectile, Extrapolation, Extremes, Heavy tails, Tail index

Newey and Powell [32] substituted the absolute deviations in the asymmetric loss function ρ_τ by squared deviations to obtain the τ th expectile of the distribution of Y as

$$(1) \quad \xi_\tau = \arg \min_{\theta \in \mathbb{R}} \mathbb{E} \{ \eta_\tau(Y - \theta) - \eta_\tau(Y) \},$$

with $\eta_\tau(y) = |\tau - \mathbb{I}(y \leq 0)| y^2$. The presence of the term $\eta_\tau(Y)$ ensures the existence of a unique solution ξ_τ for distributions with finite absolute first moments. Both quantiles and expectiles are M-quantiles as the minimizers of asymmetric convex loss functions (Breckling and Chambers [8]), but expectiles are determined by tail expectations rather than tail probabilities.

Accordingly, expectiles have been receiving a lot of attention in statistical finance and actuarial science since the pioneering papers of Taylor [38] and Kuan *et al.* [29]. They are excellent alternatives to quantiles in different aspects relevant to this kind of applications. First, expectiles depend on both the tail realizations and their probability, while quantiles only depend on the frequency of tail realizations and not on their values (Kuan *et al.* [29]). Expectiles, contrary to quantiles, thus allow to measure extreme risk based on the frequency of tail losses and their values. Second, more generally, altering the shape of the upper or lower tail of Y does not change the quantiles of the other tail, but it does impact all the expectiles (Taylor [38]). This high sensitivity of expectiles to tail behavior allows for more prudent and reactive risk management. Third, expectiles make more efficient use of the available data since they rely on the distance to observations, whereas quantiles only use the information on whether an observation is below or above the predictor (Sobotka and Kneib [37]). Fourth, inference on expectiles is much easier than inference on quantiles (Abdous and Remillard [1]). Using expectiles has the appeal of avoiding distributional assumptions (Taylor [38]) without recourse to regularity assumptions as can be seen by comparing, *e.g.*, Holzmann and Klar [25] with Zwingmann and Holzmann [42]. Most importantly, expectiles are the only M-quantiles that define a coherent risk measure in the sense of Artzner *et al.* [4] (see Bellini *et al.* [6]), and the only coherent risk measure that is elicitable (Ziegel [41]). Many other theoretical and numerical results motivate the adoption of expectiles in actuarial and financial risk management, including those of Ehm *et al.* [18] and Bellini and Di Bernardino [7].

Yet, tail expectile theory is, in comparison to tail quantile theory, relatively unexplored and still in full development. At the population level, only Bellini *et al.* [6], Mao *et al.* [30], Mao and Yang [31] and Bellini and Di Bernardino [7] have initiated the study of the connection between ξ_τ and q_τ , as $\tau \rightarrow 1$, when Y belongs to the domain of attraction of a Generalized

Extreme Value distribution. Also, for heavy-tailed distributions, Daouia *et al.* [11] have obtained an asymptotic expansion of ξ_τ/q_τ with a precise quantification of the bias term. At the sample level, attention has been mainly restricted to ordinary expectiles of fixed asymmetry level τ staying away from the distribution tails; see, *e.g.*, Holzmann and Klar [25] and Krättschmer and Zähle [28] for recent advanced theoretical developments. The extreme value analysis of asymmetric least squares estimators is a lot harder than for order statistics, mainly due to the absence of a closed form expression for expectiles. In an earlier paper, we partially solved this difficulty by proving the pointwise asymptotic normality of sample expectiles for ‘intermediate’ levels $\tau = \tau_n \rightarrow 1$ such that $n(1 - \tau_n) \rightarrow \infty$ as the sample size $n \rightarrow \infty$; see Theorem 2 of Daouia *et al.* [11]. Such a result does not, however, allow for simultaneous consideration of several intermediate sample expectiles. By contrast, Gaussian approximations of the tail empirical quantile process have been known for at least two decades; see among others, Drees [15] and Theorem 2.4.8 in de Haan and Ferreira [12]. These powerful asymptotic results, and their later generalizations, have been successfully used in the analysis of a number of complex statistical functionals, such as test statistics aimed at checking extreme value conditions (Dietrich *et al.* [14], Drees *et al.* [16], Hüsler and Li [26]), bias-corrected extreme value index estimators (de Haan *et al.* [13]) and estimators of extreme Wang distortion risk measures (El Methni and Stupfler [20, 21]).

The present paper fills this important gap in the current understanding of sample intermediate expectiles, under Pareto-type models that better describe the tail structure of most actuarial and financial data [see, *e.g.*, Embrechts *et al.* ([22], p.9) and Resnick ([33], p.1)]. In Section 2, we show that the aforementioned convergence result on single intermediate sample expectiles can vastly be generalized to the tail empirical expectile process. We first prove in Theorem 2 that the tail expectile process can be approximated by a sequence of Gaussian processes with drift and we derive its joint asymptotic behavior with the tail quantile process. Then, we analyze in Theorem 3 the difference between the tail empirical expectile process and its population counterpart. These two results constitute the major contribution of the paper; they open the door to the theoretical analysis of a wide range of functionals of the tail expectile process. Even more strongly, our joint weighted approximations of the tail empirical expectile and quantile processes make it possible to consider complex functionals of both processes.

We shall discuss below a number of applications of our main results. Section 3 applies the analysis of the tail expectile process in Theorem 2 to tail index estimation. We first construct purely expectile-based estimators of the

tail index and derive their asymptotic normality in Theorem 4. We then construct a more general class of estimators by computing a linear combination of these expectile-based estimators and of the Hill estimator (Hill [24]). This inspired the name *expectHill* estimators for this class. Thanks to the joint convergence result on the tail expectile and quantile processes in Theorem 2, we get the asymptotic normality of the *expectHill* estimators and derive their joint convergence with both intermediate quantile and expectile estimators in Theorem 5. Built on the *expectHill* estimators themselves, we propose in Section 4 general weighted estimators for intermediate expectiles ξ_{τ_n} whose asymptotic normality, obtained in Theorem 6, follows as a corollary of Theorem 5. Based on the ideas of Daouia *et al.* [11], the weighted intermediate expectile estimators are then extrapolated to the very extreme expectile levels that may approach one at an arbitrarily fast rate. The asymptotic properties of the extrapolated estimators are established in Theorem 7.

Theorem 3 is particularly important in tail risk estimation using Expected Shortfall (ES) measures. In Section 5, we show first that the expectile-based form XTCE_τ of ES introduced by Taylor [38] is not a coherent risk measure. Instead, we define a coherent alternative form that we call XES_τ . It is simply an average of tail expectiles, which is in addition asymptotically equivalent to the XTCE_τ . Asymptotic connections of XES_τ to other tail quantities, such as high quantiles q_τ and expectiles ξ_τ , are also provided before moving on to the extreme value estimation problem. XES_τ being an average of tail expectiles, it is readily estimated at an intermediate level $\tau = \tau_n$ by an average of the empirical tail expectile process, whose discrepancy with the true XES_{τ_n} can be unraveled thanks to Theorem 3. This intermediate estimator, like our generalized expectile estimators, can then be extrapolated to the very far tails of the distribution of Y where few or no data lie. Financial institutions and insurance companies are typically interested in the extreme region $\tau = \tau'_n \uparrow 1$ such that $n(1 - \tau'_n) \rightarrow c < \infty$, as $n \rightarrow \infty$ (see, for example, Cai *et al.* [9] and Daouia *et al.* [11]). In Theorems 9 and 10 we provide the asymptotic properties of the resulting extrapolated estimator, along with those of alternative plug-in estimators built on the asymptotic properties of XES_τ in Proposition 3. We conclude this section by using XES estimators as the basis for estimating the more traditional quantile-based ES (QES) itself. We derive three composite expectile-based estimators for QES at extreme levels whose asymptotic properties are established in Theorem 11.

Section 6 contains a simulation study of the estimators introduced hereafter. Applications to medical insurance data and financial returns data are presented, respectively, in Section 7 and the Supplementary Material document. The supplement also contains all proofs and auxiliary results.

2. Tail empirical expectile process. Suppose we observe independent copies $\{Y_1, \dots, Y_n\}$ of a random variable Y and denote by $Y_{1,n} \leq Y_{2,n} \leq \dots \leq Y_{n,n}$ their n th order statistics. A high expectile ξ_{τ_n} of order $\tau_n \rightarrow 1$, as $n \rightarrow \infty$, can be estimated by its empirical counterpart

$$(2) \quad \tilde{\xi}_{\tau_n} = \arg \min_{u \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n [\eta_{\tau_n}(Y_i - u) - \eta_{\tau_n}(Y_i)] = \arg \min_{u \in \mathbb{R}} \sum_{i=1}^n \eta_{\tau_n}(Y_i - u).$$

Here the expectile level τ_n approaches one at an ‘intermediate’ rate in the sense that $n(1 - \tau_n) \rightarrow \infty$ as $n \rightarrow \infty$. By analogy to the well-known *tail empirical quantile process* (see Definition 2.4.3 in de Haan and Ferreira [12])

$$(0, 1] \rightarrow \mathbb{R}, \quad s \mapsto \hat{q}_{1-ks/n} := Y_{n-[ks],n},$$

where $[\cdot]$ stands for the floor function and $k = k(n) \rightarrow \infty$ is a sequence of integers with $k/n \rightarrow 0$, we define the *tail empirical expectile process* to be the stochastic process

$$(0, 1] \rightarrow \mathbb{R}, \quad s \mapsto \tilde{\xi}_{1-(1-\tau_n)s}.$$

Note that the tail quantile process is nothing but $\{\hat{q}_{1-(1-\tau_n)s}\}_{0 < s \leq 1}$ with $\tau_n = 1 - k/n$. Our main objective in this section is to provide general asymptotic approximations of the tail expectile process by Gaussian processes, under the model assumption of heavy-tailed distributions. To this end, some preparatory remarks and work are necessary.

2.1. Statistical model and preliminary results. We focus on the maximum domain of attraction of Pareto-type distributions with tail index $0 < \gamma < 1$. The survival function of these heavy-tailed distributions can be expressed as

$$(3) \quad \bar{F}(y) := 1 - F(y) = y^{-1/\gamma} L(y),$$

for $y > 0$ large enough, where L is a slowly varying function at infinity, *i.e.*, a positive function on the positive half-line satisfying $L(ty)/L(t) \rightarrow 1$ as $t \rightarrow \infty$ for any $y > 0$. Equivalently, by Corollary 1.2.10 in de Haan and Ferreira [12], the tail quantile function of Y , defined as $U(t) := q_{1-t^{-1}} \equiv \inf\{y \in \mathbb{R} \mid 1/\bar{F}(y) \geq t\}$, satisfies

$$(4) \quad \lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma \quad \text{for all } x > 0.$$

The index γ tunes the tail heaviness of \bar{F} : the larger the index, the heavier the right tail. Let $Y_- = \min(Y, 0)$ denote the negative part of Y . Then,

together with condition $\mathbb{E}|Y_-| < \infty$, the assumption $\gamma < 1$ ensures that the first moment of Y exists, and hence expectiles of Y are well-defined. It has also been found under (3) or equivalently (4) that

$$(5) \quad \frac{\xi_\tau}{q_\tau} \sim (\gamma^{-1} - 1)^{-\gamma} \text{ as } \tau \rightarrow 1$$

(Bellini and Di Bernardino [7]). An asymptotic expansion of ξ_τ/q_τ with a precise quantification of the bias term is obtained in Corollary 1 of Daouia *et al.* [11] under the following standard second-order extreme value condition:

$\mathcal{C}_2(\gamma, \rho, A)$ For all $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{1}{A(t)} \left[\frac{U(tx)}{U(t)} - x^\gamma \right] = x^\gamma \frac{x^\rho - 1}{\rho}$$

where $\rho \leq 0$ is a constant parameter and A is a function converging to 0 at infinity and having ultimately constant sign. Hereafter, $(x^\rho - 1)/\rho$ is to be read as $\log x$ when $\rho = 0$. The meaning and the rationale behind this second-order extension of the regular variation condition (4) are extensively discussed in Beirlant *et al.* [5] and de Haan and Ferreira [12], along with abundant examples of commonly used continuous distributions satisfying $\mathcal{C}_2(\gamma, \rho, A)$. The asymptotic expansion in Daouia *et al.* [11] can actually be further strengthened to match our purposes, as follows.

