Tail expectile process and risk assessment

ABDELAATI DAOUIA[1,*] STÉPHANE GIRARD[2,**] and GILLES STUPFLER[3,†]

[1] *Toulouse School of Economics, University of Toulouse Capitole, France*
E-mail: [*]`abdelaati.daouia@tse-fr.eu`

[2] *Université Grenoble Alpes, INRIA, CNRS, Grenoble INP, LJK, 38000 Grenoble, France*
E-mail: [**]`stephane.girard@inria.fr`

[3] *School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK*
E-mail: [†]`gilles.stupfler@nottingham.ac.uk`

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 extreme expectiles and the standard quantile-based expected shortfall, 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 applications to actuarial and financial data are provided.

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

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 [34] concept of expectiles, a least squares analogue of traditional quantiles. Given an order $\tau \in (0, 1)$, Koenker and Bassett [29] elaborated an absolute error loss minimization to define the τth quantile of the distribution of a random variable Y as the minimizer

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

with equality if the distribution function of Y is increasing, where $\rho_\tau(y) = |\tau - \math{I}(y \leqslant 0)| \, |y|$ and $\math{I}(\cdot)$ is the indicator function. This successfully extends the conventional definition of quantiles as left-continuous inverse functions. Newey and Powell [34] replaced

the absolute deviations in the asymmetric loss function ρ_τ with squared deviations to obtain the τth expectile of a random variable Y with finite first moment as

$$\xi_\tau = \underset{\theta \in \mathbb{R}}{\arg\min} \, \mathbb{E} \left\{ \eta_\tau(Y - \theta) - \eta_\tau(Y) \right\}, \tag{1}$$

with $\eta_\tau(y) = |\tau - \mathbf{I}(y \leqslant 0)| \, y^2$. 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 as risk measures in statistical finance and actuarial science since the pioneering paper of Kuan *et al.* [31]. As established very recently in this literature of risk management (see, for instance [6, 7, 13, 44]), expectiles are excellent alternatives to quantiles in different aspects relevant to this kind of applications. They 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.* [31]). Expectiles, contrary to quantiles, thus allow to measure extreme risk based on both the frequency of tail losses and their severity. 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 [39]). This high sensitivity of expectiles to tail behavior allows for more prudent and reactive risk management. Another substantial difference is that expectiles 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 [38]). Also, inference on expectiles is much easier than inference on quantiles (Abdous and Remillard [1]). Using expectiles avoids distributional assumptions (Taylor [39]) without recourse to regularity assumptions as can be seen by comparing, *e.g.*, Holzmann and Klar [27] with Zwingmann and Holzmann [45]. 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 [44]). Many other theoretical and numerical results motivate the adoption of expectiles in actuarial and financial risk management, including those of Ehm *et al.* [20] 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.* [32], Mao and Yang [33] and Bellini and Di Bernardino [7] have initiated the study of the connection between ξ_τ and q_τ, as $\tau \to 1$, when Y belongs to the domain of attraction of a Generalized Extreme Value distribution. Also, for heavy-tailed distributions, Daouia *et al.* [13] have obtained an asymptotic expansion of ξ_τ/q_τ with a precise quantification of the remainder 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 [27] and Krätschmer and Zähle [30] 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 $\tau = \tau_n \to 1$ such that $n(1-\tau_n) \to \infty$ as the sample size $n \to \infty$; see Theorem 2 of Daouia *et al.* [13]. 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 three decades; see, among others, Csörgő and Horváth [10] and Einmahl and Mason [21] as well as their more modern formulations in Drees [18] and Theorem 2.4.8 in de Haan and Ferreira [15]. 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.* [17], Drees *et al.* [19], Hüsler and Li [28]), bias-corrected extreme value index estimators (de Haan *et al.* [16]) and estimators of extreme Wang distortion risk measures (El Methni and Stupfler [23, 24]).

The present paper tries to fill this gap in the current understanding of sample intermediate expectiles, under Pareto-type models that accurately describe the tail structure of most actuarial and financial data [see, *e.g.*, Embrechts *et al.* ([25], p.9) and Resnick ([35], p.1)]. In Section 2, we show that the aforementioned convergence result on single intermediate sample expectiles can be generalized to the tail empirical expectile process. We first prove in Theorem 1 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 2 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 1 to extreme expectile estimation. We first construct a general class of weighted estimators for intermediate expectiles ξ_{τ_n}, by combining nonparametric asymmetric least squares estimates with semiparametric quantile-based estimates. The latter involve the traditional Hill estimator of the tail index (Hill [26]). Thanks to the joint convergence result on the tail expectile and quantile processes, in Theorem 1, we derive the joint asymptotic normality of both empirical intermediate expectile and quantile estimators with Hill's tail index estimator in Theorem 3. Built on this theorem, we obtain the asymptotic normality of the generalized weighted ξ_{τ_n} estimators in Theorem 4. Based on the ideas of Daouia *et al.* [13], our 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 5.

Theorem 2 is particularly important in tail risk estimation using Expected Shortfall (ES) measures. In Section 4, we show first that the expectile-based form XTCE_τ of ES introduced by Taylor [39] 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 XTCE_τ. Asymptotic connections of XES_τ to other tail quantities, such as high quantiles q_τ and expectiles ξ_τ, are also provided in Proposition 3 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 from the true

XES_{τ_n} can be unraveled thanks to Theorem 2. 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 $n(1 - \tau'_n) \to c < \infty$, as $n \to \infty$ (see, for example, Cai *et al.* [9] and Daouia *et al.* [13]). In Theorems 7 and 8 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 τ'_n, whose asymptotic properties are established in Theorem 9.

Section 5 contains a simulation study of the estimators introduced hereafter. Applications to medical insurance data and financial returns data are presented in Section 6. Section 7 concludes. The Supplementary Material document contains all proofs and auxiliary results along with further simulation results.

2. Tail empirical expectile process

Suppose we observe independent copies $\{Y_1, \ldots, Y_n\}$ of a random variable Y and denote by $Y_{1,n} \leqslant Y_{2,n} \leqslant \cdots \leqslant Y_{n,n}$ their nth order statistics. A high expectile ξ_{τ_n} of order $\tau_n \to 1$, as $n \to \infty$, can be estimated by its empirical counterpart

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

Here the expectile level τ_n approaches one at an 'intermediate' rate in the sense that $n(1 - \tau_n) \to \infty$ as $n \to \infty$. By analogy to the well-known *tail empirical quantile process* (see Definition 2.4.3 in de Haan and Ferreira [15])

$$(0, 1] \to \mathbb{R}, \ s \mapsto \widehat{q}_{1-ks/n} := Y_{n-\lfloor ks \rfloor, n},$$

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

$$(0, 1] \to \mathbb{R}, \ s \mapsto \widetilde{\xi}_{1-(1-\tau_n)s}.$$

Note that the tail quantile process is nothing but $\{\widehat{q}_{1-(1-\tau_n)s}\}_{0 < s \leqslant 1}$ 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 functions of these heavy-tailed distributions can be

expressed as

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

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 $L(ty)/L(t) \to 1$ as $t \to \infty$ for any $y > 0$. Equivalently, by Corollary 1.2.10 in de Haan and Ferreira [15], the tail quantile function of Y, defined as $U(t) := q_{1-t^{-1}} \equiv \inf\{y \in \mathbb{R} \,|\, 1/\overline{F}(y) \geqslant t\}$, satisfies

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

The index γ tunes the tail heaviness of $\overline{F}$: 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

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

(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.* [13] under the following standard second-order extreme value condition:

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

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

where $\rho \leqslant 0$ 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 [15], along with abundant examples of commonly used continuous distributions satisfying $\mathcal{C}_2(\gamma, \rho, A)$. The asymptotic expansion in Daouia *et al.* [13] 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 $\tau \to 1$,

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

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

$$\frac{\xi_{1-(1-\tau_n)s}}{\xi_{\tau_n}} = s^{-\gamma} \left(1 + (s^\gamma - 1)\frac{\gamma(\gamma^{-1} - 1)^\gamma}{q_{\tau_n}}(\mathbb{E}(Y) + o(1)) \right.$$

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

Part (i) of this proposition relaxes the conditions in Corollary 1 of Daouia *et al.* [13] 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

