ExpectHill estimation, extreme risk and heavy tails

Abdelaati Daouia[a], Stéphane Girard[b] and Gilles Stupfler[c]

[a] Toulouse School of Economics, University of Toulouse Capitole, France

[b] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France

[c] Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

Abstract: Risk measures of a financial position are, from an empirical point of view, mainly based on quantiles. Replacing quantiles with their least squares analogues, called expectiles, has recently received increasing attention. The novel expectile-based risk measures satisfy all coherence requirements. We revisit their extreme value estimation for heavy-tailed distributions. First, we estimate the underlying tail index via weighted combinations of top order statistics and asymmetric least squares estimates. The resulting expectHill estimators are then used as the basis for estimating tail expectiles and Expected Shortfall. The asymptotic theory of the proposed estimators is provided, along with numerical simulations and applications to actuarial and financial data.

JEL Classifications: C13, C14.

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

1 Introduction

The risk of a financial position Y is usually summarized by a risk measure. Value at Risk (VaR) is arguably the most common risk measure used in practice. The VaR at probability level $\tau \in (0,1)$ is given by the τ-quantile $q_\tau := F_Y^{\leftarrow}(\tau) = \inf\{y \in \mathbb{R} : F(y) \geqslant \tau\}$, where F is the distribution function of Y. Koenker and Bassett [35] elaborated an absolute error loss minimization framework extending this definition of quantiles as left continuous inverse functions to the minimizers

$$q_\tau \in \underset{\theta \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left\{\varrho_\tau(Y-\theta) - \varrho_\tau(Y)\right\},$$

with equality if F is increasing, where $\varrho_\tau(y) = |\tau - 1\!\!1(y \leqslant 0)|\,|y|$ and $1\!\!1(\cdot)$ is the indicator function. There are different sign conventions for VaR which co-exist in the literature. In this paper, the position 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. In actuarial science where Y is typically a non-negative loss variable, the sign convention we have chosen implies that extreme losses also correspond to levels τ close to one. The position Y is therefore considered riskier as its risk measure gets higher.

One of the major criticisms on the VaR q_τ is its failure to fulfill the subadditivity property in general (Acerbi [1]), and hence it is not a coherent risk measure according to the axiomatic foundations in Artzner *et al.* [2]. Furthermore, it fails to account for the size of losses beyond the level τ, since quantiles only depend on the frequency of tail losses and not on their values (Daníelsson *et al.* [12]). In both of these aspects, expectiles are a perfectly reasonable alternative to quantiles as they depend on both the tail realizations and their probability (Kuan *et al.* [37]) and define a coherent risk measure for $\tau \geqslant \frac{1}{2}$ (Bellini *et al.* [5]). This is mainly due to their conception as a least squares analogue of quantiles. More precisely, by substituting the absolute deviations in the asymmetric loss function ϱ_τ with squared deviations, Newey and Powell [39] obtain the τth expectile of the distribution of Y as the minimizer

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

with $\eta_\tau(y) = |\tau - \mathbb{1}(y \leqslant 0)|\, y^2$. The additional term $\eta_\tau(Y)$ ensures the existence of a unique solution ξ_τ for distributions with finite absolute first moment. Expectiles are determined by tail expectations rather than tail probabilities, which allows for more prudent and reactive risk management. Altering the shape of extreme losses may not change the quantile-VaR, but it does impact all the expectiles (Taylor [45]). Another advantage of expectiles is that they make more efficient use of the available data since they rely on the distance to all observations and not only on the frequency of tail losses (Sobotka and Kneib [44]). Moreover, using expectiles has the appeal of avoiding recourse to regularity conditions on the underlying distribution (see *e.g.* Holzmann and Klar [32], Krätschmer and Zähle [36]). Perhaps most importantly, expectiles induce the only coherent law-invariant risk measure that is elicitable (Ziegel [49]). The property of elicitability corresponds to the existence of a natural backtesting methodology. Also, expectiles are the only M-quantiles (Breckling and Chambers [7]) that are coherent risk measures (Bellini *et al.* [5]). Further theoretical and numerical merits in favor of the adoption of expectiles in risk management can be found in Ehm *et al.* [21] and Bellini and Di Bernardino [6].

An important alternative to the VaR q_τ and its coherent least squares analogue ξ_τ is

Expected Shortfall (ES). It is favored by practitioners who are more concerned with the risk exposure to a catastrophic event that may wipe out an investment in terms of the size of potential losses. The conventional quantile-based ES at level τ equals

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

It is coherent (Acerbi [1]) and identical, when the financial position Y is continuous, to the so-called Conditional Value at Risk $\mathbb{E}[Y|Y > q_\tau]$ (Rockafellar and Uryasev [42, 43]). Similarly to this intuitive tail conditional expectation, Taylor [45] has introduced and used the expectile-based form $\mathbb{E}[Y|Y > \xi_\tau]$ as the basis for estimating the standard quantile-based measure $\mathbb{E}[Y|Y > q_\tau]$. Given that both conditional expectations $\mathbb{E}[Y|Y > q_\tau]$ and $\mathbb{E}[Y|Y > \xi_\tau]$ are not coherent risk measures in general, Daouia *et al.* [15] have suggested to estimate the coherent ES form QES_τ on the basis of its expectile-based analogue

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

obtained by substituting the expectile ξ_t in place of the quantile q_t in QES_τ. This definition is more convenient than $\mathbb{E}[Y|Y > \xi_\tau]$ as it induces a proper coherent risk measure (see Proposition 2 in Daouia *et al.* [15]).

And yet, despite this substantial body of work on expectiles and their inference, the problem of estimating tail expectiles from the perspective of extreme value theory has been much less addressed. This translates into considering both *intermediate* and *extreme* asymmetry levels, respectively, $\tau = \tau_n \to 1$ such that $n(1-\tau_n) \to \infty$ and $\tau = \tau_n' \to 1$ such that $n(1-\tau_n') \to c < \infty$, as $n \to \infty$. An appropriate distributional context is the Fréchet maximum domain of attraction of heavy-tailed distributions that describe the tail structure of most actuarial and financial data fairly well (see, *e.g.*, Embrechts *et al.* [25] and Resnick [40]). This problem is, in comparison to extreme quantile estimation, still in full development. The absence of a closed form expression for expectiles makes the extreme value analysis of their asymmetric least squares estimators a much harder mathematical problem than for order statistics. Yet, we have initiated a satisfactory solution to this problem in Daouia *et al.* [13] by proposing intermediate and extreme expectile estimators and developing their asymptotic theory. Very recently, we have come up in Daouia *et al.* [15] with powerful approximations of the tail empirical expectile process. First, Theorem 1 in Daouia *et al.* [15] derives an explicit joint asymptotic Gaussian representation of the tail expectile and quantile processes. Second, Theorem 2 in Daouia *et al.* [15] unravels the discrepancy between the tail empirical expectile process and its population counterpart. As these two theorems constitute the basic theoretical tools for our asymptotic analysis in the present paper, they are briefly described below in Proposition 1 along with the statistical model in Section 2.

Let us now highlight the contribution of this paper, which is threefold. First, building on Proposition 1, Section 3 shows that the tail index of the underlying Pareto-type distribution can be estimated in a novel manner. This index tunes the tail heaviness of F and its knowledge is of utmost interest since it makes the estimation of extreme quantiles and expectiles possible by means of appropriate extrapolation techniques. We first construct asymmetric least squares estimators of the tail index and derive their asymptotic normality in Theorem 1. We then construct a more general class of weighted estimators by computing a linear combination of these pure expectile-based estimators and of the popular Hill estimator (Hill [31]). This inspired the name *expectHill* estimators for this class. Thanks to the joint weighted Gaussian approximations of the tail expectile and quantile processes in Proposition 1, we prove the asymptotic normality of the expectHill estimators and derive their joint convergence with both intermediate quantile and expectile estimators in Theorem 2.

Second, building 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 3, follows as a corollary of Theorem 2. The weighted intermediate expectile estimators are then extrapolated to the very extreme expectile level τ_n' that may approach one at an arbitrarily fast rate. The asymptotic properties of the extrapolated $\xi_{\tau_n'}$ estimators are established in Theorem 4.

Third, we note that the proposed estimation procedures in Daouia *et al.* [15] for both extreme values $\text{XES}_{\tau_n'}$ and $\text{QES}_{\tau_n'}$ are mainly based on the classical Hill estimator of the tail index. In Section 5, we extend their extrapolation devices by using the generalized weighted expectHill estimator; see Theorems 5-6.