PROPOSITION 1. *Assume that $\mathbb{E}|Y_-| < \infty$ and condition $\mathcal{C}_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1$.*

(i) *We have, as $\tau \rightarrow 1$,*

$$\begin{aligned} \frac{\xi_\tau}{q_\tau} &= (\gamma^{-1} - 1)^{-\gamma} \left(1 + \frac{\gamma(\gamma^{-1} - 1)^\gamma}{q_\tau} (\mathbb{E}(Y) + o(1)) \right. \\ &\quad \left. + \left(\frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \gamma - \rho} + \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho} + o(1) \right) A((1 - \tau)^{-1}) \right). \end{aligned}$$

(ii) *Let $\tau_n \rightarrow 1$ be such that $n(1 - \tau_n) \rightarrow \infty$, and pick $s \in (0, 1]$. Then*

$$\begin{aligned} \frac{\xi_{1-(1-\tau_n)s}}{\xi_{\tau_n}} &= s^{-\gamma} \left(1 + (s^\gamma - 1) \frac{\gamma(\gamma^{-1} - 1)^\gamma}{q_{\tau_n}} (\mathbb{E}(Y) + o(1)) \right. \\ &\quad \left. + \frac{(1 - \gamma)(\gamma^{-1} - 1)^{-\rho}}{1 - \gamma - \rho} \times \frac{s^{-\rho} - 1}{\rho} A((1 - \tau_n)^{-1})(1 + o(1)) \right). \end{aligned}$$

Part (i) of this proposition relaxes the conditions in Corollary 1 of Daouia *et al.* [11] by removing their unnecessary assumption of strict monotonicity of F . Part (ii) gives the asymptotic expansion of intermediate expectiles akin to condition $\mathcal{C}_2(\gamma, \rho, A)$ for intermediate quantiles, which also reads as

$$\frac{q_{1-(1-\tau_n)s}}{q_{\tau_n}} = s^{-\gamma} \left(1 + \frac{s^{-\rho} - 1}{\rho} A((1 - \tau_n)^{-1})(1 + o(1)) \right).$$

2.2. Main results. It is well-known that, under condition $\mathcal{C}_2(\gamma, \rho, A)$, the tail quantile process can be approximated by a sequence of scaled Brownian motions with drift. Namely, one can construct a sequence W_n of standard Brownian motions and a suitable measurable function A_0 such that

$$s^{\gamma+1/2+\varepsilon} \left| \sqrt{k} \left(\frac{\hat{q}_{1-ks/n}}{q_{1-k/n}} - s^{-\gamma} \right) - \gamma s^{-\gamma-1} W_n(s) - \sqrt{k} A_0(n/k) s^{-\gamma} \frac{s^{-\rho} - 1}{\rho} \right|$$

converges in probability to 0 uniformly in $s \in (0, 1]$ for any sufficiently small $\varepsilon > 0$ (see Theorem 2.4.8 in de Haan and Ferreira [12]). In addition to satisfying $k \rightarrow \infty$ and $k/n \rightarrow 0$, the sequence of integers $k = k(n)$ should also satisfy $\sqrt{k} A_0(n/k) = O(1)$. The proof of this approximation result reveals that it is subject to a potential enlargement of the underlying probability space and is valid for a suitable version of the tail quantile process, equal to the original one in distribution. This result being a convergence in probability, we will not explicitly make this distinction in the sequel; full details can be found in the Supplementary Material document. Besides, the function A_0 is actually asymptotically equivalent to A . We may therefore write:

$$(6) \quad \begin{aligned} \frac{\hat{q}_{1-(1-\tau_n)s}}{q_{\tau_n}} &= s^{-\gamma} \left(1 + \frac{1}{\sqrt{n(1-\tau_n)}} \gamma s^{-1} W_n(s) + \frac{s^{-\rho} - 1}{\rho} A((1 - \tau_n)^{-1}) \right. \\ &\quad \left. + o_{\mathbb{P}} \left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}} \right) \right) \quad \text{uniformly in } s \in (0, 1], \end{aligned}$$

where we set $k = n(1 - \tau_n)$, with $\tau_n \rightarrow 1$ and $n(1 - \tau_n) \rightarrow \infty$. As regards the tail expectile process, it is instructive to start by highlighting a simpler result of pointwise asymptotic normality of sample intermediate expectiles, already proved in Theorem 2 of Daouia *et al.* [11].

THEOREM 1 (Daouia *et al.*, 2018). *Let $\tau_n \rightarrow 1$ such that $n(1 - \tau_n) \rightarrow \infty$. Assume the first-order condition (4) holds with $\gamma \in (0, 1/2)$. Suppose further that $\mathbb{E}|Y_-|^{2+\delta} < \infty$ for some $\delta > 0$. Then*

$$\sqrt{n(1-\tau_n)} \begin{pmatrix} \tilde{\xi}_{\tau_n} \\ \xi_{\tau_n} \end{pmatrix} - 1 \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3}{1-2\gamma} \right) \quad \text{as } n \rightarrow \infty.$$

Similarly to the uniform approximation (6) of the tail quantile process, Theorem 1 can be vastly generalized to a uniform approximation of the tail expectile process $s \mapsto \tilde{\xi}_{1-(1-\tau_n)s}$. Already a heuristic combination of Proposition 1, Equation (6) and Theorem 1 suggests that a similar approximation to (6) might be derived for the tail expectile process under the same assumptions that $\gamma \in (0, 1/2)$ and $\mathbb{E}|Y_-|^{2+\delta} < \infty$. These conditions essentially guarantee that the loss variable has a finite variance. This is not likely to be restrictive in practice, since in most studies on actuarial and financial data, the realized values of γ have been found to lie well below $1/2$; see, *e.g.*, the R package ‘CASdatasets’, Cai *et al.* [9], Daouia *et al.* [11] and the references therein. It should also be clear that, for any $s \in (0, 1]$, $\xi_{1-(1-\tau_n)s}$ is an asymptotically normal estimator of $\xi_{1-(1-\tau_n)s}$ whose behavior is heavily influenced by the highest values in the sample (Y_1, \dots, Y_n) . In other words, the randomness present in the tail quantile process has a direct impact on the randomness featured in the tail expectile process. As such, one should expect the existence of some relationship between Gaussian approximations for these two processes. Our first main result formalizes these intuitions.

THEOREM 2. *Suppose that $\mathbb{E}|Y_-|^2 < \infty$. Assume further that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1/2$. Let $\tau_n \rightarrow 1$ be such that $n(1 - \tau_n) \rightarrow \infty$ and $\sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) = O(1)$. Then there exists a sequence W_n of standard Brownian motions such that, for any $\varepsilon > 0$ sufficiently small,*

$$\begin{aligned} \frac{\hat{q}_{1-(1-\tau_n)s}}{q_{\tau_n}} &= s^{-\gamma} \left(1 + \frac{1}{\sqrt{n(1-\tau_n)}} \gamma \sqrt{\gamma^{-1}-1} s^{-1} W_n \left(\frac{s}{\gamma^{-1}-1} \right) \right. \\ &\quad \left. + \frac{s^{-\rho}-1}{\rho} A((1-\tau_n)^{-1}) + o_{\mathbb{P}} \left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}} \right) \right) \\ \text{and } \frac{\tilde{\xi}_{1-(1-\tau_n)s}}{\xi_{\tau_n}} &= s^{-\gamma} \left(1 + (s^\gamma - 1) \frac{\gamma(\gamma^{-1}-1)^\gamma}{q_{\tau_n}} (\mathbb{E}(Y) + o_{\mathbb{P}}(1)) \right. \\ &\quad + \frac{1}{\sqrt{n(1-\tau_n)}} \gamma^2 \sqrt{\gamma^{-1}-1} s^{\gamma-1} \int_0^s W_n(t) t^{-\gamma-1} dt \\ &\quad + \frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} \times \frac{s^{-\rho}-1}{\rho} A((1-\tau_n)^{-1}) \\ &\quad \left. + o_{\mathbb{P}} \left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}} \right) \right) \quad \text{uniformly in } s \in (0, 1]. \end{aligned}$$

In the particular case $s = 1$, Theorem 2 entails

$$\sqrt{n(1-\tau_n)} \left(\frac{\tilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 \right) \xrightarrow{d} \gamma^2 \sqrt{\gamma^{-1} - 1} \int_0^1 W(t) t^{-\gamma-1} dt$$

where W denotes a standard Brownian motion. The right-hand side is a centered Gaussian random variable, whose variance is

$$\gamma^3(1-\gamma) \int_0^1 \int_0^1 \min(s,t)(st)^{-\gamma-1} ds dt = \frac{2\gamma^3}{1-2\gamma}.$$

We do therefore recover Theorem 1, subject to the additional condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = O(1)$, but under the reduced moment condition $\mathbb{E}|Y_-|^2 < \infty$. Note that the bias condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = O(1)$ is also required in order to establish the desired approximation (6) for the tail quantile process. Because the tail expectile process is, for small s , arbitrarily close to the tail quantile process (since $\tilde{\xi}_1 = Y_{n,n} = \hat{q}_1$, and $s \mapsto \tilde{\xi}_{1-(1-\tau_n)s}$ is a sample-wise continuous function), it is unlikely that the powerful and flexible approximation in Theorem 2 can be established without resorting to the same bias condition as in the quantile case. Besides, this bias condition appears naturally in the estimation of proper expectiles ξ_{τ_n} with extreme levels $\tau_n' \rightarrow 1$ such that $n(1-\tau_n') \rightarrow c \in [0, \infty)$ (see Daouia *et al.* [11]).

Note also that the asymptotic approximation of the tail expectile process in Theorem 2 contains all the error terms obtained in the asymptotic expansion of high population expectiles (see Proposition 1(ii)). These two approximations differ only by the presence of a Gaussian process and the uniform error weighting $s^{-1/2-\varepsilon}/\sqrt{n(1-\tau_n)}$. The Gaussian process appears due to the randomness in $\tilde{\xi}_{1-(1-\tau_n)s}$, while the term $s^{-1/2-\varepsilon}/\sqrt{n(1-\tau_n)}$ appears because, in contrast to Proposition 1(ii), Theorem 2 is a uniform result. In both approximations, the presence of bias terms proportional to $1/q_{\tau_n}$ and $A((1-\tau_n)^{-1})$ may, however, be inconvenient when the ultimate interest is in comparing directly the tail empirical expectile process with its population counterpart $s \mapsto \xi_{1-(1-\tau_n)s}$. This is particularly the case when developing the asymptotic theory for integrals of the tail expectile process, as will be seen below in Section 5. Our second main result is devoted to analyzing directly the gap between the tail empirical expectile process and its population counterpart. This result cannot be obtained as a direct corollary of Theorem 2, because Proposition 1(ii) is not a uniform result.