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

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 [15]). In addition to satisfying $k \to \infty$ and $k/n \to 0$, 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 valid for a suitable version of the quantile process, equal to the original one in distribution, on a rich enough probability space (potentially different from the original space on which the Y_i's are defined). We will work in the sequel with this version of the quantile process and thus with the associated Y_i's. Our weak convergence results in Sections 3 and 4 are of course unaffected by this choice. Full details of where this suitable choice of process has to be made can be found in the Supplementary Material document. Besides, the function A_0 is actually asymptotically equivalent to A, see Theorem 2.3.9 in de Haan and Ferreira [15]. We may therefore write:

$$\frac{\widehat{q}_{1-(1-\tau_n)s}}{q_{\tau_n}} = s^{-\gamma} \left(1 + \frac{1}{\sqrt{n(1 - \tau_n)}} \gamma s^{-1} W_n(s) + \frac{s^{-\rho} - 1}{\rho} A((1 - \tau_n)^{-1}) \right.$$

$$\left. + o_{\mathbb{P}}\left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1 - \tau_n)}} \right) \right) \quad \text{uniformly in } s \in (0, 1], \tag{6}$$

where we set $k = n(1 - \tau_n)$, with $\tau_n \to 1$ and $n(1 - \tau_n) \to \infty$. As regards the tail expectile process, it is known from Theorem 2 of Daouia *et al.* [13] that $\widetilde{\xi}_{\tau_n}$ is, under certain conditions, an asymptotically normal estimator of ξ_{τ_n}. Similarly to the uniform approximation (6) of the tail quantile process, this result can be generalized to a uniform approximation of the tail expectile process $s \mapsto \widetilde{\xi}_{1-(1-\tau_n)s}$. This is the first main result of the paper.

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

$$
\frac{\widehat{q}_{1-(1-\tau_n)s}}{q_{\tau_n}} = s^{-\gamma}\left(1 + \frac{1}{\sqrt{n(1-\tau_n)}}\gamma\sqrt{\gamma^{-1}-1}\,s^{-1}W_n\left(\frac{s}{\gamma^{-1}-1}\right)\right.
$$

$$
\left. + \frac{s^{-\rho}-1}{\rho}A((1-\tau_n)^{-1}) + \mathrm{o}_{\mathbb{P}}\left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}}\right)\right)
$$

$$
and \quad \frac{\widetilde{\xi}_{1-(1-\tau_n)s}}{\xi_{\tau_n}} = s^{-\gamma}\left(1 + (s^\gamma - 1)\frac{\gamma(\gamma^{-1}-1)^\gamma}{q_{\tau_n}}(\mathbb{E}(Y) + \mathrm{o}_{\mathbb{P}}(1))\right.
$$

$$
+ \frac{1}{\sqrt{n(1-\tau_n)}}\gamma^2\sqrt{\gamma^{-1}-1}\,s^{\gamma-1}\int_0^s W_n(t)\,t^{-\gamma-1}\,dt
$$

$$
+ \frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} \times \frac{s^{-\rho}-1}{\rho}A((1-\tau_n)^{-1})
$$

$$
\left. + \mathrm{o}_{\mathbb{P}}\left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}}\right)\right) \quad uniformly\ in\ s \in (0,1].
$$

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

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

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 2 of Daouia *et al.* [13], subject to the additional condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$, but under the reduced moment condition

$\mathbb{E}|Y_-|^2 < \infty$. Note that the bias condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$ is also required in order to establish the desired approximation (6) for the tail quantile process. The latter approximation in the quantile case does not require, however, the extra assumptions that $\gamma \in (0, 1/2)$ and $\mathbb{E}|Y_-|^2 < \infty$. 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.* [13] and the references therein.

Nevertheless, Theorem 1 does not compare directly the tail expectile process with its population counterpart $s \mapsto \xi_{1-(1-\tau_n)s}$. Our second main result analyzes the discrepancy between these two quantities. Such a comparison is particularly relevant when developing the asymptotic theory for integrals of the tail expectile process, as discussed below in Section 4. This result cannot be obtained as a direct corollary of Theorem 1, because Proposition 1(ii) is not a uniform result.

Theorem 2.   *If the conditions of Theorem 1 hold with $\rho < 0$, then there exists a sequence W_n of standard Brownian motions such that, for any $\varepsilon > 0$ sufficiently small,*

$$\frac{\widetilde{\xi}_{1-(1-\tau_n)s}}{\xi_{1-(1-\tau_n)s}} = 1 + \frac{1}{\sqrt{n(1-\tau_n)}}\gamma^2\sqrt{\gamma^{-1}-1}\,s^{\gamma-1}\int_0^s W_n(t)\,t^{-\gamma-1}\,dt$$

$$+ \mathrm{o}_{\mathbb{P}}\left(\frac{s^{-1/2-\varepsilon}}{\sqrt{n(1-\tau_n)}}\right) \quad \textit{uniformly in } s \in (0,1].$$

Note that the Gaussian term appearing in Theorem 2 is exactly the same as in the approximation of the tail expectile process in Theorem 1. Both theorems 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 extreme expectiles and of an expectile-based analogue for the Expected Shortfall risk measure. Theorems 1 and 2 will be the key tools when it comes to unravel the asymptotic behavior of these estimators.

3. Extreme expectile estimation

In this section, we first return to intermediate expectile estimation by combining nonparametric asymmetric least squares estimates with semiparametric quantile-based estimates to construct a more general class of estimators for high expectiles ξ_{τ_n} such that $\tau_n \to 1$ and $n(1-\tau_n) \to \infty$ as $n \to \infty$. 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 $\widetilde{\xi}_{\tau_n}$ 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}:

$$\widehat{\xi}_{\tau_n} := \left(\widehat{\gamma}_{\tau_n}^{-1} - 1\right)^{-\widehat{\gamma}_{\tau_n}} \widehat{q}_{\tau_n},$$

where $\widehat{\gamma}_{\tau_n}$ is a suitable estimator of the tail index γ. We will consider in the sequel the Hill estimator (Hill [26])

$$\widehat{\gamma}_{\tau_n} = \frac{1}{\lfloor n(1-\tau_n) \rfloor} \sum_{i=1}^{\lfloor n(1-\tau_n) \rfloor} \log\left(\frac{\widehat{q}_{1-(i-1)/n}}{\widehat{q}_{1-\lfloor n(1-\tau_n) \rfloor/n}}\right), \tag{7}$$

which enjoys a high degree of popularity thanks to its simplicity and advantageous variance properties. Beirlant *et al.* [5] and de Haan and Ferreira [15] give an extensive overview of the asymptotic theory for this popular estimator.

More generally, one may also combine the two estimators $\widehat{\xi}_{\tau_n}$ and $\widetilde{\xi}_{\tau_n}$ to define, for $\beta \in \mathbb{R}$, the weighted estimator

$$\overline{\xi}_{\tau_n}(\beta) := \beta \widehat{\xi}_{\tau_n} + (1 - \beta) \widetilde{\xi}_{\tau_n}.$$

The two special cases $\beta = 1$ and $\beta = 0$ correspond to the unique existing intermediate expectile estimators in the literature, namely, the estimators $\widehat{\xi}_{\tau_n}$ and $\widetilde{\xi}_{\tau_n}$ first introduced in Daouia *et al.* [13]. These were coined, respectively, "indirect estimator" and "direct estimator" to reflect the pure asymmetric least squares nature of the latter and the reliance of the former on quantiles. The limit distribution of their linear combination $\overline{\xi}_{\tau_n}(\beta)$ crucially relies on the asymptotic dependence structure between the tail expectile and quantile processes established in Theorem 1, since $\overline{\xi}_{\tau_n}(\beta)$ is built on both of these processes. More specifically, it relies on the following asymptotic dependence structure between the Hill estimator $\widehat{\gamma}_{\tau_n}$ and the intermediate sample quantile $\widehat{q}_{\tau_n}$ and expectile $\widetilde{\xi}_{\tau_n}$.