In Section 6, we discuss the important issue of parameter selection in our weighted estimators. Section 7 contains our experiments with simulated data and Section 8 presents a concrete application to financial returns data. Section 9 concludes. Proofs, auxiliary results and additional simulation results are deferred to the Supplementary Material document.

2 Statistical model and basic tools

In this paper we consider the class of heavy-tailed distributions, referred to as the Fréchet maximum domain of attraction, with tail index $0 < \gamma < 1$. The survival function of these

Pareto-type distributions has the form

$$\overline{F}(y) := 1 - F(y) = y^{-1/\gamma}\ell(y), \tag{2}$$

for $y > 0$ large enough, where ℓ is a slowly varying function at infinity, *i.e.*, a positive function on $(0, \infty)$ satisfying $\ell(ty)/\ell(t) \to 1$, as $t \to \infty$, for any $y > 0$. The index γ tunes the tail heaviness of $\overline{F}$: the larger the index, the heavier the right tail. Let Y be the actuarial or financial position of interest having survival function $\overline{F}$, and let $Y_- = \min(Y, 0)$ denote the negative part of Y. Then, together with condition $\mathbb{E}|Y_-| < \infty$, the assumption $\gamma < 1$ ensures the existence of the first moment of Y, and hence the existence of expectiles. By Corollary 1.2.10 in de Haan and Ferreira [16], the model assumption (2) is equivalent to

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma \quad \text{for all} \quad x > 0, \tag{3}$$

where $U(t) := q_{1-t^{-1}} \equiv \inf\{y \in \mathbb{R} \,:\, 1/\overline{F}(y) \geqslant t\}$ stands for the tail quantile function of Y. Under (2) or equivalently (3), it has been found that

$$\frac{\xi_\tau}{q_\tau} \sim (\gamma^{-1} - 1)^{-\gamma} \quad \text{as} \quad \tau \to 1 \tag{4}$$

(Bellini and Di Bernardino [6]). A refined asymptotic expansion of ξ_τ/q_τ with a precise quantification of the error term is obtained in Mao *et al.* [38] under the following second-order regular variation condition:

$\boldsymbol{\mathcal{C}_2(\gamma, \rho, A)}$ 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 $\rho \leqslant 0$ is a constant parameter and A is an auxiliary function converging to 0 at infinity and having ultimately constant sign. Hereafter, $(x^\rho - 1)/\rho$ is to be understood as $\log x$ when $\rho = 0$.

Assumption $\mathcal{C}_2(\gamma, \rho, A)$ is a standard condition in extreme value theory, which controls the rate of convergence in (3). The monographs of Beirlant *et al.* [4, Section 3.3 and particularly 3.3.2, p.93] and de Haan and Ferreira [16, Section 2.3, p.43] give abundant examples of commonly used continuous distributions satisfying $\mathcal{C}_2(\gamma, \rho, A)$, along with thorough discussions on the interpretation and the rationale behind this second-order condition. For instance, the (Generalized) Pareto, Burr, Fréchet, Student, Fisher and Inverse-Gamma distributions all satisfy this condition, and more generally so does any distribution whose survival function has the form

$$1 - F(x) = x^{-1/\gamma} \left(a + bx^{-c} + \mathrm{o}(x^{-c}) \right) \quad \text{as} \quad x \to \infty,$$

where $a > 0$, $b \in \mathbb{R}\backslash\{0\}$ and $c > 0$ are constants. This contains in particular the Hall-Weiss class of models (see Hua and Joe [33]), where condition $\mathcal{C}_2(\gamma, \rho, A)$ is met with $\rho = -c\gamma$ and $A(t) = -a^{-c\gamma-1} bc\gamma^2 t^{-c\gamma}$.

Suppose we observe independent copies $\{Y_1, \ldots, Y_n\}$ of the random variable Y and denote by $Y_{1,n} \leqslant Y_{2,n} \leqslant \cdots \leqslant Y_{n,n}$ their nth order statistics. Let the expectile level $\tau = \tau_n$ approach one at an *intermediate* rate in the sense that $n(1 - \tau_n) \to \infty$ as $n \to \infty$. A natural estimator of the corresponding intermediate expectile ξ_{τ_n} is given by its empirical version

$$\widetilde{\xi}_{\tau_n} = \arg\min_{u\in\mathbb{R}} \sum_{i=1}^{n} \eta_{\tau_n}(Y_i - u), \tag{5}$$

where $\eta_\tau(y) = |\tau - \mathbb{1}(y \leqslant 0)|\, y^2$. Under condition $\mathcal{C}_2(\gamma, \rho, A)$, Daouia *et al.* [15] prove in their Theorem 1 that the tail empirical expectile process

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

can be approximated by a sequence of Gaussian processes with drift and derive its joint asymptotic behavior with the tail empirical quantile process

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

where $\lfloor\cdot\rfloor$ stands for the floor function. They also analyze in their Theorem 2 the difference between the tail empirical expectile process and its population counterpart. For our purposes below, we recall these two approximations in the following result.

Proposition 1 (Daouia *et al.*, 2020)**.** *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 and for appropriate versions of the* *processes involved,*

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

If in addition $\rho<0$, then

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

The assumptions that $\gamma\in(0,1/2)$ and $\mathbb{E}|Y_-|^2<\infty$ essentially guarantee that the loss variable has a finite variance. This is the case in most studies on actuarial and financial data where the estimated values of γ have been found to lie below $1/2$; see, *e.g.*, the R package `CASdatasets`, Daouia *et al.* [13] and the references therein.

The extra condition $\rho<0$, in the second part of Proposition 1, is required in most extrapolation results formulated in the extreme value literature under condition $\mathcal{C}_2(\gamma, \rho, A)$; see, *e.g.*, Chapter 4 of de Haan and Ferreira [16] regarding extreme quantile estimation and Daouia *et al.* [13] for extreme expectile estimation. Note also that, in contrast to the first part of Proposition 1, the second part avoids the error terms that are proportional to $1/q_{\tau_n}$ and $A((1-\tau_n)^{-1})$.

This result, already proved in Daouia *et al.* [15], constitutes the main intermediate theoretical tool for our ultimate interest in constructing general weighted estimators of the tail index and extreme expectiles, as well as of Expected Shortfall risk measures.

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 which suggests the following approximation:

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

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

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

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

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

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

$$\widetilde{\gamma}_{\tau_n} := \frac{1}{\lfloor n(1-\tau_n)\rfloor} \sum_{i=1}^{\lfloor n(1-\tau_n)\rfloor} \log\left(\frac{\widetilde{\xi}_{1-(i-1)/n}}{\widetilde{\xi}_{1-\lfloor n(1-\tau_n)\rfloor/n}}\right) \tag{6}$$

or, equivalently, $\widetilde{\gamma}_{\tau_n} = \widetilde{\gamma}_{1-\lfloor n(1-\tau_n)\rfloor/n,\lfloor n(1-\tau_n)\rfloor}$. This simple estimator has exactly the same form as the popular Hill estimator (Hill [31])

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

with the tail empirical quantile process $\widehat{q}$ in (7) replaced by its asymmetric least squares analogue $\widetilde{\xi}$. Beirlant *et al.* [4] and de Haan and Ferreira [16] provide an extensive overview of the asymptotic theory for the Hill estimator $\widehat{\gamma}_{\tau_n}$. The next theorem gives the asymptotic normality of the three new estimators $\breve{\gamma}_{\tau_n}$, $\widetilde{\gamma}_{\tau_n,l}$ and $\widetilde{\gamma}_{\tau_n}$. Its proof essentially consists in writing

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

before integrating and crucially using Proposition 1 twice in order to control both of the logarithms on the right-hand side.