THEOREM 3. *If the conditions of Theorem 2 hold with $\rho < 0$, then there exists a sequence W_n of standard Brownian motions such that, for any $\varepsilon > 0$*

sufficiently small,

$$\begin{aligned} \frac{\tilde{\xi}_{1-(1-\tau_n)s}}{\xi_{1-(1-\tau_n)s}} &= 1 + \frac{1}{\sqrt{n(1-\tau_n)}} \gamma^2 \sqrt{\gamma^{-1}-1} s^{\gamma-1} \int_0^s W_n(t) t^{-\gamma-1} dt \\ &+ o_{\mathbb{P}} \left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}} \right) \quad \text{uniformly in } s \in (0, 1]. \end{aligned}$$

Compared to Theorem 2, the extra condition $\rho < 0$ is unlikely to be restrictive in practical applications, since most extrapolation results formulated in the extreme value literature under condition $\mathcal{C}_2(\gamma, \rho, A)$ actually assume that $\rho < 0$ (see, *e.g.*, Chapter 4 of de Haan and Ferreira [12] regarding extreme quantile estimation and Daouia *et al.* [11] for extreme expectile estimation). This condition is needed in the proof of Theorem 3 to get a uniform equivalent for $\xi_{1-(1-\tau_n)s}/\xi_{\tau_n}$, which translates, in virtue of (5), into obtaining a uniform equivalent for $q_{1-(1-\tau_n)s}/q_{\tau_n}$. The derivation of such an equivalent is then straightforward since, under $\rho < 0$, the quantile q_{τ} is asymptotically proportional to $(1-\tau)^{-\gamma}$, as $\tau \rightarrow 1$, according to de Haan and Ferreira ([12], p.49). Most importantly and in contrast to Theorem 2, Theorem 3 avoids the error terms in the approximation of Proposition 1(ii) that are proportional to $1/q_{\tau_n}$ and $A((1-\tau_n)^{-1})$. This comes as a consequence of examining directly the difference between $\xi_{1-(1-\tau_n)s}$ and $\tilde{\xi}_{1-(1-\tau_n)s}$, the former being an asymptotically unbiased estimator of the latter. Finally, note that the Gaussian term appearing in Theorem 3 is exactly the same as in the approximation of the tail expectile process in Theorem 2.

Theorems 2 and 3 open the door to the analysis of the asymptotic properties of a vast array of functionals of the tail expectile and quantile processes. We discuss in the next sections particular examples where these results can be used to construct general weighted estimators of the tail index and extreme expectiles, as well as of an expectile-based analogue for the Expected Shortfall risk measure. Theorems 2 and 3 will be the key tools when it comes to unravel the asymptotic behavior of these estimators.

3. Estimation of the tail index. In this section, we first construct purely expectile-based estimators of the tail index γ and derive their asymptotic distributions. We shall then construct a more general class of estimators by combining both intermediate empirical expectiles and quantiles. The basic idea stems from Proposition 1(ii) which suggests the following approximation:

$$\int_0^1 \log \left(\frac{\xi_{1-(1-\tau_n)s}}{\xi_{\tau_n}} \right) ds \approx \int_0^1 \log(s^{-\gamma}) ds = \gamma$$

where $\tau_n \rightarrow 1$ is such that $n(1 - \tau_n) \rightarrow \infty$. One can then estimate γ by

$$\check{\gamma}_{\tau_n} := \int_0^1 \log \left(\frac{\tilde{\xi}_{1-(1-\tau_n)s}}{\tilde{\xi}_{\tau_n}} \right) ds.$$

A computationally more viable option is to use a discretized version of the integral estimator $\check{\gamma}_{\tau_n}$ on a regular l -grid of points in $[0, 1]$, namely:

$$\check{\gamma}_{\tau_n, l} := \frac{1}{l} \sum_{i=1}^l \log \left(\frac{\tilde{\xi}_{1-(1-\tau_n)(i-1)/l}}{\tilde{\xi}_{\tau_n}} \right)$$

where $l = l(n) \rightarrow \infty$. A particularly interesting example is

$$(7) \quad \tilde{\gamma}_{\tau_n} := \frac{1}{[n(1 - \tau_n)]} \sum_{i=1}^{\lfloor n(1-\tau_n) \rfloor} \log \left(\frac{\tilde{\xi}_{1-(i-1)/n}}{\tilde{\xi}_{1-\lfloor n(1-\tau_n) \rfloor/n}} \right)$$

or, equivalently, $\tilde{\gamma}_{\tau_n} = \tilde{\gamma}_{1-\lfloor n(1-\tau_n) \rfloor/n, \lfloor n(1-\tau_n) \rfloor}$. This simple estimator has exactly the same form as the popular Hill estimator (Hill [24])

$$(8) \quad \hat{\gamma}_{\tau_n} = \frac{1}{[n(1 - \tau_n)]} \sum_{i=1}^{\lfloor n(1-\tau_n) \rfloor} \log \left(\frac{\hat{q}_{1-(i-1)/n}}{\hat{q}_{1-\lfloor n(1-\tau_n) \rfloor/n}} \right)$$

with the tail empirical quantile process \hat{q} in (8) replaced by its asymmetric least squares analogue $\tilde{\xi}$. Beirlant *et al.* [5] and de Haan and Ferreira [12] give an extensive overview of the asymptotic theory for the Hill estimator $\hat{\gamma}_{\tau_n}$. The next theorem gives the asymptotic normality of the three new estimators $\check{\gamma}_{\tau_n}$, $\check{\gamma}_{\tau_n, l}$ and $\tilde{\gamma}_{\tau_n}$. Its proof essentially consists in writing

$$\log \left(\frac{\tilde{\xi}_{1-(1-\tau_n)s}}{\tilde{\xi}_{\tau_n}} \right) = \log \left(\frac{\tilde{\xi}_{1-(1-\tau_n)s}}{\xi_{\tau_n}} \right) - \log \left(\frac{\tilde{\xi}_{\tau_n}}{\xi_{\tau_n}} \right)$$

before integrating and crucially using Theorem 2 twice in order to control both of the logarithms on the right-hand side.

THEOREM 4. *Suppose that $\mathbb{E}|Y_-|^2 < \infty$. Assume further that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1/2$. Let $\tau_n \rightarrow 1$ be such that $n(1 - \tau_n) \rightarrow \infty$, and suppose that the bias conditions $\sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) \rightarrow \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1 - \tau_n)}/q_{\tau_n} \rightarrow \lambda_2 \in \mathbb{R}$ are satisfied. Then:*

$$(i) \quad \sqrt{n(1 - \tau_n)}(\check{\gamma}_{\tau_n} - \gamma)$$

$$\xrightarrow{d} \mathcal{N} \left(\frac{(1 - \gamma)(\gamma^{-1} - 1)^{-\rho}}{(1 - \rho)(1 - \gamma - \rho)} \lambda_1 - \mathbb{E}(Y) \frac{\gamma^2(\gamma^{-1} - 1)^\gamma}{\gamma + 1} \lambda_2, \frac{2\gamma^3}{1 - 2\gamma} \right).$$

(ii) If $l = l(n)$ fulfills $\sqrt{n(1 - \tau_n)} \log(n(1 - \tau_n))/l \rightarrow 0$, then (i) holds with $\check{\gamma}_{\tau_n}$ replaced by $\check{\gamma}_{\tau_n, l}$. Especially, (i) holds with $\check{\gamma}_{\tau_n}$ replaced by $\check{\gamma}_{\tau_n}$.

Before using the estimator $\check{\gamma}_{\tau_n}$ to construct a more general class of tail-index estimators, we formulate a couple of remarks about its theoretical and practical behavior.

REMARK 1. The conditions involving the auxiliary function A in Theorem 4 are also required to derive the asymptotic normality of the conventional Hill estimator $\hat{\gamma}_{\tau_n}$ in (8), with asymptotic bias $\lambda_2/(1 - \rho)$ and asymptotic variance γ^2 [see Theorem 3.2.5 in de Haan and Ferreira ([12], p.74)]. Theorem 4 also features a further bias condition involving the quantile function q ; this was to be expected in view of Proposition 1(ii), of which a consequence is that the remainder term in the approximation $\xi_{1-(1-\tau_n)s}/\xi_{\tau_n} \approx s^{-\gamma}$ depends on both A and q . Yet, it is straightforward to eliminate this bias component: note that the centered variable $Z = Y - \mathbb{E}(Y)$ is also heavy-tailed, with the same extreme value parameters as Y , and thus the estimator $\check{\gamma}_{\tau_n}^Z$ constructed on the $Z_i = Y_i - \mathbb{E}(Y)$ satisfies

$$\sqrt{n(1 - \tau_n)}(\check{\gamma}_{\tau_n}^Z - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{(1 - \gamma)(\gamma^{-1} - 1)^{-\rho}}{(1 - \rho)(1 - \gamma - \rho)}\lambda_1, \frac{2\gamma^3}{1 - 2\gamma}\right).$$

This suggests to define $\hat{Z}_i = Y_i - \bar{Y}_n$, where \bar{Y}_n is the sample mean, and then to consider the estimator $\check{\gamma}_{\tau_n}^{\hat{Z}}$. Due to the translation equivariance of expectiles, the gap between $\check{\gamma}_{\tau_n}^{\hat{Z}}$ and $\check{\gamma}_{\tau_n}^Z$ has the same order as $|\bar{Y}_n - \mathbb{E}(Y)| = O_{\mathbb{P}}(1/\sqrt{n})$. It follows that $\check{\gamma}_{\tau_n}^{\hat{Z}}$ has the same asymptotic distribution as $\check{\gamma}_{\tau_n}^Z$, and is therefore a bias-reduced version of $\check{\gamma}_{\tau_n}$ which eliminates the quantile component of the bias.

REMARK 2. The selection of τ_n is a difficult problem in general, since any sort of optimal choice will involve the unknown parameter ρ as well as the function A ; for a discussion about the optimal choice of τ_n in the Hill estimator based on mean-squared error, see Hall and Welsh [23]. A usual practice for selecting a reasonable estimate $\hat{\gamma}_{\tau_n}$ is, in the reparametrization $\tau_n = 1 - k/n$, to plot the graph of $k \mapsto \hat{\gamma}_{1-k/n}$ for $k \in \{1, 2, \dots, n - 1\}$, and then to pick out a value of k corresponding to the first stable part of the plot [see, e.g., de Haan and Ferreira ([12], Section 3)]. There have been a number of attempts at formalizing this procedure, including Resnick and Stărică [34], Drees *et al.* [17], and more recently El Methni and Stupfler [20, 21]. The Hill plot may be, however, so unstable that reasonable values of k (which would correspond to estimates close to the true value of γ) may be hidden in the

graph. The least squares analogue $\tilde{\gamma}_{1-k/n}$ in (7) is, in contrast to $\hat{\gamma}_{1-k/n}$, based on expectiles that enjoy superior regularity properties compared to quantiles (see Proposition 1 in Holzmam and Klar [25]). One may thus expect that $\tilde{\gamma}_{1-k/n}$ affords smoother and more stable plots compared to those of the Hill estimator $\hat{\gamma}_{1-k/n}$. This advantage is illustrated in Section A of the Supplementary Material document, where we examine the behavior of $\hat{\gamma}$ and $\tilde{\gamma}$ on two concrete actuarial and financial data sets. It can be seen thereon that the plots of $k \mapsto \tilde{\gamma}_{1-k/n}$ are indeed far smoother than the arguably wiggly plots of $k \mapsto \hat{\gamma}_{1-k/n}$.