Theorem 3. *Suppose that $\mathbb{E}|Y_-|^2 < \infty$. Assume further that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1/2$. Let $\tau_n \to 1$ be such that $n(1 - \tau_n) \to \infty$, and suppose that the bias condition $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_1 \in \mathbb{R}$ is satisfied. Then,*

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

where $\mathfrak{m}$ is the 1×3 vector

$$\mathfrak{m} := \left(\frac{\lambda_1}{1 - \rho}, 0, 0\right),$$

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

$$\mathfrak{V}(1,1) = \gamma^2, \quad \mathfrak{V}(1,2) = 0, \quad \mathfrak{V}(1,3) = \frac{\gamma^3}{(1-\gamma)^2}(\gamma^{-1} - 1)^\gamma,$$

$$\mathfrak{V}(2,2) = \gamma^2, \quad \mathfrak{V}(2,3) = \gamma^2 \left(\frac{(\gamma^{-1} - 1)^\gamma}{1 - \gamma} - 1\right), \quad \mathfrak{V}(3,3) = \frac{2\gamma^3}{1 - 2\gamma}.$$

Based on the joint asymptotic normality in Theorem 3, we obtain the following limit distribution of the weighted estimator $\overline{\xi}_{\tau_n}(\beta)$.

Theorem 4. *Suppose that the conditions of Theorem 3 hold with the additional bias condition $\sqrt{n(1-\tau_n)}/q_{\tau_n} \to \lambda_2 \in \mathbb{R}$. Then, for any $\beta \in \mathbb{R}$,*

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{\xi}_{\tau_n}(\beta)}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \beta\left(\mathfrak{b} + [(1-\gamma)^{-1} - \log(\gamma^{-1}-1)]\Psi + \Theta\right) + (1-\beta)\Xi$$

where the bias component $\mathfrak{b}$ is $\mathfrak{b} = \lambda_1 \mathfrak{b}_1 + \lambda_2 \mathfrak{b}_2$ with

$$\mathfrak{b}_1 = \frac{(1-\gamma)^{-1} - \log(\gamma^{-1}-1)}{1-\rho} - \frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} - \frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho},$$
$$\mathfrak{b}_2 = -\gamma(\gamma^{-1}-1)^{\gamma}\mathbb{E}(Y),$$

and (Ψ, Θ, Ξ) is a trivariate Gaussian centered random vector with covariance matrix $\mathfrak{V}$ as in Theorem 3.

Remark 1. *The same second-order conditions involving the auxiliary function A in Theorems 3 and 4 are also required to derive the marginal asymptotic normality of the conventional Hill estimator $\hat{\gamma}_{\tau_n}$, with asymptotic bias $\lambda_1/(1-\rho)$ and asymptotic variance γ^2 [see Theorem 3.2.5 in de Haan and Ferreira ([15], p.74)]. Theorem 4 features, however, 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.*

Remark 2. *When $\beta = 1$, we recover the convergence of the 'indirect' estimator $\hat{\widetilde{\xi}}_{\tau_n}$ obtained in Corollary 2 of Daouia et al. [13]. When $\beta = 0$, we get the convergence of the 'direct' estimator $\widetilde{\xi}_{\tau_n}$ stated in Theorem 2 of [13].*

The use of the weighted estimator $\overline{\xi}_{\tau_n}(\beta)$ is, by construction, most appropriate when it comes to deal with intermediate expectile levels $\tau = \tau_n \to 1$ such that $n(1-\tau_n) \to \infty$. In the very far tails where the expectile level $\tau = \tau_n' \to 1$ is such that $n(1-\tau_n') \to c \in [0,\infty)$, this estimator becomes unstable and inconsistent due to data sparsity. To estimate an extreme expectile $\xi_{\tau_n'}$, Daouia *et al.* [13] propose to extrapolate any consistent intermediate expectile estimator, say $\widecheck{\xi}_{\tau_n}$, to the very high level τ_n' by considering the generic class of estimators

$$\widecheck{\xi}_{\tau_n'}^{\star} := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widecheck{\gamma}_n} \widecheck{\xi}_{\tau_n},$$

where $\widecheck{\gamma}_n$ is a suitable estimator of γ. Here, we choose to use the general weighted intermediate estimator $\overline{\xi}_{\tau_n}(\beta)$ in conjunction with the Hill estimator $\hat{\gamma}_{\tau_n}$ to define the following class of extreme expectile estimators:

$$\overline{\xi}_{\tau_n'}^{\star}(\beta) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\hat{\gamma}_{\tau_n}} \overline{\xi}_{\tau_n}(\beta). \tag{8}$$

The two special cases $\beta = 1$ and $\beta = 0$ correspond to the unique existing extreme expectile estimators in the literature, namely, the extrapolated indirect and direct expectile estimators suggested in Daouia *et al.* [13]. The next theorem gives the asymptotic behavior of the generalized extreme expectile estimators $\overline{\xi}^{\star}_{\tau'_n}(\beta)$.

Theorem 5. *Suppose that the conditions of Theorem 4 hold. Assume also that $\rho < 0$ and $n(1 - \tau'_n) \to c < \infty$ with $\sqrt{n(1 - \tau_n)}/\log[(1 - \tau_n)/(1 - \tau'_n)] \to \infty$. Then, for any $\beta \in \mathbb{R}$,*

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left(\frac{\overline{\xi}^{\star}_{\tau'_n}(\beta)}{\xi_{\tau'_n}} - 1 \right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1 - \rho}, \gamma^2 \right).$$

One can observe that the limiting distribution of $\overline{\xi}^{\star}_{\tau'_n}(\beta)$ is controlled by the asymptotic distribution of $\widehat{\gamma}_{\tau_n}$. In particular, in the cases $\beta = 1$ and $\beta = 0$, we exactly recover Corollaries 3 and 4 of Daouia *et al.* [13] on the convergence of the extrapolated indirect and direct expectile estimators.

4. Estimation of tail Expected Shortfall

4.1. 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 $\mathrm{VaR}_\tau(Y) := q_\tau$. 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 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 (Daníelsson *et al.* [11]). Furthermore, VaR_τ fails to be subadditive, since the inequality $\mathrm{VaR}_\tau(Y_1 + Y_2) \leqslant \mathrm{VaR}_\tau(Y_1) + \mathrm{VaR}_\tau(Y_2)$ 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])

$$\mathrm{QES}_\tau := \frac{1}{1 - \tau} \int_\tau^1 q_t \, dt. \tag{9}$$

When Y is continuous, QES_τ is identical to the Conditional Value at Risk (Rockafellar and Uryasev [36, 37]), known also as Tail Conditional Expectation (TCE), defined as $\mathrm{QTCE}_\tau := \mathbb{E}[Y | Y > q_\tau]$. 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 [43] and Acerbi and Tasche [3]).

4.2. Expectile-based Expected Shortfall

Motivated by the merits and good properties of expectiles, Taylor [39] 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 τth expectile ξ_τ of the distribution of Y. This expectile-based TCE was actually implemented by Taylor [39] only as an intermediate instrument for the ultimate interest in estimating the conventional quantile-based form QTCE_τ, or equivalently, the coherent version QES_τ under the continuity assumption on Y. Although the interpretability of XTCE_τ is straightforward, its coherence as a proper risk measure has been an open question so far. This is now elucidated below in Proposition 2, showing the failure of XTCE_τ to fulfill the coherence property in general. It would then be awkward to use this non-coherent measure as the basis for estimating the coherent quantile-based version QES_τ. Instead, by analogy to QES_τ itself, we propose to use the new expectile-based form of ES

$$\text{XES}_\tau := \frac{1}{1-\tau} \int_\tau^1 \xi_t \, dt, \tag{10}$$

obtained by substituting the expectile ξ_t in place of the quantile q_t in the standard form (9) 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 $\tau \geqslant 1/2$,*

(i) *XES_τ induces a coherent risk measure;*
(ii) *$XTCE_\tau$ 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 $\tau \geqslant 1/2$.

Next, we show under the model assumption (3) that XES_τ is asymptotically equivalent to XTCE_τ as $\tau \to 1$, and hence inherits its direct meaning as the conditional expectation $\mathbb{E}[Y|Y > \xi_\tau]$ 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*

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

Propositions 2 and 3 then afford arguments to justify that the new form XES_τ of expectile-based ES provides a better alternative to XTCE_τ not only as a proper risk

measure, but also as an intermediate tool for estimating the classical quantile-based version QES_τ. Indeed, XES_τ is coherent and keeps the intuitive meaning of XTCE_τ as a conditional expectation when $\tau \to 1$, since $\text{XES}_\tau \sim \text{XTCE}_\tau$. Most importantly, XES_τ may be adopted as a reasonable alternative to QES_τ itself. As is the case in the duality (5) between the expectile ξ_τ and the VaR q_τ, the choice in any risk analysis between the expectile-based form XES_τ and its quantile-based analogue QES_τ will then depend on the value at hand of $\gamma \lessgtr \frac{1}{2}$. More precisely, the quantity XES_τ will be more (respectively, less) extreme than QES_τ and hence more (respectively, less) pessimistic or conservative in practice, for all τ large enough, if $\gamma > \frac{1}{2}$ (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 $\text{XES}_\tau / \xi_\tau \sim (1 - \gamma)^{-1}$.