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 suppose that the bias conditions $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}/q_{\tau_n} \to \lambda_2 \in \mathbb{R}$ are satisfied. Then:*

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

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

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

Before using the estimator $\widetilde{\gamma}_{\tau_n}$ 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 1 are also required to derive the asymptotic normality of the conventional Hill estimator $\widehat{\gamma}_{\tau_n}$ in (7), with asymptotic bias $\lambda_1/(1-\rho)$ and asymptotic variance γ^2 [see Theorem 3.2.5 in de Haan and Ferreira ([16], p.74)]. Theorem 1 also features a further bias condition involving the quantile function q; this was to be expected in view of Proposition 1, 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. 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 [30]. A usual practice for selecting a reasonable estimate $\widehat{\gamma}_{\tau_n}$ is, in the reparametrization $\tau_n = 1 - k/n$, to plot the graph of $k \mapsto \widehat{\gamma}_{1-k/n}$ for $k \in \{1, 2, \ldots, n-1\}$, 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 ([16], Section 3)]. There have been a number of attempts at formalizing this procedure, including Resnick and Stărică [41], Drees *et al.* [20], and more recently El Methni and Stupfler [23, 24]. 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 $\widetilde{\gamma}_{1-k/n}$ in (6) is, in contrast to $\widehat{\gamma}_{1-k/n}$, based on expectiles that enjoy superior regularity properties compared to quantiles (see Proposition 1 in Holzmann and Klar [32]). One may thus expect that $\widetilde{\gamma}_{1-k/n}$ affords smoother and more stable plots compared to those of the Hill estimator $\widehat{\gamma}_{1-k/n}$. This advantage is illustrated in Section A of the Supplementary Material document, where we examine the properties of $\widehat{\gamma}$ and $\widetilde{\gamma}$ on real financial data. It can be seen thereon that the plots of $k \mapsto \widetilde{\gamma}_{1-k/n}$ are indeed far smoother than the arguably wiggly plots of $k \mapsto \widehat{\gamma}_{1-k/n}$.

It could, however, happen that $\widetilde{\gamma}$ 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. A natural way to take advantage of the desirable properties of both $\widetilde{\gamma}$ and $\widehat{\gamma}$ 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

$$\overline{\gamma}_{\tau_n}(\alpha) := \alpha\widehat{\gamma}_{\tau_n} + (1-\alpha)\widetilde{\gamma}_{\tau_n}. \tag{8}$$

We shall call this linear combination the *expectHill estimator*. For example, the simple mean $\overline{\gamma}_{\tau_n}(1/2)$ would represent an equal balance between the use of large asymmetric least squares statistics in (6) and top order statistics in (7). 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 Proposition 1, 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 2. *Suppose that the conditions of Theorem 1 hold. Then, for any $\alpha \in \mathbb{R}$,*

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

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

$$b_\alpha = \frac{\lambda_1}{1-\rho}\left(\alpha + (1-\alpha)\frac{(1-\gamma)(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}\right) - (1-\alpha)\mathbb{E}(Y)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2, \tag{9}$$

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

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

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

$$\sqrt{n(1-\tau_n)}\left(\overline{\gamma}_{\tau_n}(\alpha) - \gamma\right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha) \quad \text{where} \quad v_\alpha = \mathfrak{V}_\alpha(1,1). \tag{10}$$

This remains valid if $\widetilde{\gamma}_{\tau_n}$ is replaced in (8) by the continuous version $\breve{\gamma}_{\tau_n}$, or any other discretized version $\widetilde{\gamma}_{\tau_n,l}$ provided $\sqrt{n(1-\tau_n)}\log(n(1-\tau_n))/l \to 0$.

In this situation where the estimator $\overline{\gamma}_{\tau_n}(\alpha)$ depends on a weighting parameter $\alpha \in \mathbb{R}$, a reasonable question is to seek the value(s) (if any) of the parameter α giving in some sense the "best" performing estimator in the class $(\overline{\gamma}_{\tau_n}(\alpha))_{\alpha\in\mathbb{R}}$. A standard measure of the quality of the estimator is the Asymptotic Mean-Squared Error (AMSE). Minimizing this quantity for the estimator $\overline{\gamma}_{\tau_n}(\alpha)$ would amount to minimizing the quantity $b_\alpha^2 + v_\alpha$ with respect to α. This is a degree 2 convex polynomial in α, and therefore this minimization is theoretically completely straightforward. In practice though, computing the value of the optimal α^* minimizing this AMSE requires the knowledge of ρ, λ_1, λ_2, γ and $\mathbb{E}(Y)$. The accurate estimation of the particular quantities ρ and λ_1 is known to be difficult to implement in practice and requires involved methodologies, see *e.g.* the Introduction of [8]. In contrast to the sum $b_\alpha^2 + v_\alpha$, the calculation of the single asymptotic variance term v_α, which also defines a degree 2 convex polynomial in α, requires only the estimation of the parameter γ for which we can simply plug in, for instance, the Hill estimator already in use. Focusing on the minimization of the variance term v_α and ignoring the bias term b_α may therefore be a plausible pragmatic strategy. We expand upon this choice of the parameter α in our next remark.

Remark 3. The value of the weighting coefficient α in (8) which minimizes the asymptotic variance v_α of $\overline{\gamma}_{\tau_n}(\alpha)$, 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 Figure 1(a). Interestingly, this optimal value of α is negative for small values of γ, say $\gamma \leqslant 0.2$. By contrast, for large values of γ (close to 1/2), it tends to one, favoring thus the robustness of order statistics over the tail sensitivity of asymmetric least squares. In the special case of stock returns, where realized values of the tail index were found in Gabaix [26] to be $\gamma \approx 1/3$, the corresponding variance-optimal combination parameter $\alpha(\gamma)$ varies around $\alpha(1/3) \approx 0.9$. It can also be seen that the simple mean $\overline{\gamma}_{\tau_n}(1/2)$ of $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$, with $\alpha = 1/2$, minimizes the asymptotic variance v_α for $\gamma = 1/4$. This is unsurprising since both $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$ have the same asymptotic variance in this case, as illustrated in Figure 2 in the Supplementary Material document. It can be seen thereon that the simple mean $\overline{\gamma}_{\tau_n}(1/2)$ affords a middle course between $\widehat{\gamma}_{\tau_n} \equiv \overline{\gamma}_{\tau_n}(1)$ and $\widetilde{\gamma}_{\tau_n} \equiv \overline{\gamma}_{\tau_n}(0)$ in terms of asymptotic variance. In terms of smoothness, $\overline{\gamma}_{\tau_n}(1/2)$ offers a middle course as well, as shown in Section A of the Supplementary Material document.

Remark 4. Let us comment on the covariance of $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$, as well as the variance of the expectHill estimator given by

$$\mathbb{V}(\overline{\gamma}_{\tau_n}(\alpha)) = \alpha^2\mathbb{V}(\widehat{\gamma}_{\tau_n}) + (1-\alpha)^2\mathbb{V}(\widetilde{\gamma}_{\tau_n}) + 2\alpha(1-\alpha)\mathbb{C}ov(\widehat{\gamma}_{\tau_n}, \widetilde{\gamma}_{\tau_n}).$$

An inspection of the proof of Theorem 2 (see page 22 of the Supplementary Material docu-

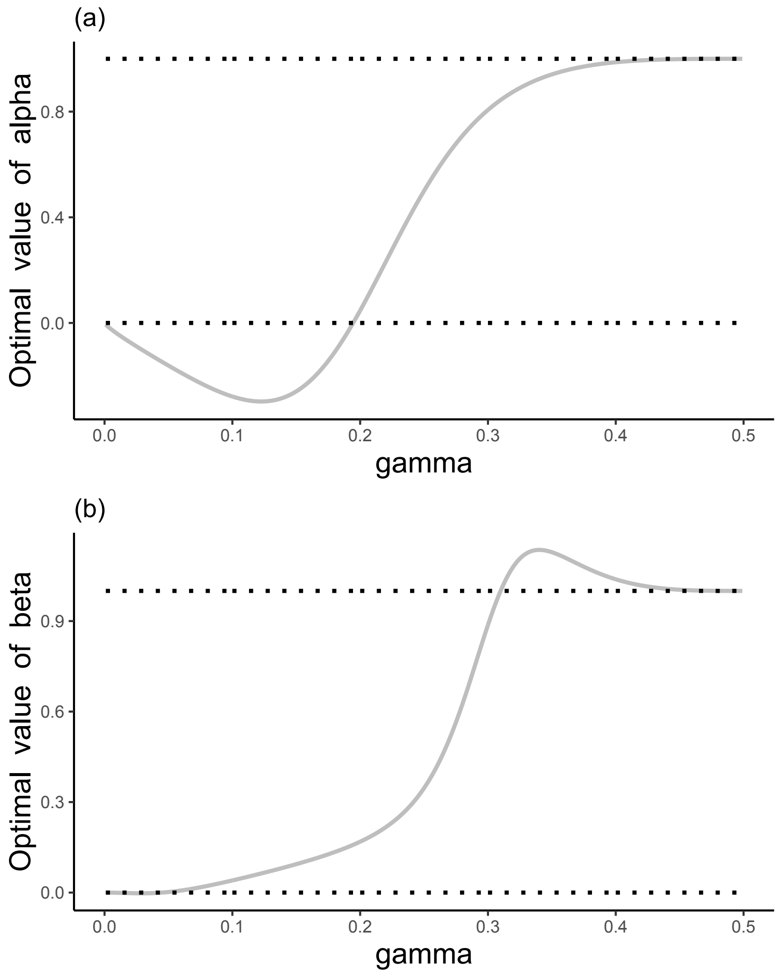

Figure 1: *(a) — Evolution of the variance-optimal value $\alpha(\gamma)$ against $\gamma \in (0, 1/2)$. The dotted lines represent the values $\alpha = 0$ and $\alpha = 1$. (b) — Evolution of the variance-optimal value $\beta_{\alpha(\gamma)}(\gamma)$ against $\gamma \in (0, 1/2)$. The dotted lines represent the values $\beta = 0$ and $\beta = 1$.*