It could, however, happen that $\tilde{\gamma}$ has a higher bias than the Hill estimator. This is for instance the case if $|\rho|$ is large, since a large $|\rho|$ means that the underlying distribution is, in its right tail, very close to a multiple of the Pareto distribution for which the Hill estimator is unbiased. An efficient way to take advantage of the desirable properties of both $\tilde{\gamma}$ and $\hat{\gamma}$ in a large class of models is by using their linear combination for estimating γ . For $\alpha \in \mathbb{R}$, we then define the more general estimator

$$(9) \quad \bar{\gamma}_{\tau_n}(\alpha) := \alpha \hat{\gamma}_{\tau_n} + (1 - \alpha) \tilde{\gamma}_{\tau_n}.$$

We shall call this linear combination the *expectHill estimator*. For example, the simple mean $\bar{\gamma}_{\tau_n}(1/2)$ would represent an equal balance between the use of large asymmetric least squares statistics in (7) and top order statistics in (8). The convergence of the *expectHill* estimator is, however, a highly non-trivial problem as it hinges, by construction, on both the tail expectile and quantile processes. The explicit joint asymptotic Gaussian representation of these two processes, obtained in Theorem 2, is a pivotal tool for our analysis, and enables us to address the convergence problem in its full generality. We establish below the asymptotic normality of the *expectHill* estimator, along with its joint convergence with intermediate sample quantiles and expectiles.

THEOREM 5. *Suppose that the conditions of Theorem 4 hold. Then, for any $\alpha \in \mathbb{R}$,*

$$\sqrt{n(1 - \tau_n)} \left(\bar{\gamma}_{\tau_n}(\alpha) - \gamma, \frac{\hat{q}_{\tau_n}}{q_{\tau_n}} - 1, \frac{\tilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(\mathbf{m}_\alpha, \mathfrak{V}_\alpha)$$

where \mathbf{m}_α is the 1×3 vector $\mathbf{m}_\alpha := (b_\alpha, 0, 0)$, with

$$(10) \quad \begin{aligned} b_\alpha &= \frac{\lambda_1}{1 - \rho} \left(\alpha + (1 - \alpha) \frac{(1 - \gamma)(\gamma^{-1} - 1)^{-\rho}}{1 - \gamma - \rho} \right) \\ &\quad - (1 - \alpha) \mathbb{E}(Y) \frac{\gamma^2(\gamma^{-1} - 1)^\gamma}{\gamma + 1} \lambda_2, \end{aligned}$$

and \mathfrak{V}_α is the 3×3 symmetric matrix with entries

$$\begin{aligned}\mathfrak{V}_\alpha(1,1) &= \gamma^2 \left(\alpha^2 \left[\frac{3-4\gamma}{1-2\gamma} - 2 \frac{(\gamma^{-1}-1)^\gamma}{1-\gamma} \right] - 2\alpha \left[\frac{1}{1-2\gamma} - \frac{(\gamma^{-1}-1)^\gamma}{1-\gamma} \right] \right. \\ &\quad \left. + \frac{2\gamma}{1-2\gamma} \right), \\ \mathfrak{V}_\alpha(1,2) &= (1-\alpha)\gamma[(\gamma^{-1}-1)^\gamma - 1 - \gamma \log(\gamma^{-1}-1)], \\ \mathfrak{V}_\alpha(1,3) &= \frac{\gamma^3}{(1-\gamma)^2} \left[\alpha(\gamma^{-1}-1)^\gamma + (1-\alpha) \frac{1-\gamma}{1-2\gamma} \right], \\ \mathfrak{V}_\alpha(2,2) &= \gamma^2, \quad \mathfrak{V}_\alpha(2,3) = \gamma^2 \left(\frac{(\gamma^{-1}-1)^\gamma}{1-\gamma} - 1 \right), \quad \mathfrak{V}_\alpha(3,3) = \frac{2\gamma^3}{1-2\gamma}.\end{aligned}$$

As an immediate consequence, we have for any $\alpha \in \mathbb{R}$,

$$(11) \quad \sqrt{n(1-\tau_n)} (\bar{\gamma}_{\tau_n}(\alpha) - \gamma) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha) \quad \text{where } v_\alpha = \mathfrak{V}_\alpha(1,1).$$

This remains valid if $\tilde{\gamma}_{\tau_n}$ is replaced in (9) by the continuous version $\check{\gamma}_{\tau_n}$, or any other discretized version $\tilde{\gamma}_{\tau_n,l}$ provided $\sqrt{n(1-\tau_n)} \log(n(1-\tau_n))/l \rightarrow 0$.

REMARK 3. The optimal value of the weighting coefficient α in (9), which minimizes the asymptotic variance v_α of $\bar{\gamma}_{\tau_n}(\alpha)$, only depends on the tail index γ and has the explicit expression

$$\alpha(\gamma) = \frac{(1-\gamma) - (1-2\gamma)(\gamma^{-1}-1)^\gamma}{(1-\gamma)(3-4\gamma) - 2(1-2\gamma)(\gamma^{-1}-1)^\gamma}.$$

Its plot against $\gamma \in (0, 1/2)$ is given in Section B of the Supplementary Material document. It can be seen thereon that the simple mean $\bar{\gamma}_{\tau_n}(1/2)$ of $\hat{\gamma}_{\tau_n}$ and $\tilde{\gamma}_{\tau_n}$, with $\alpha = 1/2$, affords a middle course between $\hat{\gamma}_{\tau_n} \equiv \bar{\gamma}_{\tau_n}(1)$ and $\tilde{\gamma}_{\tau_n} \equiv \bar{\gamma}_{\tau_n}(0)$ in terms of asymptotic variance. In terms of smoothness, $\bar{\gamma}_{\tau_n}(1/2)$ offers a middle course as well, as shown in Section A of the Supplementary Material document, where the plot of $\bar{\gamma}_{\tau_n}(1/2)$ is superimposed in green line with the plots of $\hat{\gamma}_{\tau_n}$ and $\tilde{\gamma}_{\tau_n}$.

4. Extreme expectile estimation. In this section, we first return to intermediate expectile estimation by making use of the general class of γ estimators $\{\bar{\gamma}_{\tau_n}(\alpha)\}_{\alpha \in \mathbb{R}}$ to construct alternative estimators for high expectiles ξ_{τ_n} such that $\tau_n \rightarrow 1$ and $n(1-\tau_n) \rightarrow \infty$ as $n \rightarrow \infty$. Then we extrapolate the obtained estimators to the very high expectile levels that may approach one at an arbitrarily fast rate.

Alternatively to the direct nonparametric estimator $\tilde{\xi}_{\tau_n}$ defined in (2), one may use the asymptotic connection $\xi_{\tau_n} \sim (\gamma^{-1} - 1)^{-\gamma} q_{\tau_n}$ between ξ_{τ_n} , γ and the intermediate quantile q_{τ_n} , described in (5), to define the following indirect semiparametric estimator of ξ_{τ_n} :

$$\hat{\xi}_{\tau_n}(\alpha) := (\bar{\gamma}_{\tau_n}(\alpha)^{-1} - 1)^{-\bar{\gamma}_{\tau_n}(\alpha)} \hat{q}_{\tau_n}.$$

More generally, one may also combine the two estimators $\hat{\xi}_{\tau_n}(\alpha)$ and $\tilde{\xi}_{\tau_n}$ to define, for $\beta \in \mathbb{R}$, the weighted estimator

$$\bar{\xi}_{\tau_n}(\alpha, \beta) := \beta \hat{\xi}_{\tau_n}(\alpha) + (1 - \beta) \tilde{\xi}_{\tau_n}.$$

The two special cases $\alpha = \beta = 1$ and $\beta = 0$ correspond to the unique existing intermediate expectile estimators in the literature, namely, the estimators $\hat{\xi}_{\tau_n}$ and $\tilde{\xi}_{\tau_n}$ in Daouia *et al.* [11]. These were coined, respectively, “indirect estimator” and “direct estimator” to reflect the asymmetric least squares nature of the latter and the reliance of the former on quantiles. The next result provides the limit distribution of $\bar{\xi}_{\tau_n}(\alpha, \beta)$ for an intermediate level τ_n . Its proof crucially relies on the asymptotic dependence structure between the tail expectile and quantile processes established in Theorem 2, since $\bar{\xi}_{\tau_n}(\alpha, \beta)$ is built on both of these processes. More specifically, we use the asymptotic dependence structure between $\bar{\gamma}_{\tau_n}(\alpha)$, \hat{q}_{τ_n} and $\tilde{\xi}_{\tau_n}$ proved in Theorem 5.

THEOREM 6. *Suppose that the conditions of Theorem 4 hold. Then, for any $\alpha, \beta \in \mathbb{R}$,*

$$\begin{aligned} & \sqrt{n(1 - \tau_n)} \left(\frac{\bar{\xi}_{\tau_n}(\alpha, \beta)}{\xi_{\tau_n}} - 1 \right) \\ & \xrightarrow{d} \beta (\mathbf{b}_\alpha + [(1 - \gamma)^{-1} - \log(\gamma^{-1} - 1)]\Psi_\alpha + \Theta) + (1 - \beta)\Xi \end{aligned}$$

where the bias component \mathbf{b}_α is $\mathbf{b}_\alpha = \lambda_1 \mathbf{b}_{1,\alpha} + \lambda_2 \mathbf{b}_{2,\alpha}$ with

$$\begin{aligned} \mathbf{b}_{1,\alpha} &= \frac{(1 - \gamma)^{-1} - \log(\gamma^{-1} - 1)}{1 - \rho} \left[\alpha + (1 - \alpha) \frac{(1 - \gamma)(\gamma^{-1} - 1)^{-\rho}}{1 - \gamma - \rho} \right] \\ & \quad - \frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \gamma - \rho} - \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho}, \\ \mathbf{b}_{2,\alpha} &= -\gamma(\gamma^{-1} - 1)^\gamma \mathbb{E}(Y) \left(1 + (1 - \alpha)[(1 - \gamma)^{-1} - \log(\gamma^{-1} - 1)] \frac{\gamma}{\gamma + 1} \right), \end{aligned}$$

and $(\Psi_\alpha, \Theta, \Xi)$ is a trivariate Gaussian centered random vector with covariance matrix \mathfrak{V}_α as in Theorem 5.

REMARK 4. When $\alpha = \beta = 1$, we recover the convergence of the “indirect estimator” $\bar{\xi}_{\tau_n}(1, 1)$ obtained in Corollary 2 of Daouia *et al.* [11]. When $\beta = 0$, we recover the convergence of the “direct estimator” $\tilde{\xi}_{\tau_n}$ stated in Theorem 2 of [11].