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 $\tau \to 1$,*

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

This result will prove instrumental when examining the asymptotic properties of our tail expectile-based ES estimators in the next section.

4.3. 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 $\tau'_n \to 1$ and $n(1 - \tau'_n) \to c < \infty$, 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:

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

obtained simply by replacing ξ_t in (10) with its sample version $\widetilde{\xi}_t$ described in (2). Since this estimator is a linear functional of the tail empirical expectile process, Theorem 2 is more adapted than Theorem 1 for the analysis of its asymptotic distribution.

Theorem 6. *Under the conditions of Theorem 2,*

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

On the basis of Proposition 1(ii) and then Proposition 3, we have for $\tau_n < \tau_n' \to 1$ that

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

Therefore, to estimate $\mathrm{XES}_{\tau_n'}$ at an arbitrary extreme level τ_n', we replace γ by the Hill estimator $\widehat{\gamma}_{\tau_n}$ and XES_{τ_n} at an intermediate level τ_n by the estimator $\widetilde{\mathrm{XES}}_{\tau_n}$ to get

$$\widetilde{\mathrm{XES}}_{\tau_n'}^{\star} := \left(\frac{1 - \tau_n'}{1 - \tau_n}\right)^{-\widehat{\gamma}_{\tau_n}} \widetilde{\mathrm{XES}}_{\tau_n}. \tag{11}$$

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

Theorem 7. *Assume that the conditions of Theorem 5 hold. Then*

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau_n')]} \left(\frac{\widetilde{\mathrm{XES}}_{\tau_n'}^{\star}}{\mathrm{XES}_{\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1 - \rho}, \gamma^2\right).$$

One can also design alternative options for estimating $\mathrm{XES}_{\tau_n'}$ by using the asymptotic connections in Proposition 3. For example, the asymptotic equivalence $\mathrm{XES}_{\tau_n'} \sim (1 - \gamma)^{-1} \xi_{\tau_n'}$, established therein, suggests that $\mathrm{XES}_{\tau_n'}$ can be estimated consistently by substituting the tail quantities γ and $\xi_{\tau_n'}$ with their consistent estimators $\widehat{\gamma}_{\tau_n}$ and $\overline{\xi}_{\tau_n'}^{\star}(\beta)$ described in (7) and (8), respectively. This yields the extrapolated estimator

$$\overline{\mathrm{XES}}_{\tau_n'}^{\star}(\beta) := [1 - \widehat{\gamma}_{\tau_n}]^{-1} \overline{\xi}_{\tau_n'}^{\star}(\beta), \tag{12}$$

with a weight $\beta \in \mathbb{R}$, whose asymptotic normality is established in the following.

Theorem 8. *Assume that the conditions of Theorem 5 hold. Then, for any $\beta \in \mathbb{R}$,*

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau_n')]} \left(\frac{\overline{\mathrm{XES}}_{\tau_n'}^{\star}(\beta)}{\mathrm{XES}_{\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1 - \rho}, \gamma^2\right).$$

Theorems 7 and 8 are, like Theorem 5, derived by noticing that, on the one hand, the asymptotic behaviors of $\widetilde{\mathrm{XES}}_{\tau_n'}^{\star}$ and $\overline{\xi}_{\tau_n'}^{\star}(\beta)$ are controlled by the asymptotic behavior of $\{(1 - \tau_n')/(1 - \tau_n)\}^{-\widehat{\gamma}_{\tau_n}}$, which is itself governed by that of $\widehat{\gamma}_{\tau_n}$ only. This is why, in Theorems 7 and 8, the asymptotic distributions of $\widetilde{\mathrm{XES}}_{\tau_n'}^{\star}$ and $\overline{\mathrm{XES}}_{\tau_n'}^{\star}(\beta)$ coincide. On the other hand, the nonrandom remainder term coming from the use of Proposition 3 can be controlled thanks to Proposition 4.

4.4. 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 $\mathrm{XES}_{\tau_n'}$. Since $\mathrm{XES}_{\tau_n'} \sim \mathbb{E}[Y|Y > \xi_{\tau_n'}]$, this problem translates into choosing $\xi_{\tau_n'}$ itself.

When moving from the conventional VaR q_{p_n}, for a pre-specified tail probability $p_n \to 1$ with $n(1 - p_n) \to c < \infty$, 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.* [13] 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

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

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

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

Under the model assumption of Pareto-type tails (3), it turns out that

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

The proof of this result can be found in Daouia *et al.* ([13], Proposition 3). Built on the Hill estimator $\widehat{\gamma}_{\tau_n}$ of γ, we can then define a natural estimator of $\tau'_n(p_n)$ as

$$\widehat{\tau}'_n(p_n) := 1 - (1 - p_n)\frac{\widehat{\gamma}_{\tau_n}}{1 - \widehat{\gamma}_{\tau_n}}. \tag{13}$$

By substituting this estimated value in place of $\tau'_n \equiv \tau'_n(p_n)$ in the extrapolated estimators $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}$ and $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\beta)$ described in (11) and (12), we obtain composite estimators that estimate $\mathrm{XES}_{\tau'_n(p_n)} \sim \mathrm{QES}_{p_n}$, by Proposition 3. The convergence results in Theorems 7 and 8 of the extrapolated estimators $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}$ and $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\beta)$ still hold true for their composite versions as estimators of QES_{p_n}, with the same technical conditions.

Theorem 9. *Suppose the conditions of Theorem 5 hold with p_n in place of τ'_n. Then, for any $\beta \in \mathbb{R}$,*

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - p_n)]}\left(\frac{\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}}{\mathrm{QES}_{p_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1 - \rho}, \gamma^2\right),$$

$$and \quad \frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - p_n)]}\left(\frac{\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\beta)}{\mathrm{QES}_{p_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1 - \rho}, \gamma^2\right).$$

5. Numerical simulations

In order to illustrate the behavior of the presented estimation procedures of the two expected shortfall forms $\mathrm{XES}_{\tau_n'}$ 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. The accuracy of the weighted estimators is investigated for various values of the weight $\beta \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. All the experiments have sample size $n = 500$ and tail index $\gamma \in \{0.05, 0.25, 0.45\}$. In our simulations we used the extreme levels $\tau_n' = p_n = 1 - \frac{1}{n}$ and the intermediate level $\tau_n = 1 - \frac{k}{n}$, where the integer k can be viewed as the effective sample size for tail extrapolation.

5.1. Estimates of $\mathrm{XES}_{\tau_n'}$

We evaluated the finite-sample performance of $\widetilde{\mathrm{XES}}^{\star}_{\tau_n'}$ and $\overline{\mathrm{XES}}^{\star}_{\tau_n'}(\beta)$, described in (11) and (12), as estimators of $\mathrm{XES}_{\tau_n'}$, for the different chosen scenarios and values of β.

Figures 1 and 2 give, respectively, the MSE (on a logarithmic scale) and bias estimates of $\overline{\mathrm{XES}}^{\star}_{\tau_n'}(\beta)/\mathrm{XES}_{\tau_n'}$, as functions of the sample fraction k (each curve corresponds to a value of β as indicated by the colour-scheme). The Monte-Carlo estimates obtained for $\widetilde{\mathrm{XES}}^{\star}_{\tau_n'}/\mathrm{XES}_{\tau_n'}$ are superimposed in black curves. In the case of Student distribution (top panels), the latter estimates perform clearly better in terms of both MSE and bias, for all values of γ. In the case of Fréchet distribution (panels in the middle), it may be seen that the orange curves ($\beta = 1$) behave quite well in terms of MSE and bias, for the three values of γ. In the case of Pareto distribution (bottom panels), it may be seen from left to right that the green ($\beta = 0.4$), blue ($\beta = 0.8$) and orange ($\beta = 1$) curves are superior, respectively, for $\gamma = 0.05, 0.25, 0.45$.