ment) reveals that the asymptotic covariance of $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$ is

$$\mathbb{C}ov(\widehat{\gamma}_{\tau_n}, \widetilde{\gamma}_{\tau_n}) = \gamma[(\gamma^{-1} - 1)^{\gamma - 1} - \gamma] = \gamma^2 \left(\frac{(\gamma^{-1} - 1)^{\gamma}}{1 - \gamma} - 1 \right).$$

Since $\mathbb{V}(\widehat{\gamma}_{\tau_n}) = \gamma^2$ and $\mathbb{V}(\widetilde{\gamma}_{\tau_n}) = 2\gamma^3/(1 - 2\gamma)$, it is arguably more instructive to study the correlation term

$$\operatorname{corr}(\widehat{\gamma}_{\tau_n}, \widetilde{\gamma}_{\tau_n}) = \frac{\mathbb{C}ov(\widehat{\gamma}_{\tau_n}, \widetilde{\gamma}_{\tau_n})}{\sqrt{\mathbb{V}(\widehat{\gamma}_{\tau_n})\mathbb{V}(\widetilde{\gamma}_{\tau_n})}} = \sqrt{\frac{1-2\gamma}{2\gamma}} \left(\frac{(\gamma^{-1}-1)^{\gamma}}{1-\gamma} - 1 \right).$$

This correlation as a function of $\gamma \in (0, 1/2]$ is represented on Figure 2. It seems to be a concave function of γ, attaining a maximum of approximately 0.8 at around $\gamma = 1/6$. Note though that for $\alpha = 1/2$ and $\gamma = 1/6$, we have $\mathbb{V}(\overline{\gamma}_{\tau_n}(\alpha))/\mathbb{V}(\widehat{\gamma}_{\tau_n}) \approx 2/3$. Consequently, even for values of γ where the correlation between $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$ is high, the improvement brought in terms of variance by considering the expectHill estimator can be very substantial.

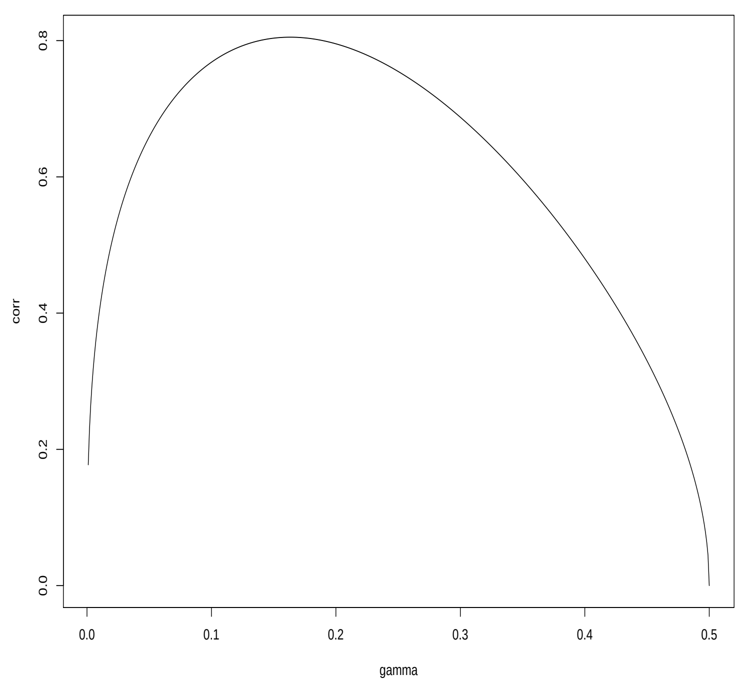

Figure 2: Correlation term $\operatorname{corr}(\widehat{\gamma}_{\tau_n}, \widetilde{\gamma}_{\tau_n})$ as a function of $\gamma \in (0, 1/2)$.

4 Extreme expectile estimation

In this section, we first return to intermediate expectile estimation by making use of the general class of γ estimators $\{\overline{\gamma}_{\tau_n}(\alpha)\}_{\alpha\in\mathbb{R}}$ to construct alternative 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 asymmetric least squares estimator $\widetilde{\xi}_{\tau_n}$ defined in (5), one may use the asymptotic connection $\xi_{\tau_n} \sim (\gamma^{-1}-1)^{-\gamma} q_{\tau_n}$, described in (4), to define the following semiparametric estimator of ξ_{τ_n}:

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

In the special case $\alpha = 1$, we recover the purely quantile-based estimator $\widehat{\xi}_{\tau_n}(1)$ suggested in Daouia *et al.* [13]. The asymmetric least squares estimator $\widetilde{\xi}_{\tau_n}$ inherits the requisite property of coherency of the true risk measure ξ_{τ_n} and is superior to $\widehat{\xi}_{\tau_n}(1)$ in terms of asymptotic variance when the tail index $\gamma < 0.3$, as visualized in Figure 3 (right-hand side). By contrast, $\widehat{\xi}_{\tau_n}(1)$ is more efficient over the range of values of $\gamma > 0.3$ that are common in actuarial and financial applications, as can be seen from Figure 3 (left-hand side). The asymptotic variances of both $\widetilde{\xi}_{\tau_n}$ and $\widehat{\xi}_{\tau_n}(1)$ can be found in Daouia *et al.* [13] and follow as special cases from Theorem 3 below.

In order to obtain the best of both $\widehat{\xi}_{\tau_n}(1)$ and $\widetilde{\xi}_{\tau_n}$, it is then natural to consider their linear combination, or more generally, one may combine the two estimators $\widehat{\xi}_{\tau_n}(\alpha)$ and $\widetilde{\xi}_{\tau_n}$

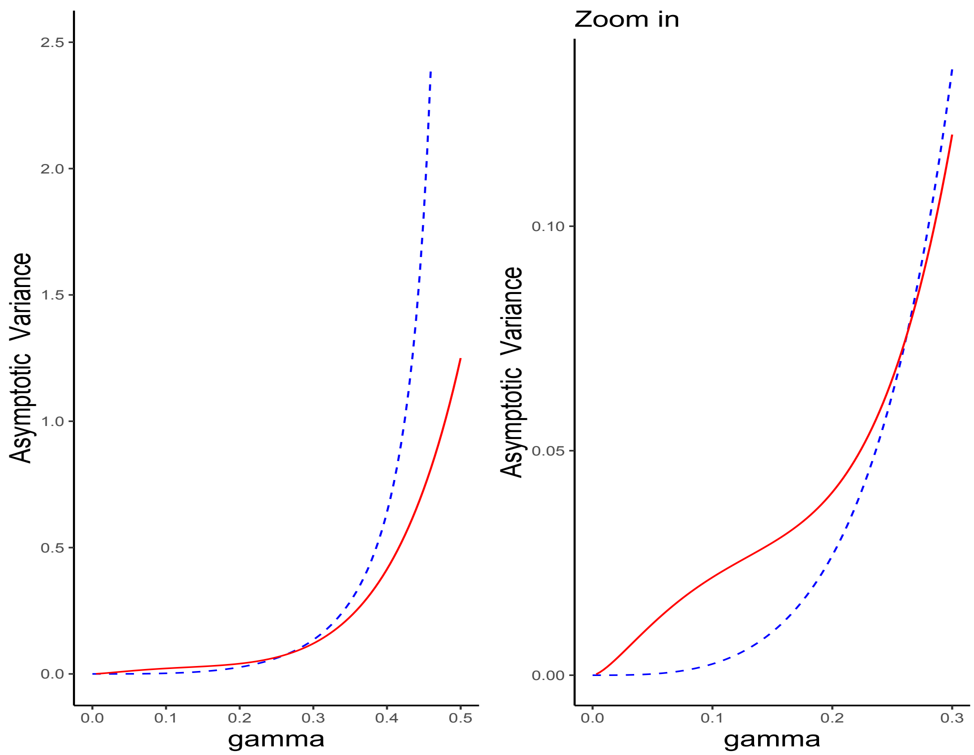

Figure 3: *Asymptotic variances of* $\widetilde{\xi}_{\tau_n}$ *in dashed blue line and* $\widehat{\xi}_{\tau_n}(1)$ *in solid red line, with* $\gamma \in (0, 1/2)$.