The use of the weighted estimator $\bar{\xi}_{\tau_n}(\alpha, \beta)$ is, by construction, most appropriate when it comes to deal with intermediate expectile levels $\tau = \tau_n \rightarrow 1$ such that $n(1 - \tau_n) \rightarrow \infty$. In the very far tails where the expectile level $\tau = \tau'_n \rightarrow 1$ is such that $n(1 - \tau'_n) \rightarrow c \in [0, \infty)$, this estimator becomes unstable and inconsistent due to data sparsity. To estimate an extreme expectile $\xi_{\tau'_n}$, Daouia *et al.* [11] propose to extrapolate any consistent intermediate expectile estimator, say $\check{\xi}_{\tau_n}$, to the very high level τ'_n by considering the generic class of estimators

$$(12) \quad \check{\xi}_{\tau'_n}^* := \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\check{\gamma}_n} \check{\xi}_{\tau_n},$$

where $\check{\gamma}_n$ is a suitable estimator of γ . The rationale behind the formulation (12) is to first use the regular variation condition (3), or equivalently (4), that entails the following classical extrapolation formula for high quantiles:

$$\frac{q_{\tau'_n}}{q_{\tau_n}} = \frac{U((1 - \tau'_n)^{-1})}{U((1 - \tau_n)^{-1})} \approx \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\gamma}$$

as τ_n and τ'_n approach one (Weissman [39]). Then, by applying the asymptotic connection (5) between quantiles and expectiles, we get $\xi_{\tau'_n}/\xi_{\tau_n} \sim q_{\tau'_n}/q_{\tau_n}$, and hence

$$(13) \quad \xi_{\tau'_n} \approx \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\gamma} \xi_{\tau_n}.$$

This motivates the extrapolated Weissman-type estimator (12) obtained by replacing in the approximation (13) the intermediate expectile ξ_{τ_n} and the tail index γ with consistent estimators $\check{\xi}_{\tau_n}$ and $\check{\gamma}_n$, respectively. Here, we choose to use the general *expectHill* estimator $\check{\gamma}_n := \bar{\gamma}_{\tau_n}(\alpha)$ and the weighted intermediate estimator $\bar{\xi}_{\tau_n}(\alpha, \beta)$ to define the following class of extreme expectile estimators:

$$(14) \quad \bar{\xi}_{\tau'_n}^*(\alpha, \beta) := \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\bar{\gamma}_{\tau_n}(\alpha)} \bar{\xi}_{\tau_n}(\alpha, \beta).$$

The two special cases $\alpha = \beta = 1$ and $\alpha = 1 - \beta = 1$ correspond to the unique existing extreme expectile estimators in the literature, namely, the

extrapolated indirect and direct expectile estimators suggested in Daouia *et al.* [11]. The next theorem gives the asymptotic behavior of the generalized extreme expectile estimators $\bar{\xi}_{\tau'_n}^*(\alpha, \beta)$.

THEOREM 7. *Suppose that the conditions of Theorem 4 hold. Assume also that $\rho < 0$ and $n(1 - \tau'_n) \rightarrow c < \infty$ with $\sqrt{n(1 - \tau_n)}/\log[(1 - \tau_n)/(1 - \tau'_n)] \rightarrow \infty$. Then, for any $\alpha, \beta \in \mathbb{R}$,*

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left(\frac{\bar{\xi}_{\tau'_n}^*(\alpha, \beta)}{\xi_{\tau'_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha)$$

with (b_α, v_α) as in (10) and (11).

One can observe that the limiting distribution of $\bar{\xi}_{\tau'_n}^*(\alpha, \beta)$ is controlled by the asymptotic distribution of $\bar{\gamma}_{\tau_n}(\alpha)$. In particular, in the cases $\alpha = \beta = 1$ and $\alpha = 1 - \beta = 1$, we exactly recover Corollaries 3 and 4 of Daouia *et al.* [11] on the convergence of the extrapolated indirect and direct expectile estimators. This is a consequence of the fact that the convergence of $\bar{\xi}_{\tau'_n}^*(\alpha, \beta)$ is governed by that of the extrapolation factor $[(1 - \tau'_n)/(1 - \tau_n)]^{-\bar{\gamma}_{\tau_n}(\alpha)}$. The latter approximates the theoretical factor $[(1 - \tau'_n)/(1 - \tau_n)]^{-\gamma}$ in the extrapolation (13) at a slower rate than both the speed of convergence of $\bar{\xi}_{\tau_n}(\alpha, \beta)$ to ξ_{τ_n} , given by Theorem 6, and the speed of convergence to 0 of the bias term that is incurred by the use of (13) and that can be controlled by Proposition 1(ii).

5. Estimation of tail Expected Shortfall.

5.1. Background. The risk of a financial position Y is usually summarized by a risk measure $\varrho(Y)$, where ϱ is a mapping from a space of random variables to the real line. Value at Risk (VaR) is arguably the most common risk measure used in practice. It is given at probability level $\tau \in (0, 1)$ by the τ -quantile $\text{VaR}_\tau(Y) := q_\tau$. Hereafter, we adopt the convention that Y is a real-valued random variable whose values are the negative of financial returns. The right-tail of the distribution of Y , for levels τ close to one, then corresponds to the negative of extreme losses.

One of the main criticisms of VaR_τ is that it does not account for the size of losses beyond the level τ , since it only depends on the frequency of tail losses and not on their values (Dánielsson *et al.* [10]). Furthermore, VaR_τ fails to be subadditive, since the inequality $\text{VaR}_\tau(Y_1 + Y_2) \leq \text{VaR}_\tau(Y_1) + \text{VaR}_\tau(Y_2)$ does not hold in general (Acerbi [2]). It is therefore not a coherent

risk measure in the sense of Artzner *et al.* [4], which is problematic in risk management.

An important alternative to VaR_τ is Expected Shortfall at level τ . This risk measure is defined as (Acerbi [2])

$$(15) \quad \text{QES}_\tau := \frac{1}{1-\tau} \int_\tau^1 q_t dt.$$

When Y is continuous, QES_τ is identical to the Conditional Value at Risk (Rockafellar and Uryasev [35, 36]), known also as Tail Conditional Expectation (TCE), defined as $\text{QTCE}_\tau := \mathbb{E}[Y|Y > q_\tau]$. Both QES_τ and QTCE_τ can then be interpreted as the average loss incurred in the event of a loss higher than VaR_τ . We note, however, that QES_τ defines a coherent risk measure but QTCE_τ does not in general (see Wirch and Hardy [40] and Acerbi and Tasche [3]).

5.2. *Expectile-based Expected Shortfall.* Motivated by the merits and good properties of expectiles, Taylor [38] has introduced an expectile-based form of Expected Shortfall (ES) as the expectation $\text{XTCE}_\tau := \mathbb{E}[Y|Y > \xi_\tau]$ of exceedances beyond the τ th expectile ξ_τ of the distribution of Y . The interpretability of this risk measure is therefore straightforward, but its coherence has been an open problem so far. This problem is now elucidated below in Proposition 2, showing the failure of XTCE_τ to fulfill the coherence property in general. Alternatively, by analogy to the coherent quantile-based version QES_τ , we propose to use the new expectile-based form of ES

$$(16) \quad \text{XES}_\tau := \frac{1}{1-\tau} \int_\tau^1 \xi_t dt,$$

obtained by substituting the expectile ξ_t in place of the quantile q_t in the standard form (15) of ES. It turns out that, in contrast to XTCE_τ , the new risk measure XES_τ is coherent in general.

PROPOSITION 2. *For all $\tau \geq 1/2$,*

- (i) *XES_τ induces a coherent risk measure;*
- (ii) *XTCE_τ is neither monotonic nor subadditive in general, and hence does not induce a coherent risk measure.*

The coherence property of XES_τ , contrary to that of QES_τ , is actually a straightforward consequence of the coherence of the expectile-based risk measure ξ_τ , for $\tau \geq 1/2$.

Next, we show under the model assumption (3) that XES_τ is asymptotically equivalent to $XTCE_\tau$ as $\tau \rightarrow 1$, and hence inherits its direct meaning as a conditional expectation for all τ large enough.

PROPOSITION 3. *Assume that $\mathbb{E}|Y_-| < \infty$ and that Y has a Pareto-type distribution (3) with tail index $0 < \gamma < 1$. Then*

$$\frac{XES_\tau}{QES_\tau} \sim \frac{\xi_\tau}{q_\tau} \sim \frac{XTCE_\tau}{QTCE_\tau} \quad \text{and} \quad \frac{XES_\tau}{\xi_\tau} \sim \frac{1}{1-\gamma} \sim \frac{XTCE_\tau}{\xi_\tau} \quad \text{as } \tau \rightarrow 1.$$

Propositions 2(i) and (3) then afford additional convincing arguments that the expectile-based ES may be a reasonable alternative to the classical quantile-based version. Indeed, the new form XES_τ is coherent and keeps the intuitive meaning of $XTCE_\tau$ as a conditional expectation when $\tau \rightarrow 1$, since $XES_\tau \sim XTCE_\tau$. As is the case in the duality (5) between the expectile ξ_τ and the VaR q_τ , the choice in practice between the expectile-based form of ES and its quantile-based analogue will then depend on the value at hand of $\gamma \lesseqgtr \frac{1}{2}$. More precisely, the quantity XES_τ is more extreme (respectively, less extreme) than its quantile-based version QES_τ , for all τ large enough, when $\gamma > \frac{1}{2}$ (respectively, $\gamma < \frac{1}{2}$).

The connections in Proposition 3 are very useful when it comes to interpreting and proposing estimators for XES_τ . Also, by considering the second-order regular variation condition $\mathcal{C}_2(\gamma, \rho, A)$, one may establish a precise control of the remainder term which arises in the asymptotic equivalent $XES_\tau/\xi_\tau \sim (1-\gamma)^{-1}$. This will prove instrumental when examining asymptotic properties of our tail expectile-based ES estimators.

PROPOSITION 4. *Assume that $\mathbb{E}|Y_-| < \infty$. Assume further that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1$. Then, as $\tau \rightarrow 1$,*

$$\begin{aligned} \frac{XES_\tau}{\xi_\tau} &= \frac{1}{1-\gamma} \left(1 - \frac{\gamma^2(\gamma^{-1}-1)^\gamma}{q_\tau} (\mathbb{E}(Y) + o(1)) \right. \\ &\quad \left. + \frac{1-\gamma}{(1-\gamma-\rho)^2} (\gamma^{-1}-1)^{-\rho} A((1-\tau)^{-1})(1+o(1)) \right). \end{aligned}$$

The asymptotic expansion in Proposition 4, like the asymptotic expansion of expectiles in Proposition 1(i), includes a term proportional to the function A and another term proportional to the inverse of the quantile function q . This guarantees that the bias conditions in the estimation of an extreme XES_τ , at a proper extreme level $\tau = \tau'_n \rightarrow 1$ such that $n(1-\tau'_n) \rightarrow c < \infty$, will be similar to those assumed in the estimation of an extreme expectile ξ_τ .

5.3. *Estimation and asymptotics.* Propositions 1(i) and 4 indicate that the expectile-based ES satisfies a regular variation property in the same way as quantiles and expectiles do. To estimate an extreme value $\text{XES}_{\tau'_n}$, where $\tau'_n \rightarrow 1$ and $n(1 - \tau'_n) \rightarrow c < \infty$, we may therefore start by estimating XES_{τ_n} , with τ_n being an intermediate level, before extrapolating this estimator to the far tail using an estimator of the tail index γ . A natural estimator of XES_{τ_n} is its direct empirical counterpart:

$$\widetilde{\text{XES}}_{\tau_n} := \frac{1}{1 - \tau_n} \int_{\tau_n}^1 \tilde{\xi}_t dt,$$

obtained simply by replacing ξ_t in (16) with its sample version $\tilde{\xi}_t$ described in (2). Since this estimator is a linear functional of the tail empirical expectile process, Theorem 3 is more adapted than Theorem 2 for the analysis of its asymptotic distribution. Technically, using Theorem 2 would involve connecting $\widetilde{\text{XES}}_{\tau_n}$ to ξ_{τ_n} , before connecting ξ_{τ_n} back to XES_{τ_n} via Proposition 4. In doing so, the second step adds another level of complexity related to the control of bias terms proportional to the auxiliary function A and to the inverse of the quantile function q . Using Theorem 3 is much less demanding as it allows us to avoid these superfluous bias terms and to address the convergence problem directly by comparing $\widetilde{\text{XES}}_{\tau_n}$ with XES_{τ_n} .