To summarize, $\widetilde{\mathrm{XES}}^{\star}_{\tau_n'}$ is the winner in the case of the real-valued Student distribution, while $\overline{\mathrm{XES}}^{\star}_{\tau_n'}(\beta)$ appears to be the most efficient in the case of the non-negative Fréchet and Pareto distributions. The best choice of the weight β in the case of Fréchet distribution is globally $\beta = 1$. By contrast, the best choice of β in the case of Pareto distribution seems to increase with γ.

5.2. Estimates of QES_{p_n}

We have also undertaken simulation experiments to evaluate the finite-sample performance of the composite versions $\widetilde{\mathrm{XES}}^{\star}_{\hat{\tau}_n'(p_n)}$ and $\overline{\mathrm{XES}}^{\star}_{\hat{\tau}_n'(p_n)}(\beta)$ studied in Theorem 9, with $\hat{\tau}_n'(p_n)$ being described in (13). These composite expectile-based estimators estimate the same conventional expected shortfall QES_{p_n} as the direct quantile-based estimator

$$\widehat{\mathrm{QES}}^{\star}_{p_n} := \left(\frac{1-p_n}{1-\tau_n}\right)^{-\hat{\gamma}_{\tau_n}} \frac{1}{\lfloor n(1-\tau_n)\rfloor} \sum_{i=1}^{\lfloor n(1-\tau_n)\rfloor} Y_{n-i+1,n} \tag{14}$$
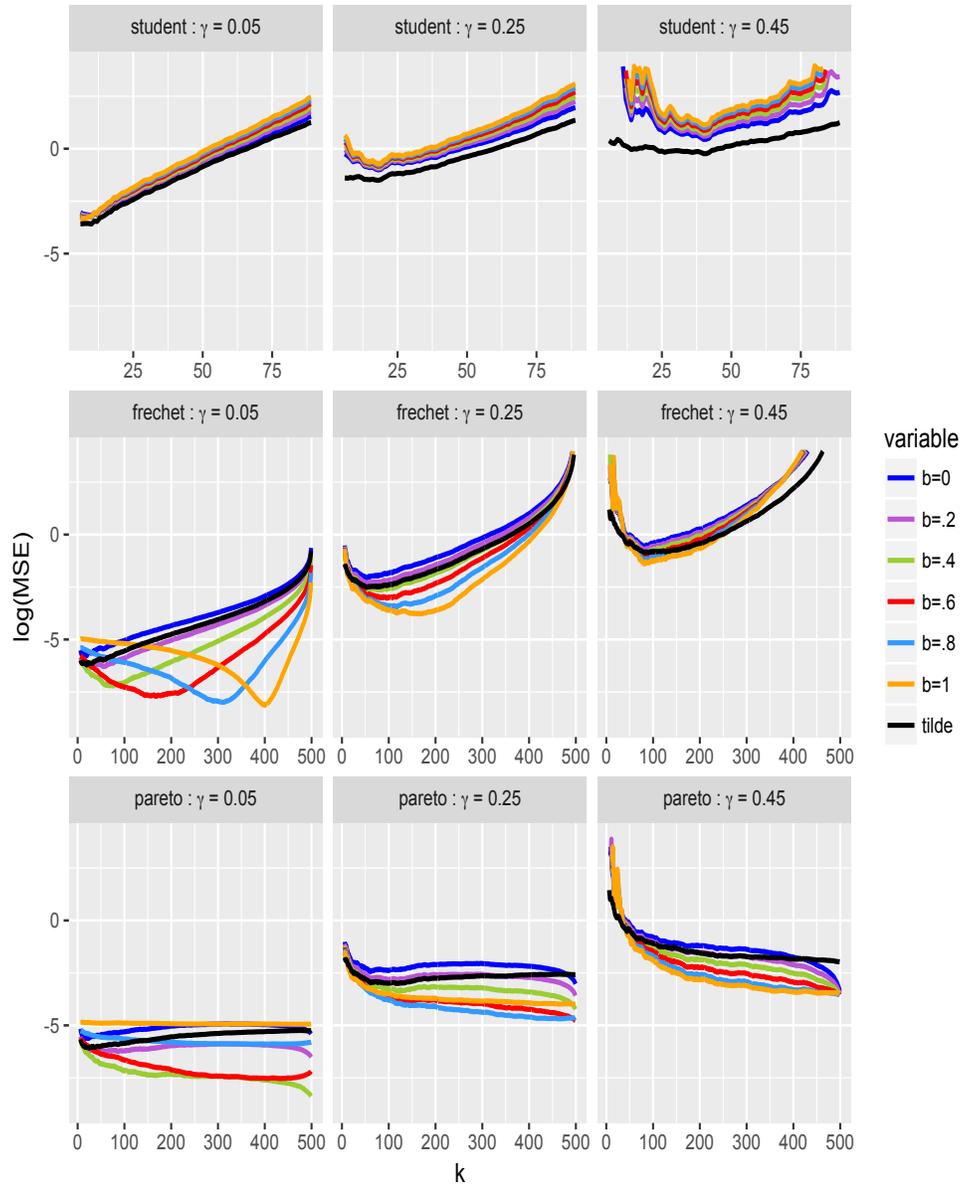
Figure 1. *MSE estimates (in log scale) of $\widetilde{XES}^{\star}_{\tau'_n}/XES_{\tau'_n}$ (black) and $\overline{XES}^{\star}_{\tau'_n}(\beta)/XES_{\tau'_n}$ (colour-scheme), against k, for Student (top), Fréchet (middle) and Pareto (bottom) distributions, with $\gamma = 0.05$ (left), $\gamma = 0.25$ (middle) and $\gamma = 0.45$ (right).*
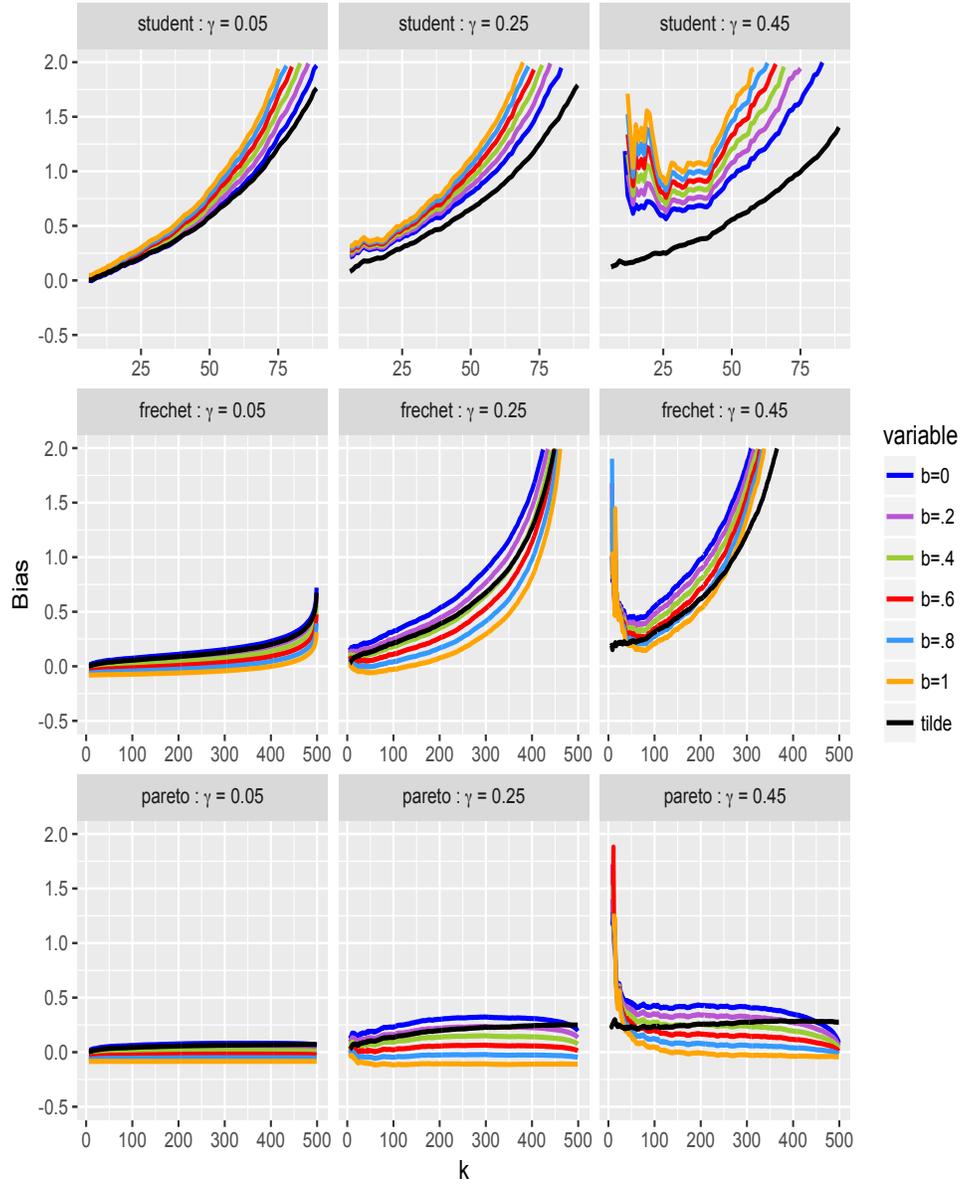
Figure 2. *Bias estimates of $\widehat{XES}^{\star}_{\tau'_n}/XES_{\tau'_n}$ (black) and $\overline{XES}^{\star}_{\tau'_n}(\beta)/XES_{\tau'_n}$ (colour-scheme).*