to define, for $\beta \in \mathbb{R}$, the weighted estimator

$$\overline{\xi}_{\tau_n}(\alpha, \beta) := \beta\, \widehat{\xi}_{\tau_n}(\alpha) + (1 - \beta)\, \widetilde{\xi}_{\tau_n}. \tag{11}$$

When $\alpha = 1$, we recover the particular expectile estimator $\overline{\xi}_{\tau_n}(\beta) := \overline{\xi}_{\tau_n}(1, \beta)$ introduced in Daouia *et al.* [15]. The limit distribution of the more general variant $\overline{\xi}_{\tau_n}(\alpha, \beta)$ crucially relies on the asymptotic dependence structure in Theorem 2 between $\overline{\gamma}_{\tau_n}(\alpha)$, $\widehat{q}_{\tau_n}$ and $\widetilde{\xi}_{\tau_n}$.

Theorem 3. *Suppose that the conditions of Theorem 1 hold. Then, for any $\alpha, \beta \in \mathbb{R}$,*

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

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

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

and $(\Psi_\alpha, \Theta, \Xi)$ is a trivariate Gaussian centered random vector with covariance matrix $\mathfrak{V}_\alpha$ as in Theorem 2.

Similarly to the tail index estimator $\overline{\gamma}_{\tau_n}(\alpha)$, the expectile estimator

$$\overline{\xi}_{\tau_n}(\alpha,\beta) \equiv \beta\,\left(\overline{\gamma}_{\tau_n}(\alpha)^{-1} - 1\right)^{-\overline{\gamma}_{\tau_n}(\alpha)}\,\widehat{q}_{\tau_n} + (1-\beta)\,\widetilde{\xi}_{\tau_n}$$

depends on a weighting parameter $(\alpha,\beta) \in \mathbb{R}^2$. The optimal value (α^*, β^*) minimizing the AMSE of $\overline{\xi}_{\tau_n}(\alpha,\beta)$ is also difficult to estimate, as it depends on the elusive parameters ρ and λ_1. Our strategy here will thus be to set first $\alpha = \alpha(\gamma)$, the variance-optimal choice of the weighting parameter in the estimator $\overline{\gamma}_{\tau_n}(\alpha)$ at the heart of the construction of $\overline{\xi}_{\tau_n}(\alpha,\beta)$, and then to determine the parameter β minimizing the asymptotic variance of $\overline{\xi}_{\tau_n}(\alpha(\gamma),\beta)$. This is a sensible approach to find a low-variance estimator within the class $(\overline{\xi}_{\tau_n}(\alpha,\beta))_{(\alpha,\beta)\in\mathbb{R}^2}$.

Remark 5. Like the variance-optimal weighting coefficient α defining the expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ in (8) [see Remark 3], the combination parameter β, which minimizes the asymptotic variance of the intermediate expectile estimator $\overline{\xi}_{\tau_n}(\alpha,\beta)$ in (11), has a closed form expression that only depends on the tail index γ. Indeed, this optimal value of β is obtained by minimizing the variance of the random quantity

$$\beta\left([(1-\gamma)^{-1}-\log(\gamma^{-1}-1)]\Psi_\alpha+\Theta\right)+(1-\beta)\Xi,$$

where (Ψ_α,Θ,Ξ) is a trivariate Gaussian centered random vector with the covariance matrix $\mathfrak{V}_\alpha$ given in Theorem 2. Setting

$$m(\gamma):=(1-\gamma)^{-1}-\log(\gamma^{-1}-1),$$

the above random quantity equals $\beta\,(m(\gamma)\Psi_\alpha+\Theta-\Xi)+\Xi$. Its variance is then

$$\beta^2\operatorname{Var}(m(\gamma)\Psi_\alpha+\Theta-\Xi)+2\beta\operatorname{Cov}(m(\gamma)\Psi_\alpha+\Theta-\Xi,\,\Xi)+\operatorname{Var}\Xi.$$

The minimizer $\beta_\alpha(\gamma)$ of this variance is

$$\beta_\alpha(\gamma)=-\frac{\operatorname{Cov}(m(\gamma)\Psi_\alpha+\Theta-\Xi,\,\Xi)}{\operatorname{Var}(m(\gamma)\Psi_\alpha+\Theta-\Xi)},\tag{12}$$

where the numerator can be rewritten explicitly as

$$\operatorname{Cov}(m(\gamma)\Psi_\alpha+\Theta-\Xi,\,\Xi)=m(\gamma)\mathfrak{V}_\alpha(1,3)+\mathfrak{V}_\alpha(2,3)-\mathfrak{V}_\alpha(3,3),$$

and the denominator as

$$\begin{aligned}\mathrm{Var}\left(m(\gamma)\Psi_\alpha + \Theta - \Xi\right) &= [m(\gamma)]^2\mathfrak{V}_\alpha(1,1) + \mathfrak{V}_\alpha(2,2) + \mathfrak{V}_\alpha(3,3) \\ &+ 2m(\gamma)\mathfrak{V}_\alpha(1,2) - 2m(\gamma)\mathfrak{V}_\alpha(1,3) - 2\mathfrak{V}_\alpha(2,3).\end{aligned}$$

Taking the variance-optimal weight $\alpha \equiv \alpha(\gamma)$ in the expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$, the plot of the resulting variance-optimal value $\beta_\alpha(\gamma) \equiv \beta_{\alpha(\gamma)}(\gamma)$ against $\gamma \in (0, 1/2)$ is graphed in Figure 1(b). Interestingly, this quantity exceeds one for large values of γ, say $\gamma \geqslant 0.31$.

Let us now extend the estimation procedure far into the right tail, where few or no observations are available. This translates into considering the expectile level $\tau = \tau_n' \to 1$ such that $n(1-\tau_n') \to c \in [0, \infty)$, as $n \to \infty$. To estimate the extreme expectile $\xi_{\tau_n'}$, the basic idea is to extrapolate a consistent expectile estimator of intermediate order τ_n to the very high level τ_n'. To do so, note that on the one hand we have $\xi_{\tau_n'}/\xi_{\tau_n} \sim q_{\tau_n'}/q_{\tau_n}$ in view of (4). On the other hand, we have the classical Weissman extrapolation formula

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

as τ_n and τ_n' approach one (Weissman [48]). Thus, we arrive at the expectile approximation

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

By substituting our expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ and the general weighted intermediate estimator $\overline{\xi}_{\tau_n}(\alpha, \beta)$, respectively, in place of γ and ξ_{τ_n}, we get the extrapolated expectile estimator

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

The special case $\alpha = 1$ corresponds to the estimator $\overline{\xi}^{\star}_{\tau'_n}(\beta) := \overline{\xi}^{\star}_{\tau'_n}(1,\beta)$ introduced by Daouia *et al.* [15]. We extend this estimator by using the generalized expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ instead of the Hill estimator $\widehat{\gamma}_{\tau_n}$. The next theorem gives the asymptotic behavior of $\overline{\xi}^{\star}_{\tau'_n}(\alpha,\beta)$.

Theorem 4. *Suppose that the conditions of Theorem 1 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 $\alpha, \beta \in \mathbb{R}$,*

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

with (b_α, v_α) as in (9) and (10).