THEOREM 8. *Under the conditions of Theorem 3,*

$$\sqrt{n(1 - \tau_n)} \left(\frac{\widetilde{\text{XES}}_{\tau_n}}{\text{XES}_{\tau_n}} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3(1 - \gamma)(3 - 4\gamma)}{(1 - 2\gamma)^3} \right).$$

On the basis of Proposition 3 and then of the approximation (13), we have for $\tau_n < \tau'_n \rightarrow 1$ that

$$\frac{\text{XES}_{\tau'_n}}{\text{XES}_{\tau_n}} \sim \frac{\xi_{\tau'_n}}{\xi_{\tau_n}} \approx \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\gamma}.$$

Therefore, to estimate $\text{XES}_{\tau'_n}$ at an arbitrary extreme level τ'_n , we replace γ by the *expectHill* estimator $\bar{\gamma}_{\tau_n}(\alpha)$ and XES_{τ_n} at an intermediate level τ_n by the estimator $\widetilde{\text{XES}}_{\tau_n}$ to get

$$(17) \quad \widetilde{\text{XES}}_{\tau'_n}^*(\alpha) := \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\bar{\gamma}_{\tau_n}(\alpha)} \widetilde{\text{XES}}_{\tau_n}.$$

The next result analyzes the convergence of this Weissman-type estimator.

THEOREM 9. *Assume that the conditions of Theorem 7 hold. Then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \left(\frac{\widetilde{XES}_{\tau'_n}^*(\alpha)}{XES_{\tau'_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha),$$

with b_α and v_α as in (10) and (11).

One can also design alternative options for estimating $XES_{\tau'_n}$ by using the asymptotic connections in Proposition 3. The asymptotic equivalence $XES_{\tau'_n} \sim (1-\gamma)^{-1} \xi_{\tau'_n}$, established therein, suggests that $XES_{\tau'_n}$ can be estimated consistently by substituting the tail quantities γ and $\xi_{\tau'_n}$ with their consistent estimators described in (9) and (14), respectively. This yields the following extrapolated estimator:

$$(18) \quad \overline{XES}_{\tau'_n}^*(\alpha, \beta) := [1 - \bar{\gamma}_{\tau_n}(\alpha)]^{-1} \bar{\xi}_{\tau'_n}^*(\alpha, \beta)$$

for the weights $\alpha, \beta \in \mathbb{R}$. Another option motivated by the second asymptotic equivalence $XES_{\tau'_n} \sim \frac{\xi_{\tau'_n}}{q_{\tau'_n}} QES_{\tau'_n}$, as established in Proposition 3, would be to estimate $XES_{\tau'_n}$ by

$$(19) \quad \widehat{XES}_{\tau'_n}^*(\alpha, \beta) := \frac{\widehat{QES}_{\tau'_n}^*(\alpha)}{\widehat{q}_{\tau'_n}^*(\alpha)} \bar{\xi}_{\tau'_n}^*(\alpha, \beta)$$

for the estimators $\widehat{q}_{\tau'_n}^*(\alpha)$ of $q_{\tau'_n}$ and $\widehat{QES}_{\tau'_n}^*(\alpha)$ of $QES_{\tau'_n}$ defined as

$$(20) \quad \widehat{q}_{\tau'_n}^*(\alpha) := \left(\frac{1-\tau'_n}{1-\tau_n} \right)^{-\bar{\gamma}_{\tau_n}(\alpha)} \widehat{q}_{\tau_n},$$

$$(21) \quad \widehat{QES}_{\tau'_n}^*(\alpha) := \left(\frac{1-\tau'_n}{1-\tau_n} \right)^{-\bar{\gamma}_{\tau_n}(\alpha)} \frac{1}{[n(1-\tau_n)]} \sum_{i=1}^{\lfloor n(1-\tau_n) \rfloor} Y_{n-i+1,n}.$$

In the special case $\alpha = 1$, the latter estimators are identical to the popular $q_{\tau'_n}$ estimator of Weissman [39] and to the extrapolated $QES_{\tau'_n}$ estimator of El Methni *et al.* [19], respectively.

THEOREM 10. *Assume that the conditions of Theorem 7 hold. Then, for any $\alpha, \beta \in \mathbb{R}$,*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \left(\frac{\overline{XES}_{\tau'_n}^*(\alpha, \beta)}{XES_{\tau'_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha)$$

and

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \left(\frac{\widehat{XES}_{\tau'_n}^*(\alpha, \beta)}{XES_{\tau'_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha)$$

with (b_α, v_α) as in (10) and (11).

Theorems 9 and 10 are, like Theorem 7, derived by noticing that, on the one hand, the asymptotic behaviors of $\widehat{\text{XES}}_{\tau'_n}^*(\alpha)$, $\widehat{\xi}_{\tau'_n}^*(\alpha, \beta)$, $\widehat{\text{QES}}_{\tau'_n}^*(\alpha)$ and $\widehat{q}_{\tau'_n}^*(\alpha)$ are controlled by the asymptotic behavior of $\{(1 - \tau'_n)/(1 - \tau_n)\}^{-\bar{\gamma}_{\tau_n}(\alpha)}$, which is itself governed by that of $\bar{\gamma}_{\tau_n}(\alpha)$. On the other hand, the nonrandom remainder term coming from the use of Proposition 3 can be controlled thanks to Proposition 4.

5.4. *Extreme level selection.* A major practical question that remains to be addressed is the choice of the extreme level τ'_n in the tail risk measure $\text{XES}_{\tau'_n}$. Since $\text{XES}_{\tau'_n} \sim \mathbb{E}[Y|Y > \xi_{\tau'_n}]$, this problem translates into choosing $\xi_{\tau'_n}$ itself.

When moving from the conventional VaR q_{p_n} , for a pre-specified tail probability $p_n \rightarrow 1$ with $n(1 - p_n) \rightarrow c < \infty$, to the expectile $\xi_{\tau'_n}$, Bellini and Di Bernardino [7] have suggested to pick out τ'_n so that $\xi_{\tau'_n} \equiv q_{p_n}$. The expectile $\xi_{\tau'_n}$ then inherits the same intuitive probabilistic interpretation as the quantile q_{p_n} while keeping its coherence. This idea was, however, implemented for a normally distributed Y . Instead, Daouia *et al.* [11] have suggested to estimate nonparametrically the level τ'_n that satisfies $\xi_{\tau'_n} = q_{p_n}$, without recourse to any *a priori* distributional specification apart from the standard assumption (3) of heavy tails. By taking the derivative with respect to θ in the L^2 criterion (1) and setting it to zero, we get

$$\tau = \frac{\mathbb{E}\{|Y - \xi_\tau| \mathbb{I}(Y \leq \xi_\tau)\}}{\mathbb{E}|Y - \xi_\tau|} \quad \text{for all } \tau \in (0, 1).$$

The extreme expectile level $\tau'_n(p_n) := \tau'_n$ such that $\xi_{\tau'_n} \equiv q_{p_n}$ then satisfies

$$1 - \tau'_n(p_n) = \frac{\mathbb{E}\{|Y - q_{p_n}| \mathbb{I}(Y > q_{p_n})\}}{\mathbb{E}|Y - q_{p_n}|}.$$

Under the model assumption of Pareto-type tails (3), it turns out that the resulting level $\tau'_n(p_n)$ asymptotically depends only on the quantile level p_n and on the tail index γ , but not on the value of the extreme quantile q_{p_n} .

PROPOSITION 5 (Daouia *et al.*, 2018). *If Y has a Pareto-type distribution (3) with tail index $0 < \gamma < 1$, then*

$$1 - \tau'_n(p_n) \sim (1 - p_n) \frac{\gamma}{1 - \gamma} \quad \text{as } n \rightarrow \infty.$$

The proof of this result can be found in Daouia *et al.* ([11], Proposition 3). Built on the *expectHill* estimator $\bar{\gamma}_{\tau_n}(\alpha)$ of γ , we can then define a natural estimator of $\tau'_n(p_n)$ as

$$(22) \quad \hat{\tau}'_n(p_n) := 1 - (1 - p_n) \frac{\bar{\gamma}_{\tau_n}(\alpha)}{1 - \bar{\gamma}_{\tau_n}(\alpha)}.$$

By substituting this estimated value in place of $\tau'_n \equiv \tau'_n(p_n)$ in the extrapolated estimators $\widehat{\text{XES}}_{\tau'_n}^*(\alpha)$, $\overline{\text{XES}}_{\tau'_n}^*(\alpha, \beta)$ and $\widehat{\text{XES}}_{\tau'_n}^*(\alpha, \beta)$ described in (17), (18) and (19), we obtain composite estimators that estimate $\text{XES}_{\tau'_n(p_n)} \sim \text{QES}_{p_n}$, by Proposition 3. It is actually easily seen that the quantile-based estimator $\widehat{\text{QES}}_{p_n}^*(\alpha)$, defined in (21), is identical to the composite expectile-based estimator $\widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha, 1)$, obtained for the special weight $\beta = 1$. The convergence results in Theorems 9 and 10 of the extrapolated estimators $\widehat{\text{XES}}_{\tau'_n}^*(\alpha)$, $\overline{\text{XES}}_{\tau'_n}^*(\alpha, \beta)$ and $\widehat{\text{XES}}_{\tau'_n}^*(\alpha, \beta)$ still hold true for their composite versions as estimators of QES_{p_n} , with the same technical conditions.

THEOREM 11. *Suppose the conditions of Theorem 7 hold with p_n in place of τ'_n . Then, for any $\alpha, \beta \in \mathbb{R}$,*

$$\begin{aligned} & \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-p_n)]} \left(\frac{\widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha)}{\text{QES}_{p_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha), \\ & \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-p_n)]} \left(\frac{\overline{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha, \beta)}{\text{QES}_{p_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha), \\ \text{and} \quad & \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-p_n)]} \left(\frac{\widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha, \beta)}{\text{QES}_{p_n}} - 1 \right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha) \end{aligned}$$

with (b_α, v_α) as in (10) and (11).

6. Numerical simulations. In order to illustrate the behavior of the presented estimation procedures of the tail index γ and the two expected shortfall forms $\text{XES}_{\tau'_n}$ and QES_{p_n} , we consider the Student t -distribution with degree of freedom $1/\gamma$, the Fréchet distribution $F(x) = e^{-x^{-1/\gamma}}$, $x > 0$, and the Pareto distribution $F(x) = 1 - x^{-1/\gamma}$, $x > 1$. The finite-sample performance of the different estimators is evaluated through their relative Mean-Squared Error (MSE) and bias, computed over 200 replications. All the experiments have sample size $n = 500$ and true tail index $\gamma \in \{0.35, 0.45\}$ (motivated by our real data applications where the realized values of γ were found to vary between 0.35 and 0.45). In our estimators we used the extreme

levels $\tau'_n = p_n = 1 - 1/n$ and the intermediate level $\tau_n = 1 - k/n$, where the integer k can be viewed as the effective sample size for tail extrapolation. To save space, all figures illustrating our simulation results are deferred to Section C of the Supplementary Material document.