introduced by El Methni *et al.* [22]. To save space, all figures illustrating our simulation results here are deferred to the Supplementary Material document. In Supplement A, we arrive at the following tentative conclusions:

- In the case of the (real-valued) Student distribution, the best estimator seems to be $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$;
- In the cases of Fréchet and Pareto distributions (both positive), the best estimators seem to be $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\beta = 1)$ and/or $\widehat{\mathrm{QES}}^{\star}_{p_n}$.

6. Applications

6.1. 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 3(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 (9). 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 $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\beta = 1)$ and $\widehat{\mathrm{QES}}^{\star}_{p_n}$ 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, $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\beta = 1)$ will be denoted in the sequel by $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$.

The path of the composite expectile-based estimator $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ against the sample fraction k is shown in Figure 3(b) as rainbow curve, for the selected range of intermediate values of $k = 10, 11, \ldots, 700$. The effect of the Hill estimate $\widehat{\gamma}_{1-k/n}$ on $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ is highlighted by a colour-scheme, ranging from dark red (low $\widehat{\gamma}_{1-k/n}$) to dark violet (high $\widehat{\gamma}_{1-k/n}$). This γ estimate seems to mainly vary within the interval $[0.35, 0.36]$, which corresponds to the stable (blue-green) part of the plot over $k \in [150, 500]$. The curve $k \mapsto \overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ 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 similar to that of El Methni *et al.* [23]. This consists first in computing the standard deviations of $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ over a moving window large enough to cover 20% of the possible values of k in the selected range $10 \leqslant k \leqslant 700$. Then one selects the first window over which the standard deviation has a local minimum, and is less than the average standard deviation across all windows. The desired sample

fraction is finally defined as the value of k for which $\overline{\text{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ is the median estimate within this window. The resulting estimate $\overline{\text{XES}}^{\star}_{\widehat{\tau}'_n(p_n)} = 6.42$ million is obtained for the value $k = 222$ in the window $[119, 259]$.

The path of the direct quantile-based estimator $\widehat{\text{QES}}^{\star}_{p_n}$ against k is graphed in the same figure as dashed black curve. It is very similar to that of $\overline{\text{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$. The pointwise estimate $\widehat{\text{QES}}^{\star}_{p_n} = 6.37$ million is indicated by the minimal standard deviation achieved at $k = 222$ over the window $[119, 259]$. By Theorem 9 we have

$$\frac{\sqrt{k}}{\log[k/n(1-p_n)]} \left(\frac{\overline{\text{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}}{\text{QES}_{p_n}} - 1 \right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1-\rho}, \gamma^2 \right).$$

Under the bias condition $\lambda_1 = 0$ in Theorem 3, the asymptotic bias reduces to zero. With this condition, the (symmetric) expectile-based asymptotic confidence interval with confidence level $100\vartheta\%$ has the form $\overline{\text{CI}}_\vartheta(k) = \overline{\text{XES}}^{\star}_{\widehat{\tau}'_n(p_n)} \times \mathcal{I}$, where $\mathcal{I}$ stands for the interval

$$\mathcal{I} := \left[1 \pm z_{(1+\vartheta)/2} \, \log\left(\frac{k}{n(1-p_n)} \right) \frac{\widehat{\gamma}_{1-k/n}}{\sqrt{k}} \right], \tag{15}$$

with $z_{(1+\vartheta)/2}$ being the $(1+\vartheta)/2-$quantile of the standard Gaussian distribution. The confidence interval derived from the asymptotic normality of $\widehat{\text{QES}}^{\star}_{p_n}$ in [22] can be expressed as $\widehat{\text{CI}}_\vartheta(k) = \widehat{\text{QES}}^{\star}_{p_n} \times \mathcal{I}$.

The two asymptotic 95% confidence intervals $\overline{\text{CI}}_{0.95}(k)$ and $\widehat{\text{CI}}_{0.95}(k)$ are superimposed in Figure 3(b) as well, respectively, in dotted blue and solid grey lines. Clearly, they 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 to the maximum recorded claim amount.

Finally, we would like to comment on the estimator $\widehat{\tau}'_n(p_n)$ of the extreme expectile level $\tau'_n(p_n)$ which ensures that $\overline{\text{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ estimates $\text{XES}_{\tau'_n(p_n)} \sim \text{QES}_{p_n}$. The plot of $\widehat{\tau}'_n(p_n)$ versus k is graphed in Figure 3(c) as rainbow curve, and the corresponding optimal pointwise estimate is indicated by the horizontal dashed black line. This selected optimal level $\widehat{\tau}'_n(p_n) = 0.9999941$ is 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

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

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 Holzmann and Klar [27].

6.2. Application to financial data

In this section, we apply our method to estimate the ES for three large US financial institutions. We consider the same investment banks as in the study of Cai *et al.* [9], namely Goldman Sachs, Morgan Stanley and T. Rowe Price. All of these banks had a market capitalization greater than 5 billion USD at the end of June 2007. The dataset consists of the negative log-returns (Y_i) on their equity prices at a weekly frequency during 10 years from July 3rd, 2000, to June 30th, 2010. The choice of the frequency of data and time horizon follows the same set-up as in Cai *et al.* [9] and Daouia *et al.* [13]. Our theorems being derived under the assumption that the data $Y_1, \ldots, Y_n$ are independent and identically distributed, we use here weekly rather than daily loss returns to reduce substantially the potential serial dependence. This results in the sample size $n = 522$. We use our composite expectile-based method to estimate the standard quantile-based expected shortfall QES_{p_n}, or equivalently the expectile-based expected shortfall $\mathrm{XES}_{\tau'_n(p_n)}$, with an extreme relative frequency $p_n = 1 - \frac{1}{n}$ that corresponds to a once-per-decade rare event.

In this situation of real-valued profit-loss distributions, our experience with simulated data indicates that the composite estimator $\widetilde{\mathrm{XES}}^\star_{\widehat{\tau}'_n(p_n)}$ provides the best QES_{p_n} estimates in terms of MSE and bias. In the estimation, we employ the intermediate sequence $\tau_n = 1 - k/n$ as before, for the selected range of values $k = 1, \ldots, 80$. The confidence interval with confidence level $100\vartheta\%$ derived from the asymptotic normality of $\widetilde{\mathrm{XES}}^\star_{\widehat{\tau}'_n(p_n)}$, in Theorem 9, has the form $\widetilde{\mathrm{CI}}_\vartheta(k) = \widetilde{\mathrm{XES}}^\star_{\widehat{\tau}'_n(p_n)} \times \mathcal{I}$, where $\mathcal{I}$ is described in (15). For our comparison purposes, we use as a benchmark the direct quantile-based estimator $\widehat{\mathrm{QES}}^\star_{p_n}$ of El Methni *et al.* [22], along with the corresponding asymptotic confidence interval $\widehat{\mathrm{CI}}_\vartheta(k)$.