One can observe that the limiting distribution of $\overline{\xi}^{\star}_{\tau'_n}(\alpha,\beta)$ is controlled by the asymptotic distribution of $\overline{\gamma}_{\tau_n}(\alpha)$. This is a consequence of the fact that the convergence of $\overline{\xi}^{\star}_{\tau'_n}(\alpha,\beta)$ is governed by that of the extrapolation factor $[(1-\tau'_n)/(1-\tau_n)]^{-\overline{\gamma}_{\tau_n}(\alpha)}$. 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 $\overline{\xi}_{\tau_n}(\alpha,\beta)$ to ξ_{τ_n}, given by Theorem 3, 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.

5 Estimation of tail Expected Shortfall

This section aims to estimate both expectile- and quantile-based forms of Expected Shortfall,

$$\mathrm{XES}_\tau := \frac{1}{1-\tau}\int_\tau^1 \xi_t\,dt, \quad \mathrm{QES}_\tau := \frac{1}{1-\tau}\int_\tau^1 q_t\,dt, \tag{15}$$

at a very extreme security level τ that may approach one at an arbitrarily fast rate. To do so, Daouia *et al.* [15] have already suggested to start by estimating these risk measures at an intermediate level $\tau_n \to 1$ such that $n(1-\tau_n) \to \infty$, before extrapolating the resulting estimates to the far tail by making use of the traditional Hill estimator $\widehat{\gamma}_{\tau_n}$ of the tail index γ. Here, we extend their device by using the generalized expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ in place of $\widehat{\gamma}_{\tau_n}$. The following asymptotic connections, established in Proposition 3 of Daouia *et al.* [15], will prove instrumental in the estimation procedure.

Proposition 2 (Daouia *et al.*, 2020)**.** *Assume that* $\mathbb{E}|Y_-| < \infty$ *and that* Y *has a Pareto-type distribution* (2) *with tail index* $0 < \gamma < 1$*. Then*

$$\frac{XES_\tau}{QES_\tau} \sim \frac{\xi_\tau}{q_\tau} \sim \frac{\mathbb{E}[Y|Y>\xi_\tau]}{\mathbb{E}[Y|Y>q_\tau]} \quad \textit{and} \quad \frac{XES_\tau}{\xi_\tau} \sim \frac{1}{1-\gamma} \sim \frac{\mathbb{E}[Y|Y>\xi_\tau]}{\xi_\tau}, \quad \tau \to 1.$$

5.1 Expectile-based Expected Shortfall

Under the model assumptions that $\mathbb{E}|Y_-| < \infty$ and Y has a heavy-tailed distribution (2), we wish to estimate an extreme value of the expectile-based form $\mathrm{XES}_{\tau_n'}$, where $\tau_n' \to 1$ and $n(1-\tau_n') \to c < \infty$. By Proposition 2, we have

$$\frac{\mathrm{XES}_{\tau_n'}}{\mathrm{XES}_{\tau_n}} \sim \frac{\xi_{\tau_n'}}{\xi_{\tau_n}} \quad \text{as} \quad n \to \infty.$$

It follows from the approximation (13) that $\mathrm{XES}_{\tau_n'} \approx \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma} \mathrm{XES}_{\tau_n}$. Then, by replacing γ with $\overline{\gamma}_{\tau_n}(\alpha)$ and XES_{τ_n} with its empirical counterpart

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

we obtain the extrapolated $\mathrm{XES}_{\tau_n'}$ estimator

$$\widetilde{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} \widetilde{\mathrm{XES}}_{\tau_n}. \tag{16}$$

One may also estimate $\mathrm{XES}_{\tau_n'}$ by using the asymptotic equivalence $\mathrm{XES}_{\tau_n'} \sim (1-\gamma)^{-1}\xi_{\tau_n'}$ in Proposition 2. By substituting γ and $\xi_{\tau_n'}$ with their estimators $\overline{\gamma}_{\tau_n}(\alpha)$ and $\overline{\xi}^{\star}_{\tau_n'}(\alpha,\beta)$, described respectively in (8) and (14), we define the alternative $\mathrm{XES}_{\tau_n'}$ estimator

$$\overline{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha,\beta) := [1-\overline{\gamma}_{\tau_n}(\alpha)]^{-1}\, \overline{\xi}^{\star}_{\tau_n'}(\alpha,\beta) \tag{17}$$

for the weights $\alpha, \beta \in \mathbb{R}$. The next result provides the convergence of the two competing estimators $\widetilde{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha)$ and $\overline{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha,\beta)$ of $\mathrm{XES}_{\tau_n'}$.

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

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha)}{\mathrm{XES}_{\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha),$$
$$\text{and} \qquad \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\overline{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha,\beta)}{\mathrm{XES}_{\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}(b_\alpha, v_\alpha),$$

with (b_α, v_α) as in (9) and (10).

The two estimators share the same asymptotic behavior from a theoretical point of view. However, our experience with simulated data in Section 7.3 indicates that $\widetilde{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha)$ tends to be more efficient in the case of real-valued profit-loss distributions with heavy left and right tails, while $\overline{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha,\beta)$ affords advantageous estimates in the case of non-negative heavy-tailed loss distributions.

5.2 Quantile-based Expected Shortfall

In this section, we return to the estimation of the usual form QES_{p_n} of tail Expected Shortfall, for a pre-specified tail probability $p_n \to 1$ with $n(1-p_n) \to c < \infty$. We wish to derive composite expectile-based estimators from the two $\mathrm{XES}_{\tau_n'}$ estimators introduced above, where $\tau_n' = \tau_n'(p_n)$ is to be determined. The starting point is the asymptotic equivalences $\mathrm{QES}_{p_n} \sim \mathbb{E}[Y|Y > q_{p_n}]$ and $\mathrm{XES}_{\tau_n'} \sim \mathbb{E}[Y|Y > \xi_{\tau_n'}]$ in Proposition 2. The basic idea is then to pick out τ_n' so that $\xi_{\tau_n'} \equiv q_{p_n}$, and hence $\mathrm{QES}_{p_n} \sim \mathrm{XES}_{\tau_n'}$. In this way, QES_{p_n} inherits the extreme value estimators of $\mathrm{XES}_{\tau_n'}$ itself, namely $\widetilde{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha)$ in (16) and $\overline{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha,\beta)$ in (17). Yet, it remains to estimate the extreme expectile level $\tau_n'(p_n) := \tau_n'$ such that $\xi_{\tau_n'} = q_{p_n}$. It has been found in Proposition 3 of Daouia *et al.* [13] that such a level satisfies

$$1 - \tau_n'(p_n) \sim (1-p_n)\,\frac{\gamma}{1-\gamma} \quad \text{as} \quad n \to \infty,$$

under the model assumption of heavy tails (2) with tail index $0 < \gamma < 1$. Built on our novel expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ of γ, we can then estimate $\tau_n'(p_n)$ by

$$\widehat{\tau}_n'(p_n) := 1 - (1-p_n)\,\frac{\overline{\gamma}_{\tau_n}(\alpha)}{1-\overline{\gamma}_{\tau_n}(\alpha)}. \tag{18}$$

By substituting this estimated value in place of $\tau_n'(p_n) \equiv \tau_n'$ in the extrapolated estimators $\widetilde{\text{XES}}^{\star}_{\tau_n'}(\alpha)$ and $\overline{\text{XES}}^{\star}_{\tau_n'}(\alpha,\beta)$, we obtain composite estimators that estimate $\text{XES}_{\tau_n'(p_n)} \sim \text{QES}_{p_n}$. The asymptotic properties of $\widetilde{\text{XES}}^{\star}_{\tau_n'}(\alpha)$ and $\overline{\text{XES}}^{\star}_{\tau_n'}(\alpha,\beta)$, stated in Theorem 5, still hold true for their composite versions as estimators of QES_{p_n}, with the same conditions.

Theorem 6. *Suppose the conditions of Theorem 4 hold with p_n in place of τ_n'. Then, for any $\alpha, \beta \in \mathbb{R}$,*

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

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

with (b_α, v_α) as in (9) and (10).