6.1. *Estimation of the tail index.* Our Monte-Carlo simulations in Supplement C.1 indicate that the *expectHill* estimator $\bar{\gamma}_{1-k/n}(\alpha)$, introduced in (9) with the weight $\alpha = 1/2$, is more efficient relative to the standard Hill estimator $\hat{\gamma}_{1-k/n}$, given in (8), for both Student and Fréchet distributions. In the case of the real-valued Student distribution, it may be seen therein that $\bar{\gamma}_{1-k/n}(\frac{1}{2})$ performs better than $\hat{\gamma}_{1-k/n}$ in terms of MSE, for all values of k , without sacrificing too much quality in terms of bias, especially for the larger value of γ . We arrive at the same tentative conclusion in the case of the Fréchet distribution. By contrast, in the special case of the Pareto distribution, the Hill estimator $\hat{\gamma}_{1-k/n}$ is exactly the maximum likelihood estimator of γ and is unbiased, whereas the *expectHill* estimator $\bar{\gamma}_{1-k/n}(\frac{1}{2}) = \frac{1}{2}(\hat{\gamma}_{1-k/n} + \tilde{\gamma}_{1-k/n})$ is biased in this case. Unsurprisingly, the Monte Carlo results obtained in this case indicate that $\hat{\gamma}_{1-k/n}$ is, as expected, the winner.

6.2. *Expected Shortfall estimation.*

6.2.1. *Estimates of $XES_{\tau'_n}$.* Before comparing the finite-sample performance of $\widetilde{XES}_{\tau'_n}^*(\alpha)$ described in (17), $\overline{XES}_{\tau'_n}^*(\alpha, \beta)$ in (18) and $\widehat{XES}_{\tau'_n}^*(\alpha, \beta)$ in (19), as estimators of $XES_{\tau'_n}$, we first investigated the accuracy of each estimator in terms of the associated weights α and β . Then we compared the three estimators with each other by using the best choice of α and β in each scenario; see Supplement C.2. In particular, we arrive at the following tentative conclusion: $\widetilde{XES}_{\tau'_n}^*(\alpha)$ seems to be the winner in the case of the real-valued Student distribution for $\alpha = 1$, while $\widehat{XES}_{\tau'_n}^*(\alpha, \beta)$ appears to be the most efficient in the case of the non-negative Fréchet and Pareto distributions, for $\alpha \in \{0.5, 1\}$ and $\beta = 1$.

6.2.2. *Estimates of QES_{p_n} .* We have also undertaken simulation experiments to evaluate the finite-sample performance of the composite versions $\widetilde{XES}_{\hat{\tau}'_n(p_n)}^*(\alpha)$, $\overline{XES}_{\hat{\tau}'_n(p_n)}^*(\alpha, \beta)$ and $\widehat{XES}_{\hat{\tau}'_n(p_n)}^*(\alpha, \beta)$ studied in Theorem 11, with $\hat{\tau}'_n(p_n)$ being described in (22). These composite expectile-based estimators estimate the same conventional expected shortfall QES_{p_n} as the direct quantile-based estimator $\widehat{QES}_{p_n}^*(\alpha)$ defined in (21). In Supplement C.3, we first examined the accuracy of each estimator for various values of α and β ,

and then we compared the four estimators with each other. We arrive at the following tentative conclusions:

- In the case of the (real-valued) Student distribution, the best estimator seems to be $\widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha = 0)$;
- In the cases of Fréchet and Pareto distributions (both positive), the best estimators seem to be, respectively, $\widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha = 0.5, \beta = 1)$ and $\widehat{\text{QES}}_{p_n}^*(\alpha = 1) \equiv \widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha = 1, \beta = 1)$.

6.2.3. *Confidence intervals for QES_{p_n} .* By Theorem 11 we have

$$\frac{\sqrt{k}}{\log[k/n(1-p_n)]} \left(\frac{\widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha)}{\widehat{\text{QES}}_{p_n}^*} - 1 \right) \xrightarrow{d} \mathcal{N}(b_\alpha(\gamma), v_\alpha(\gamma)),$$

where $b_\alpha(\gamma) := b_\alpha$ and $v_\alpha(\gamma) := v_\alpha$ are described in (10) and (11), respectively. Under the bias condition $\lambda_1 = \lambda_2 = 0$ in Theorem 4, the asymptotic bias in (10) reduces to $b_\alpha(\gamma) = 0$. With this condition, the (symmetric) expectile-based asymptotic confidence interval with confidence level $100\vartheta\%$ has the form $\widetilde{\text{CI}}_\vartheta(k) = \widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha) \times \mathcal{I}$, where \mathcal{I} stands for the interval

$$\mathcal{I} := \left[1 \pm z_{(1+\vartheta)/2} \log \left(\frac{k}{n(1-p_n)} \right) \sqrt{v_\alpha \left(\bar{\gamma}_{1-k/n}(\alpha) \right) / k} \right],$$

with $z_{(1+\vartheta)/2}$ being the $(1+\vartheta)/2$ -quantile of the standard Gaussian distribution. Likewise, the confidence intervals derived from the asymptotic normality of $\widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha)$ and $\widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha, \beta)$, in Theorem 11, can be expressed respectively as

$$\overline{\text{CI}}_\vartheta(k) = \widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha, \beta) \times \mathcal{I}, \quad \widehat{\text{CI}}_\vartheta(k) = \widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha, \beta) \times \mathcal{I}.$$

Note also that the quantile-based confidence interval, derived from the asymptotic normality of $\widehat{\text{QES}}_{p_n}^*(\alpha) \equiv \widehat{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha, 1)$, is just $\widehat{\text{CI}}_\vartheta(k)$ for $\beta = 1$. In Supplement C.4, we compared the average lengths and the achieved coverages of the three 95% asymptotic confidence intervals $\widetilde{\text{CI}}_{0.95}(k)$, $\overline{\text{CI}}_{0.95}(k)$ and $\widehat{\text{CI}}_{0.95}(k)$. It follows that

- $\widetilde{\text{CI}}_{0.95}(k)$ performs best in the case of the Student distribution, for the selected weight $\alpha = 1$;
- $\widehat{\text{CI}}_{0.95}(k)$ performs quite well in the case of the Fréchet distribution, for the selected weights $\alpha = 1$ and $\beta = 1$;
- $\overline{\text{CI}}_{0.95}(k)$ performs quite well in the case of the Pareto distribution, for the selected weights $\alpha = 1$ and $\beta = 0.5$.

7. Application to medical insurance data. The Society of Actuaries (SOA) group Medical Insurance large claims database contains 75,789 claim amounts exceeding 25,000 USD, collected over the year 1991 from 26 insurers. The full database which records about 3 million claims over the period 1991-92 is available at <http://www.soa.org>. The scatterplot and histogram of the 1991 log-claim amounts, displayed in Figure 1(a), exhibit a considerable right-skewness. Beirlant *et al.* ([5], p.123) have argued that the underlying distribution is heavy-tailed with a γ estimate around 0.35. A traditional instrument to assess the magnitude of future unexpected higher claim amounts is the expected shortfall QES_{p_n} described in (15). Insurance companies typically are interested in an extremely low exceedance probability of the order of $1/n$, say, $1 - p_n = 1/100,000$ for the sample size $n = 75,789$. This corresponds to rare events that happen on average only once in 100,000 cases.

In this situation of non-negative data with heavy right tail, our experience with simulated data indicates that $\overline{\text{XES}}_{\hat{\gamma}'_n(p_n)}^*(\alpha = 0.5, \beta = 1)$ and $\widehat{\text{QES}}_{p_n}^*(\alpha = 1)$ provide the best extrapolated pointwise estimates of the extreme value QES_{p_n} in terms of MSE and bias. As such, these are the estimates we adopt here. For the sake of simplicity, they will be denoted by $\overline{\text{XES}}_{\hat{\gamma}'_n(p_n)}^*$ and $\widehat{\text{QES}}_{p_n}^*$, respectively.

The path of the composite expectile-based estimator $\overline{\text{XES}}_{\hat{\gamma}'_n(p_n)}^*$ against the sample fraction k is shown in Figure 1(b) as the rainbow curve, for the selected range of intermediate values of $k = 10, 11, \dots, 700$. The effect of the *expectHill* estimate $\bar{\gamma}_{1-k/n}(\alpha = 0.5)$ on $\overline{\text{XES}}_{\hat{\gamma}'_n(p_n)}^*$ is highlighted by a colour-scheme, ranging from dark red (low $\bar{\gamma}_{1-k/n}$) to dark violet (high $\bar{\gamma}_{1-k/n}$). This γ estimate seems to mainly vary within the interval $[0.35, 0.36]$, which corresponds to the stable (green) part of the plot. The curve $k \mapsto \overline{\text{XES}}_{\hat{\gamma}'_n(p_n)}^*$ exceeds overall the sample maximum $Y_{n,n} = 4.51$ million (indicated by the horizontal pink dashed line). To select a reasonable pointwise estimate, we applied a simple automatic data-driven device that consists first in computing the standard deviations of $\overline{\text{XES}}_{\hat{\gamma}'_n(p_n)}^*$ over a moving window large enough to cover 20% of the possible values of k in the selected range $10 \leq k \leq 700$. Then the k where the standard deviation is minimal defines the desired sample fraction. The resulting estimate $\overline{\text{XES}}_{\hat{\gamma}'_n(p_n)}^* = 5.99$ million is obtained for the value $k = 208$ in the window $[119, 259]$.

The path of the direct quantile-based estimator $\widehat{\text{QES}}_{p_n}^*$ against k is graphed in the same figure as dashed black curve. It is broadly similar to that of $\overline{\text{XES}}_{\hat{\gamma}'_n(p_n)}^*$, but the latter is smoother and more stable. The pointwise estimate $\widehat{\text{QES}}_{p_n}^* = 6.37$ million is indicated by the minimal standard deviation

achieved at $k = 222$ over the window $[119, 259]$. It is more pessimistic (in risk assessment terminology) than $\overline{\text{XES}}_{\hat{\tau}'_n(p_n)}^* = 5.99$ million, probably due to the instability of the quantile-based plot in dashed black.

Our experience with simulated data also indicates that reasonably good asymptotic 95% confidence intervals for QES_{p_n} , in terms of average lengths and achieved coverages, are provided by $\widehat{\text{CI}}_{0.95}(k)$, constructed via $\widehat{\text{QES}}_{p_n}^*$, and $\overline{\text{CI}}_{0.95}(k)$ constructed on $\overline{\text{XES}}_{\hat{\tau}'_n(p_n)}^*(\alpha = 1, \beta = 0.5)$. The two confidence intervals $\overline{\text{CI}}_{0.95}(k)$ and $\widehat{\text{CI}}_{0.95}(k)$ are superimposed in Figure 1(b) as well, respectively, in dotted blue and solid grey lines. Though $\overline{\text{CI}}_{0.95}(k)$ gives slightly more pessimistic confidence bounds than $\widehat{\text{CI}}_{0.95}(k)$, both confidence intervals point towards similar conclusions. In particular, the stable parts of their lower boundaries (around $k \in [100, 500]$) remain quite conservative as they are very close from the maximum recorded claim amount.

Finally, we would like to comment on the estimator $\hat{\tau}'_n(p_n)$ of the extreme expectile level $\tau'_n(p_n)$ which ensures that $\overline{\text{XES}}_{\hat{\tau}'_n(p_n)}^*$ estimates $\text{XES}_{\tau'_n(p_n)} \sim \text{QES}_{p_n}$. The plot of $\hat{\tau}'_n(p_n)$ versus k is graphed in Figure 1(c) as rainbow curve, and the corresponding optimal pointwise estimate is indicated by the horizontal dashed black line. This selected optimal level $\hat{\tau}'_n(p_n) = 0.9999944$ is much higher than the pre-specified relative frequency $p_n = 0.99999$ indicated by the horizontal dashed pink line. This is actually in line with our theoretical findings in Proposition 3 that lead in conjunction with (5) to

$$\frac{\text{XES}_{p_n}}{\text{QES}_{p_n}} \sim \frac{\xi_{p_n}}{q_{p_n}} \sim (\gamma^{-1} - 1)^{-\gamma} \quad \text{as } p_n \rightarrow 1.$$

Since $\gamma < 1/2$, it follows that XES_{p_n} is less extreme than $\text{QES}_{p_n} \sim \text{XES}_{\tau'_n(p_n)}$, for all p_n large enough. Therefore $p_n < \tau'_n(p_n)$ by monotonicity of $\tau \mapsto \text{XES}_\tau$, which follows from the fact that $\text{XES}_\tau = (1 - \tau)^{-1} \int_\tau^1 \xi_t dt$, where the expectile function $t \mapsto \xi_t$ is continuous and strictly increasing by Proposition 1 in Holzmam and Klar [25].