For each bank, we superimpose in Figure 4 the plots of the two competing estimates $\widetilde{\mathrm{XES}}^\star_{\widehat{\tau}'_n(p_n)}$ and $\widehat{\mathrm{QES}}^\star_{p_n}$ against k, as rainbow and dashed black curves respectively, along with their associated 95% confidence intervals $\widetilde{\mathrm{CI}}_{0.95}(k)$ in dotted blue lines and $\widehat{\mathrm{CI}}_{0.95}(k)$ in solid grey lines. The effect of the Hill estimate $\widehat{\gamma}_{1-k/n}$ on $\widetilde{\mathrm{XES}}^\star_{\widehat{\tau}'_n(p_n)}$ is highlighted by a colour-scheme, ranging from dark red (low $\widehat{\gamma}_{1-k/n}$) to dark violet (high $\widehat{\gamma}_{1-k/n}$).

We have already provided Monte Carlo evidence that the composite expectile-based estimator $\widetilde{\mathrm{XES}}^\star_{\widehat{\tau}'_n(p_n)}$ is efficient relative to the pure quantile-based estimator $\widehat{\mathrm{QES}}^\star_{p_n}$. Its superiority in terms of plots' stability, including confidence intervals, can clearly be visualized in Figure 4 for the three banks. The final ES levels based on the selection criterion of Section 6.1, computed over a moving window covering 20% of the possible values of k, are reported in Table 1, along with the asymptotic 95% confidence intervals of the ES. Based on the $\widetilde{\mathrm{XES}}^\star_{\widehat{\tau}'_n(p_n)}$ estimates (in the second column), the ES levels for Goldman Sachs and T. Rowe Price seem to be close (around -38% to -44%), whereas the ES level for Morgan Stanley is almost twice higher (around -81%). It is worth noticing that the difference between the $\widetilde{\mathrm{XES}}^\star_{\widehat{\tau}'_n(p_n)}$ levels for Goldman Sachs and T. Rowe Price is very close to the difference between their respective maxima $Y_{n,n}$. The $\widehat{\mathrm{QES}}^\star_{p_n}$ estimates

(in the fourth column) point towards slightly more pessimistic risk measures for the three
banks.

Bank	$\widehat{\text{XES}}^{\star}_{\hat{\tau}'_n(p_n)}$	$\widetilde{\text{CI}}_{0.95}$	$\widehat{\text{QES}}^{\star}_{p_n}$	$\widehat{\text{CI}}_{0.95}$	$Y_{n,n}$
Goldman Sachs	0.445	(0.194, 0.620)	0.495	(0.226, 0.680)	0.365
Morgan Stanley	0.817	(0.384, 1.305)	0.883	(0.366, 1.478)	0.904
T. Rowe Price	0.386	(0.213, 0.511)	0.407	(0.216, 0.548)	0.305

Table 1. *ES levels of the three investment banks, with the 95% confidence intervals and the sample maxima. Results based on weekly loss returns, with $n = 522$ and $p_n = 1 - \frac{1}{n}$.*

Our two applications with real data seem to indicate that the more accurate composite
expectile-based estimates tend to be less (respectively, more) conservative than the pure
quantile-based estimates $\widehat{\text{QES}}^{\star}_{p_n}$ in the case of real-valued profit-loss (respectively, non-
negative loss) distributions.

7. Concluding remarks

Originally introduced in Newey and Powell [34], expectiles have found recently an increas-
ing usage in finance and actuarial science as alternative instruments of risk protection to
quantiles. Their estimation *via* the method of asymmetric least squares is still in full de-
velopment in the areas of risk management and extreme value statistics. Our general joint
theory for tail empirical expectile and quantile processes opens new horizons for a wide
variety of tail risk estimation problems. This is illustrated through a fruitful estimation
of Expected Shortfall (ES), based on general weighted combinations of both top order
statistics and high expectiles. Interestingly, pure asymmetric least squares estimators are
particularly advantageous in the case of real-valued profit-loss distributions.

In our motivating application we focus on both quantile and expectile-based forms
of ES, but our weighted approximation theorems can be applied to other complex risk
functionals of both expectile and quantile processes. This includes the wider class of co-
herent spectral risk measures (Acerbi [2]), but also the very recent concept of *extremiles*
(Daouia *et al.* [12]). The latter concept defines a new least squares analogue of quantiles,
which is motivated via several angles and includes the family of expected minima and
expected maxima. Its characterization as weighted average of all quantiles, as well as its
specific merits and strengths, raise the question of extending this concept by replacing
quantiles with expectiles. The class of Wang [41] distortion risk measures with concave
distortion functions is another concrete example of genuine interest for future research.
The formulation and estimation of extreme versions of these coherent risk measures, tack-
led for instance in Vandewalle and Beirlant [40] and El Methni and Stupfler [23, 24], can
also be adapted and extended by substituting expectiles in place of quantiles and apply-
ing our general theory. Yet another example of generalized quantile-based risk measures
where similar considerations may be relevant is the family of so-called L^p-quantile risk
measures studied recently from the perspective of extreme values in Daouia *et al.* [14].

Supplementary Material

The supplement to this article contains simulation results along with technical lemmas and the proofs of all our theoretical results.

Acknowledgements

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] Csörgő, M. and Horváth, L. (1987). Approximation of intermediate quantile processes, *J. Multivariate Anal.*, **21**, 250–262.

[11] Daníelsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, 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.

[12] Daouia, A., Gijbels, I. and Stupfler, G. (2018). Extremiles: A new perspective on asymmetric least squares, *J. Amer. Statist. Assoc.*, to appear. Available at https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1498348.

[13] 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.

[14] Daouia, A., Girard, S. and Stupfler, G. (2019). Extreme M-quantiles as risk measures: From L^1 to L^p optimization, *Bernoulli*, **25**, 264–309.

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

[16] 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.

[17] Dietrich, D., de Haan, L. and Hüsler, J. (2002). Testing extreme value conditions, *Extremes*, **5**, 71–85.

[18] Drees, H. (1998). On smooth statistical tail functionals, *Scand. J. Stat.*, **25**, 187–210.

[19] 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.

[20] Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations, and forecast rankings (with discussion), *J. R. Stat. Soc. Ser. B*, **78**, 505–562.

[21] Einmahl, J.H.J. and Mason, D.M. (1988). Strong limit theorems for weighted quantile processes, *Ann. Probab.*, **16**, 1623–1643.

[22] 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.

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

[24] El Methni, J. and Stupfler, G. (2018). Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions, *Econometrics and Statistics*, **6**, 129–148.

[25] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer.

[26] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**, 1163–1174.

[27] Holzmann, H. and Klar, B. (2016). Expectile asymptotics, *Electron. J. Stat.*, **10**, 2355–2371.

[28] Hüsler, J. and Li, D. (2006). On testing extreme value conditions, *Extremes*, **9**, 69–86.

[29] Koenker, R. and Bassett, G.S. (1978). Regression quantiles, *Econometrica*, **46**, 33–50.

[30] Krätschmer, V. and Zähle, H. (2017). Statistical inference for expectile-based risk measures, *Scand. J. Stat.*, **44**, 425–454.

[31] 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.

[32] 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.

[33] Mao, T. and Yang, F. (2015). Risk concentration based on expectiles for extreme risks under FGM copula, *Insurance Math. Econom.*, **64**, 429–439.

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

[35] Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer.

[36] Rockafellar, R.T. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk, *Journal of Risk*, **2**, 21–42.

[37] Rockafellar, R.T. and Uryasev, S. (2002). Conditional Value-at-Risk for General Loss Distributions, *Journal of Banking and Finance*, **26**, 1443–1471.

[38] Sobotka, F. and Kneib, T. (2012). Geoadditive expectile regression, *Comput. Statist. Data Anal.*, **56**, 755–767.

[39] Taylor, J. (2008). Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics*, **6**, 231–252.

[40] Vandewalle, B. and Beirlant, J. (2006). On univariate extreme value statistics and the estimation of reinsurance premiums, *Insurance Math. Econom.*, **38**, 441–459.

[41] Wang, S.S. (1996). Premium calculation by transforming the layer premium density, *ASTIN Bull.*, **26**, 71–92.

[42] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.

[43] Wirch, J.L. and Hardy, M.R. (1999). A synthesis of risk measures for capital adequacy, *Insurance Math. Econom.*, **25**, 337–347.