6 Selection of the weights

Since the seminal works of Crane and Crotty [11] and Bates and Granger [3], combining estimators or forecasts has come to be viewed as a simple and effective way to improve and robustify the estimation or forecasting accuracy over that offered by individual models. Two extensive reviews of the literature, techniques and applications of forecast combinations are Clemen [10] and Timmermann [46], see also Weiss *et al.* [47] for a recent survey. An important step beyond designing the individual competing estimators and their combination is how to weight them, or equivalently, how to assign in our setup appropriate values to the combination parameters α in the expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ and β in the intermediate expectile estimator $\overline{\xi}_{\tau_n}(\alpha,\beta)$, described respectively in (8) and (11). One way to address this issue is by setting α to be a suitable estimate of the weight $\alpha(\gamma)$ that minimizes the asymptotic variance of $\overline{\gamma}_{\tau_n}(\alpha)$. Given that $\alpha(\gamma)$ has a closed form expression that only depends on γ itself (see Remark 3), this suggests using the following two-step estimation procedure:

- In a first step, one may estimate γ by the hybrid version $\overline{\gamma}_{\tau_n}(\frac{1}{2}) = [\widehat{\gamma}_{\tau_n} + \widetilde{\gamma}_{\tau_n}]/2$, for $\alpha = \frac{1}{2}$. Any convex combination would have sufficed at this preliminary stage, but we do not see any reason to bias $\overline{\gamma}_{\tau_n}(\alpha)$ one way or the other;

- In a second step, one may use the consistent estimator $\overline{\alpha}_{\tau_n} := \alpha(\overline{\gamma}_{\tau_n}(\frac{1}{2}))$ of $\alpha(\gamma)$ as the desired combination parameter α in the expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$, resulting in

$$\overline{\gamma}_{\tau_n} := \overline{\gamma}_{\tau_n}(\overline{\alpha}_{\tau_n}), \text{ with } \overline{\gamma}_{\tau_n}(\alpha) = \alpha\widehat{\gamma}_{\tau_n} + (1-\alpha)\widetilde{\gamma}_{\tau_n}. \tag{19}$$

Section 7.1 provides Monte Carlo evidence that the finite-sample performance of the two-step estimator $\overline{\gamma}_{\tau_n}$ is quite remarkable in comparison with the best (in terms of asymptotic variance) version $\overline{\gamma}_{\tau_n}(\alpha(\gamma))$ that is calculated with the true variance-optimal weight $\alpha(\gamma)$ itself. Section 8 shows how these practical guidelines can easily be implemented and applied through empirical data.

Let us now turn to the choice of the second combination parameter β in the intermediate expectile estimator $\overline{\xi}_{\tau_n}(\alpha, \beta)$ and other related expectile and expected shortfall estimators. Once the first combination parameter α is chosen as the optimal value $\alpha(\gamma)$ that minimizes the asymptotic variance of the expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$, the second weight β can similarly be set as the optimal value $\beta_\alpha(\gamma)$ which minimizes the asymptotic variance of $\overline{\xi}_{\tau_n}(\alpha, \beta)$, with $\alpha = \alpha(\gamma)$. The explicit expression of the variance-optimal weight $\beta_{\alpha(\gamma)}(\gamma)$, derived in

Remark 5, motivates the plug-in estimator

$$\overline{\beta}_{\tau_n} := \beta_{\overline{\alpha}_{\tau_n}}(\overline{\gamma}_{\tau_n}), \tag{20}$$

obtained by substituting the estimated values $\overline{\alpha}_{\tau_n}$ and $\overline{\gamma}_{\tau_n}$ in place of the population values $\alpha(\gamma)$ and γ.

When it comes to compute the intermediate expectile estimator $\overline{\xi}_{\tau_n}(\alpha, \beta)$ in (11), which is in turn used to compute the extreme expectile and expected shortfall estimators $\overline{\xi}^{\star}_{\tau_n'}(\alpha, \beta)$ in (14), $\widetilde{\text{XES}}^{\star}_{\tau_n'}(\alpha)$ in (16), $\overline{\text{XES}}^{\star}_{\tau_n'}(\alpha, \beta)$ in (17), and $\widetilde{\text{XES}}^{\star}_{\widehat{\tau}_n'(p_n)}(\alpha)$ and $\overline{\text{XES}}^{\star}_{\widehat{\tau}_n'(p_n)}(\alpha, \beta)$ in Theorem 6, we can consider using the following two-step procedure:

- First, estimate the combination parameters α and β by $\overline{\alpha}_{\tau_n}$ and $\overline{\beta}_{\tau_n}$, respectively;
- Second, use the tail expectile and expected shortfall estimators above, as if α and β were known, by substituting in the estimated values $\overline{\alpha}_{\tau_n}$ and $\overline{\beta}_{\tau_n}$.

Our experiments with simulated data in Sections 7.2-7.4 provide Monte Carlo evidence that the resulting two-step estimators perform remarkably well compared with their corresponding variance-optimal versions using the theoretical weights $\alpha = \alpha(\gamma)$ and $\beta = \beta_{\alpha(\gamma)}(\gamma)$.

Remark 6. From the theoretical standpoint, in view of the consistency of $\overline{\gamma}_{\tau_n}$ in Theorem 2 and the continuous dependence of $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$ viewed as functions of γ, it is straightforward to show that the adaptive estimator $\overline{\gamma}_{\tau_n}(\overline{\alpha}_{\tau_n})$ has the same asymptotic distribution as $\overline{\gamma}_{\tau_n}(\alpha(\gamma))$. Indeed

$$\overline{\gamma}_{\tau_n}(\overline{\alpha}_{\tau_n}) - \overline{\gamma}_{\tau_n}(\alpha(\gamma)) = (\overline{\alpha}_{\tau_n} - \alpha(\gamma))(\widehat{\gamma}_{\tau_n} - \gamma - [\widetilde{\gamma}_{\tau_n} - \gamma])$$

and therefore, under the conditions of Theorem 2, $\overline{\gamma}_{\tau_n}(\overline{\alpha}_{\tau_n}) - \overline{\gamma}_{\tau_n}(\alpha(\gamma)) = \mathrm{o}_{\mathbb{P}}(1/\sqrt{n(1-\tau_n)})$, by the consistency of $\overline{\alpha}_{\tau_n}$ and the $\sqrt{n(1-\tau_n)}-$asymptotic normality of both $\widehat{\gamma}_{\tau_n}$ and $\widetilde{\gamma}_{\tau_n}$ stated in Theorem 1. Similarly, the adaptive estimator $\overline{\xi}_{\tau_n}(\overline{\alpha}_{\tau_n}, \overline{\beta}_{\tau_n})$ has the same asymptotic distribution as $\overline{\xi}_{\tau_n}(\alpha(\gamma), \beta_{\alpha(\gamma)}(\gamma))$.

7 Numerical simulations

In order to illustrate the behavior of the presented estimation procedures of the tail index γ and the two expected shortfall forms $\mathrm{XES}_{\tau_n'}$ and QES_{p_n}, we consider the Student t-distribution with $1/\gamma$ degrees of freedom, 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 $n = 2{,}500$ and true tail index $\gamma \in \{0.33, 0.48\}$ (motivated by a number of actuarial and financial applications where the realized values of γ were found to vary between 0.33 and 0.48, see Gabaix [26] for a nice survey and Cai *et al.* [9] and Daouia *et al.* [13] for very recent applications). 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 B of the Supplementary Material document.

7.1 Tail index estimation

We investigated the finite-sample performance of the two-step expectHill estimator $\overline{\gamma}_{\tau_n}$ of the tail index γ, obtained in (19) by substituting in the estimate $\overline{\alpha}_{\tau_n} = \alpha(\overline{\gamma}_{\tau_n}(\frac{1}{2}))$ of the theoretical optimal weight $\alpha(\gamma)$ described in Remark 3. As a first benchmark, we used the 'variance-optimal' expectHill version $\overline{\gamma}_{\tau_n}(\alpha(\gamma))$ that is defined in (8) in the same way as $\overline{\gamma}_{\tau_n}$, but calculated with the true weight $\alpha(\gamma)$ itself (rather than its estimate $\overline{\alpha}_{\tau_n}$). A second benchmark is the 'hybrid' expectHill estimator $\overline{\gamma}_{\tau_n}(1/2)$ obtained with the average weight $\alpha = 1/2$. The last and most important benchmark is the 'oracle' expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ obtained by selecting the value of α which minimizes its MSE. Remarkably, the Monte Carlo estimates, obtained in Supplement B.1, indicate that both 'oracle' and 'variance-optimal' expectHill estimators have very close MSE in all cases, which is good news for our variance-optimal selection device (though the oracle procedure may provide slightly better estimates, in terms of bias, for large values of γ). Moreover, in terms of both bias and MSE, the Monte Carlo estimates indicate that the accuracy of the two-step expectHill estimator $\overline{\gamma}_{\tau_n}$ is quite respectable in comparison with the theoretical version $\overline{\gamma}_{\tau_n}(\alpha(\gamma))$. Finally, while the 'hybrid' expectHill estimator $\overline{\gamma}_{\tau_n}(1/2)$ performs quite well in the Student scenario, it is clearly outperformed by our two-step estimator $\overline{\gamma}_{\tau_n}$ in both Fréchet and Pareto scenarios.