Supplementary Material. The supplement to this article contains simulation results and a second application to financial data, along with technical lemmas and the proofs of all our theoretical results.

REFERENCES

- [1] Abdous, B. and Remillard, B. (1995). Relating quantiles and expectiles under weighted-symmetry, *Ann. Inst. Statist. Math.*, **47**, 371–384.
- [2] Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking and Finance*, **26**, 1505–1518.

- [3] Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall, *Journal of Banking and Finance*, **26**, 1487–1503.
- [4] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent Measures of Risk, *Math. Finance*, **9**, 203–228.
- [5] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley.
- [6] Bellini, F., Klar, B., Müller, A. and Gianina, E.R. (2014). Generalized quantiles as risk measures, *Insurance Math. Econom.*, **54**, 41–48.
- [7] Bellini, F. and Di Bernardino, E. (2017). Risk management with expectiles, *The European Journal of Finance*, **23**, 487–506.
- [8] Breckling, J. and Chambers, R. (1988). M-quantiles, *Biometrika*, **75**, 761–772.
- [9] Cai, J.J., Einmahl, J.H.J., de Haan, L. and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme, *J. R. Stat. Soc. Ser. B*, **77**, 417–442.
- [10] Daniélsso, J., Embrechts, P., Goodhart, C., Keating, C., Muenich, F., Renault, O. and Shin, H.S. (2001). An Academic Response to Basel II. Special paper no. 130, Financial Markets Group, London School of Economics.
- [11] Daouia, A., Girard, S. and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles, *J. R. Stat. Soc. Ser. B*, **80**, 263–292.
- [12] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer-Verlag, New York.
- [13] de Haan, L., Mercadier, C. and Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence, *Finance Stoch.*, **20**, 321–354.
- [14] Dietrich, D., de Haan, L. and Hüslér, J. (2002). Testing extreme value conditions, *Extremes*, **5**, 71–85.
- [15] Drees, H. (1998). On smooth statistical tail functionals, *Scand. J. Stat.*, **25**, 187–210.
- [16] Drees, H., de Haan, L. and Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions, *J. Statist. Plann. Inference*, **136**, 3498–3538.
- [17] Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot, *Ann. Statist.*, **28**, 254–274.
- [18] Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations, and forecast rankings, *J. R. Stat. Soc. Ser. B*, **78**, 505–562.
- [19] El Methni, J., Gardes, L. and Girard, S. (2014). Nonparametric estimation of extreme risks from conditional heavy-tailed distributions, *Scand. J. Stat.*, **41**, 988–1012.
- [20] El Methni, J. and Stupfler, G. (2017a). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions, *Statist. Sinica*, **27**, 907–930.
- [21] El Methni, J. and Stupfler, G. (2017b). Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions, *Econometrics and Statistics* (to appear).
- [22] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer.
- [23] Hall, P. and Welsh, A.W. (1985). Adaptive estimates of parameters of regular variation, *Ann. Statist.*, **13**, 331–341.

- [24] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**, 1163–1174.
- [25] Holzmam, H. and Klar, B. (2016). Expectile asymptotics, *Electron. J. Stat.*, **10**, 2355–2371.
- [26] Hüslér, J. and Li, D. (2006). On testing extreme value conditions, *Extremes*, **9**, 69–86.
- [27] Koenker, R. and Bassett, G.S. (1978). Regression quantiles, *Econometrica*, **46**, 33–50.
- [28] Krätshmer, V. and Zähle, H. (2017). Statistical inference for expectile-based risk measures, *Scand. J. Stat.*, **44**, 425–454.
- [29] Kuan, C.-M., Yeh, J.-H. and Hsu, Y.-C. (2009). Assessing value at risk with CARE, the Conditional Autoregressive Expectile models, *J. Econometrics*, **2**, 261–270.
- [30] Mao, T., Ng, K. and Hu, T. (2015). Asymptotic expansions of generalized quantiles and expectiles for extreme risks, *Probab. Engrg. Inform. Sci.*, **29**, 309–327.
- [31] Mao, T. and Yang, F. (2015). Risk concentration based on expectiles for extreme risks under FGM copula, *Insurance Math. Econom.*, **64**, 429–439.
- [32] Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing, *Econometrica*, **55**, 819–847.
- [33] Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer.
- [34] Resnick, S. and Stărică, C. (1997). Smoothing the Hill estimator, *Adv. in Appl. Probab.*, **29**, 271–293.
- [35] Rockafellar, R.T. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk, *Journal of Risk*, **2**, 21–42.
- [36] Rockafellar, R.T. and Uryasev, S. (2002). Conditional Value-at-Risk for General Loss Distributions, *Journal of Banking and Finance*, **26**, 1443–1471.
- [37] Sobotka, F. and Kneib, T. (2012). Geoadditive expectile regression, *Comput. Statist. Data Anal.*, **56**, 755–767.
- [38] Taylor, J. (2008). Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics*, **6**, 231–252.
- [39] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.
- [40] Wirch, J.L. and Hardy, M.R. (1999). A synthesis of risk measures for capital adequacy, *Insurance Math. Econom.*, **25**, 337–347.
- [41] Ziegel, J.F. (2016). Coherence and elicibility, *Math. Finance*, **26**, 901–918.
- [42] Zwingmann, T. and Holzmam, H. (2016). Asymptotics for the expected shortfall, preprint.

TOULOUSE SCHOOL OF ECONOMICS
 UNIVERSITY OF TOULOUSE CAPITOLE
 MANUFACTURE DES TABACS,
 AILE JEAN-JACQUES LAFFONT,
 21 ALLÉE DE BRIENNE, 31000 TOULOUSE, FRANCE
 E-MAIL: abdelaati.daouia@tse-fr.eu

INRIA GRENOBLE RHÔNE-ALPES, MISTIS
 655, AVENUE DE L'EUROPE - CS 90051
 38334 MONTBONNOT CEDEX, FRANCE
 E-MAIL: stephane.girard@inria.fr

SCHOOL OF MATHEMATICAL SCIENCES
 UNIVERSITY OF NOTTINGHAM
 UNIVERSITY PARK CAMPUS
 NOTTINGHAM NG7 2RD, UNITED KINGDOM
 E-MAIL: gilles.stupfler@nottingham.ac.uk

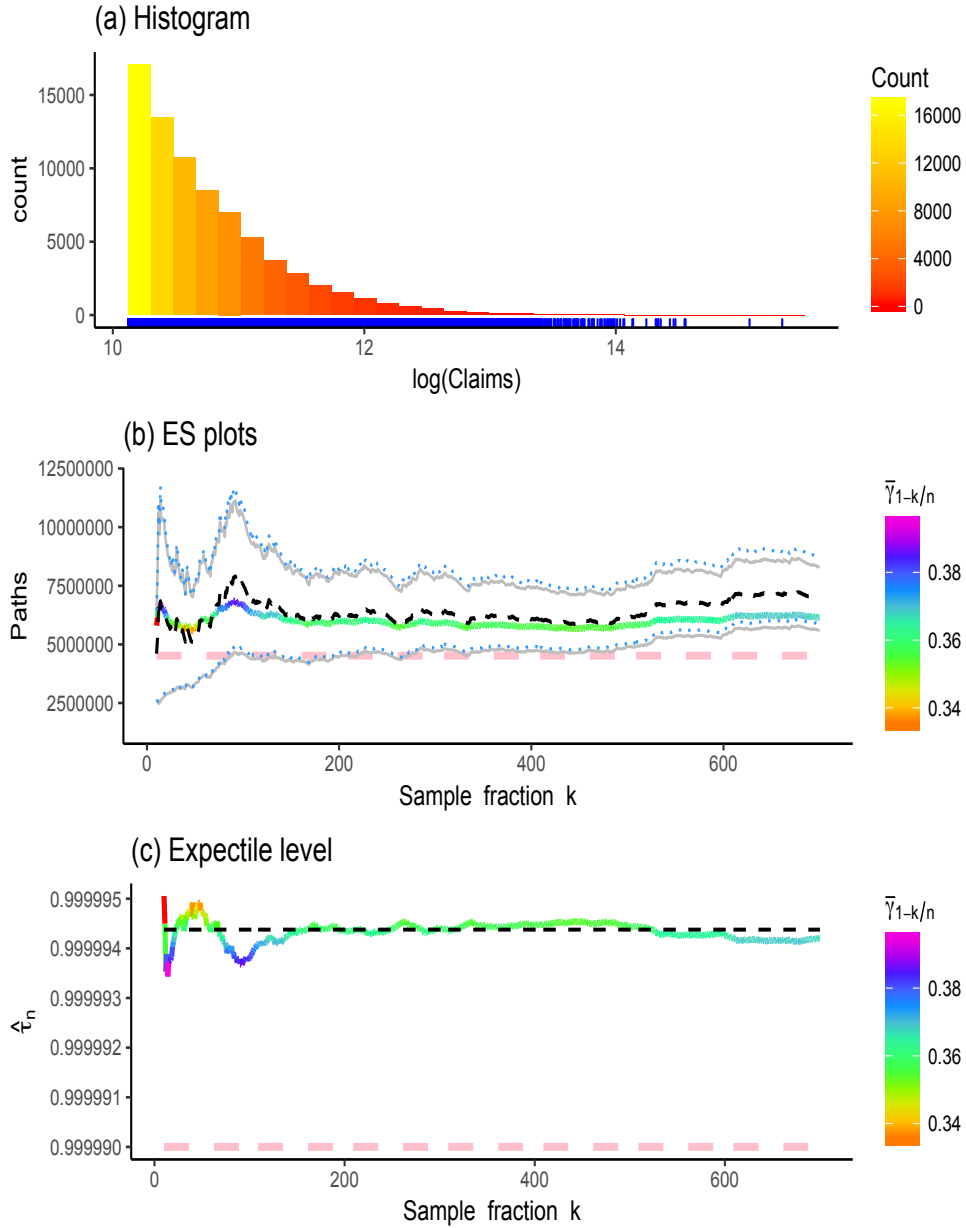


FIG 1. (a) Scatterplot and histogram of the log-claim amounts. (b) The ES plots $k \mapsto \overline{XES}_{\hat{\tau}'_n(p_n)}^*(\alpha = 0.5, \beta = 1)$ as rainbow curve, and $k \mapsto \widehat{QES}_{p_n}^*(\alpha = 1)$ in dashed black, along with the constant sample maximum $Y_{n,n}$ in horizontal dashed pink. The confidence intervals $\overline{CI}_{0.95}(k)$ in dotted blue lines and $\widehat{CI}_{0.95}(k)$ in solid grey lines. (c) The plot of $k \mapsto \hat{\tau}'_n(p_n)$ as rainbow curve, along with the selected optimal pointwise estimate in horizontal dashed black line, and the constant tail probability p_n in horizontal dashed pink.