[44] Ziegel, J.F. (2016). Coherence and elicitability, *Math. Finance*, **26**, 901–918.

[45] Zwingmann, T. and Holzmann, H. (2016). Asymptotics for the expected shortfall, preprint.
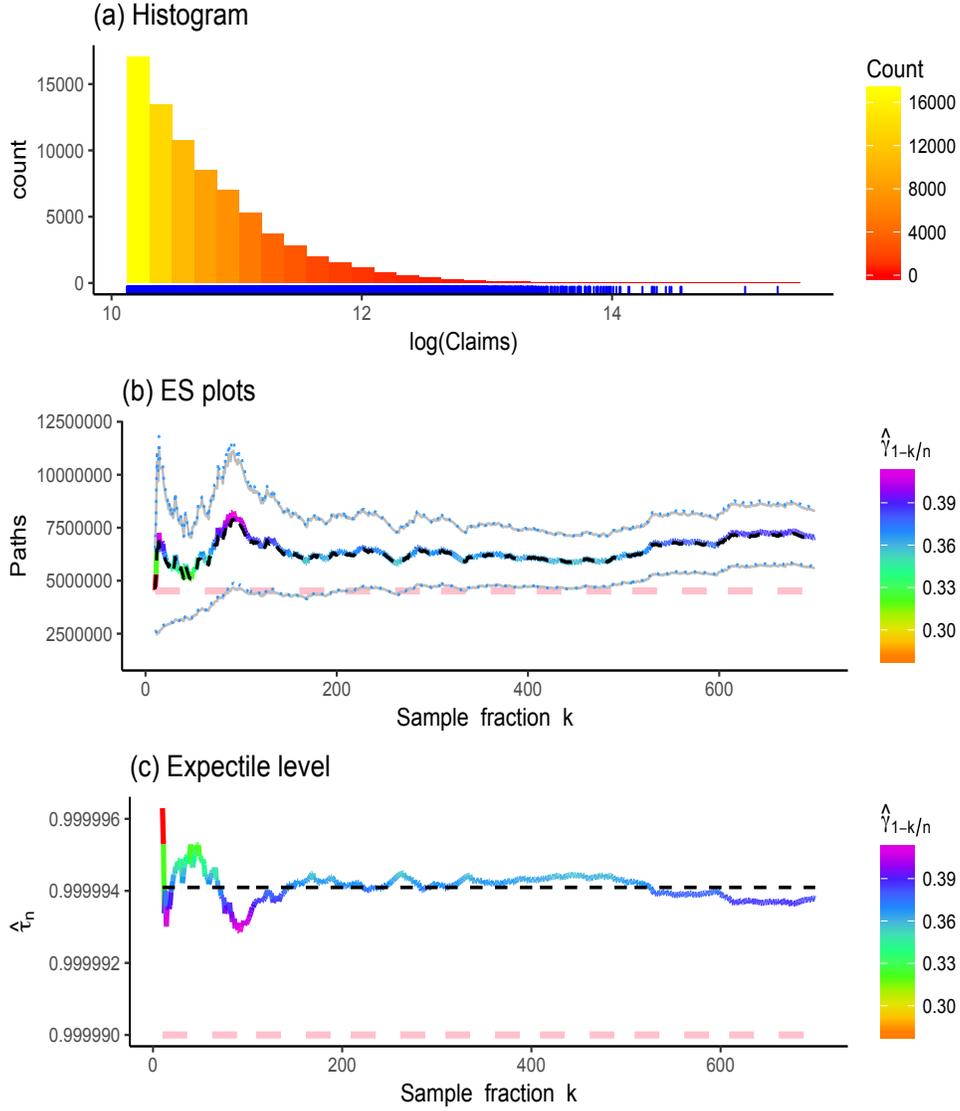
Figure 3. *(a) Scatterplot and histogram of the log-claim amounts. (b) The ES plots $k \mapsto \overline{XES}^{\star}_{\hat{\tau}'_n(p_n)}(\beta = 1)$ as rainbow curve, and $k \mapsto \widehat{QES}^{\star}_{p_n}$ in dashed black, along with the constant sample maximum $Y_{n,n}$ in horizontal dashed pink. 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.*
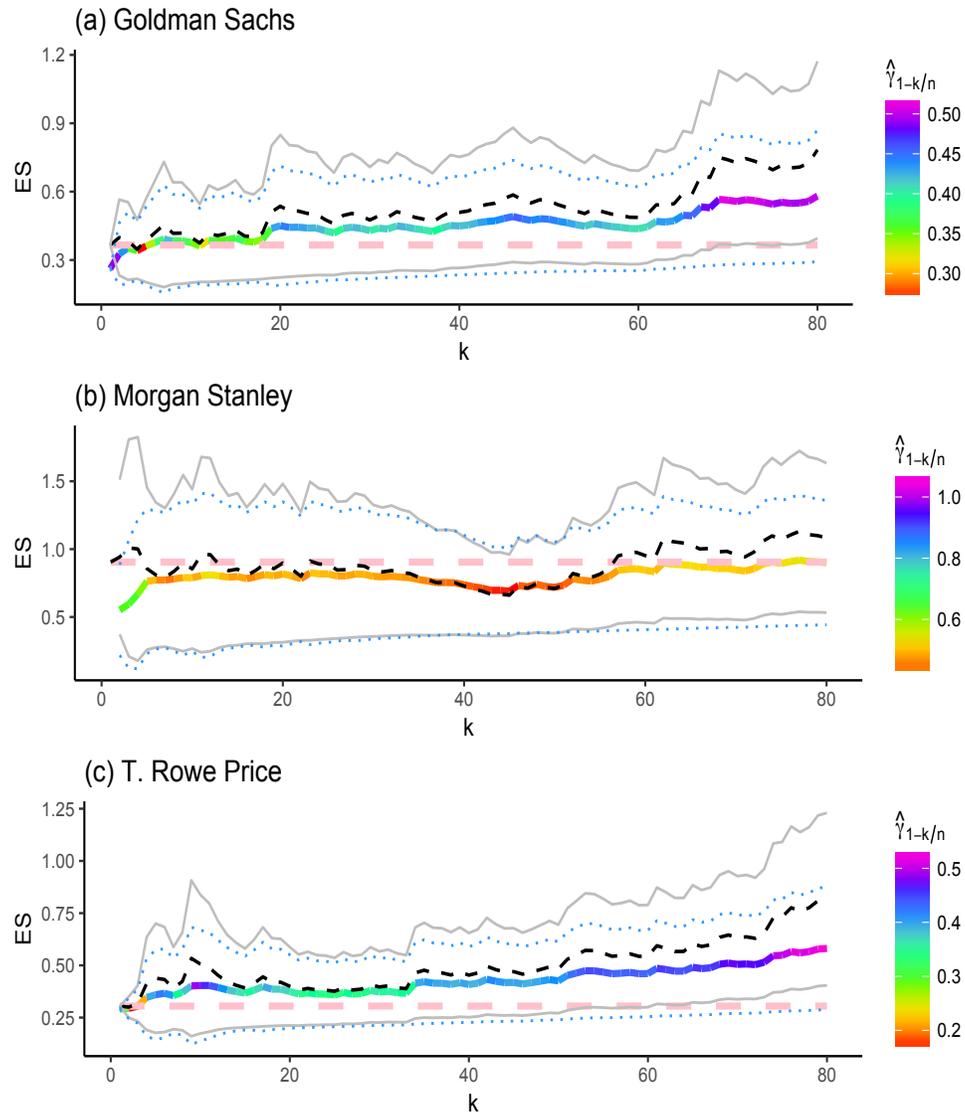
Figure 4. *Results based on weekly loss returns of the three investment banks: (a) Goldman Sachs, (b) Morgan Stanley, and (c) T. Rowe Price, with $n = 522$ and $p_n = 1 - 1/n$. The estimates $\widetilde{XES}^{\star}_{\hat{\tau}'_n(p_n)}$ as rainbow curve and $\widehat{QES}^{\star}_{p_n}$ as dashed black curve, along with the asymptotic 95% confidence intervals $\widetilde{CI}_{0.95}(k)$ in dotted blue lines and $\widehat{CI}_{0.95}(k)$ in solid grey lines. The sample maximum $Y_{n,n}$ indicated in horizontal dashed pink line.*