7.2 Extreme expectile estimation

The simulation experiments undertaken here are concerned with the two-step estimator

$$\overline{\xi}^{\star}_{\tau_n'} := \overline{\xi}^{\star}_{\tau_n'}(\alpha, \beta) = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} \overline{\xi}_{\tau_n}(\alpha, \beta) \tag{21}$$

of the extreme expectile $\xi_{\tau_n'}$. It is computed by substituting in the estimated values $\alpha = \overline{\alpha}_{\tau_n}$ and $\beta = \overline{\beta}_{\tau_n}$ of the theoretical variance-optimal weights $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$, as described in Section 6. Its accuracy is evaluated in comparison with the variance-optimal version $\overline{\xi}^{\star}_{\tau_n'}(\alpha(\gamma), \beta_{\alpha(\gamma)}(\gamma))$ itself that is obtained by replacing the combination parameters α and β with the theoretical values $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$. We also considered two additional benchmark estimators: the 'hybrid' version corresponding to the average weights $\alpha = \beta = 1/2$, and the 'oracle' version obtained by selecting the values of α and β which minimize the MSE estimates. The Monte Carlo results we obtained in Supplement B.2 show that the use of the estimated values $\alpha = \overline{\alpha}_{\tau_n}$ and $\beta = \overline{\beta}_{\tau_n}$ provides, in all cases, very similar results, in terms of both MSE and bias, to the variance-optimal weights $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$ themselves. Most importantly, the MSE and bias estimates based on our variance-optimal selection of weights appear to be quite good in comparison with the oracle estimates. Finally, although the hybrid version $\overline{\xi}^{\star}_{\tau_n'}(1/2, 1/2)$ exhibits a slightly better bias relative to the variance-optimal estimates, the latter are superior in terms of MSE.

7.3 Expected shortfall $\mathrm{XES}_{\tau_n'}$ estimation

We also compared the finite-sample performance of the two-step $\mathrm{XES}_{\tau_n'}$ estimators

$$\widetilde{\mathrm{XES}}^{\star}_{\tau_n'} := \widetilde{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha) = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} \widetilde{\mathrm{XES}}_{\tau_n}, \tag{22}$$

$$\overline{\mathrm{XES}}^{\star}_{\tau_n'} := \overline{\mathrm{XES}}^{\star}_{\tau_n'}(\alpha, \beta) = [1 - \overline{\gamma}_{\tau_n}(\alpha)]^{-1}\, \overline{\xi}^{\star}_{\tau_n'}(\alpha, \beta) \tag{23}$$

that are computed by substituting in the estimated values $\alpha = \overline{\alpha}_{\tau_n}$ and $\beta = \overline{\beta}_{\tau_n}$ of the weights $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$. In addition, we compared these competing estimators with their benchmark versions $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha(\gamma))$ and $\overline{\mathrm{XES}}^{\star}_{\tau'_n}(\alpha(\gamma), \beta_{\alpha(\gamma)}(\gamma))$ that are obtained by substituting in the variance-optimal weights $\alpha = \alpha(\gamma)$ and $\beta = \beta_{\alpha(\gamma)}(\gamma)$. The Monte Carlo estimates of MSE and bias, obtained in Supplement B.3, indicate that the two-step estimators are very accurate with respect to their variance-optimal versions in all cases. Also, the estimates $\widetilde{\mathrm{XES}}^{\star}_{\tau'_n}$ seem to perform better in the case of the real-valued Student distribution, while their competitors $\overline{\mathrm{XES}}^{\star}_{\tau'_n}$ appear to be the most efficient in the case of the non-negative Fréchet and Pareto distributions. It should be noted that the central part and the left tail of the underlying distribution have an impact on the behavior of the *expectile-based* estimators at the right tail. This effect would not occur in the case of pure *quantile-based* estimators that correspond to the combination weights $\alpha = \beta = 1$. The reason for this is that quantiles only depend on the frequency of tail observations. By contrast, expectiles (for any asymmetry level τ) rely on the distance to "all" observations due to their L^2-nature. Accordingly, shifting mass in the lower tail of a distribution has no impact on the quantiles of the upper tail, but it does have an impact on all the expectiles.

7.4 Expected shortfall QES_{p_n} estimation

We have also undertaken simulation experiments to evaluate the finite-sample performance of the composite expectile-based estimators

$$\begin{aligned}
\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)} := \widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha) &= \left(\frac{1-\widehat{\tau}'_n(p_n)}{1-\tau_n}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} \widetilde{\mathrm{XES}}_{\tau_n} \qquad (24)\\
&= \left(\frac{\overline{\gamma}_{\tau_n}(\alpha)}{1-\overline{\gamma}_{\tau_n}(\alpha)}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} \widetilde{\mathrm{XES}}^{\star}_{p_n}(\alpha),\\
\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)} := \overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha,\beta) &= [1-\overline{\gamma}_{\tau_n}(\alpha)]^{-1}\, \overline{\xi}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha,\beta) \qquad (25)\\
&= \left(\frac{\overline{\gamma}_{\tau_n}(\alpha)}{1-\overline{\gamma}_{\tau_n}(\alpha)}\right)^{-\overline{\gamma}_{\tau_n}(\alpha)} \overline{\mathrm{XES}}^{\star}_{p_n}(\alpha,\beta),
\end{aligned}$$

where $\widehat{\tau}'_n(p_n)$ is defined in (18), with $\alpha = \overline{\alpha}_{\tau_n}$ and $\beta = \overline{\beta}_{\tau_n}$. Note that, in view of Proposition 2 and (4), we have $\mathrm{QES}_{p_n} \sim (\gamma^{-1}-1)^{\gamma}\, \mathrm{XES}_{p_n}$, as $n \to \infty$. Then, by replacing in this asymptotic equivalence γ and XES_{p_n} with their respective estimators $\overline{\gamma}_{\tau_n}(\alpha)$ and $\widetilde{\mathrm{XES}}^{\star}_{p_n}(\alpha)$ or $\overline{\mathrm{XES}}^{\star}_{p_n}(\alpha,\beta)$, we get directly the composite estimators in (24) and (25). The latter estimate the same conventional expected shortfall QES_{p_n} as the purely quantile-based estimator

$$\widehat{\mathrm{QES}}^{\star}_{p_n} := \left(\frac{1-p_n}{1-\tau_n}\right)^{-\widehat{\gamma}_{\tau_n}} \frac{1}{\lfloor n(1-\tau_n)\rfloor} \sum_{i=1}^{\lfloor n(1-\tau_n)\rfloor} Y_{n-i+1,n} \qquad (26)$$

proposed by El Methni *et al.* [22]. We compared their MSE and bias in Supplement B.4 with those of $\widehat{\mathrm{QES}}^{\star}_{p_n}$ and those of their benchmark versions that are obtained with $\alpha = \alpha(\gamma)$ and $\beta = \beta_{\alpha(\gamma)}(\gamma)$. We arrive at the following tentative conclusions:

- In the case of the (real-valued) Student distribution, the best estimator is clearly $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$;

- In the cases of Fréchet and Pareto distributions (both positive), the other composite expectile-based estimator $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}_n'(p_n)}$ seems to be the winner.

8 Financial returns data

This section applies our expectHill-based method to estimate the tail expected shortfall on financial returns data. We use the same trade data as in the study of Kim and Meddahi [34] on the SPDR S&P 500 ETF (SPY), which is an exchange traded fund (ETF) that tracks the S&P 500 index. The dataset comprises 10 years of trade data on SPY starting from June 15th, 2004, to June 13th, 2014. The choice of the frequency of data, trading days and time horizon follows the same setup as in Kim and Meddahi [34]. This results in 2,497 days of trade data. Our sample consists of the negative returns (Y_i) depicted on Figure 4.

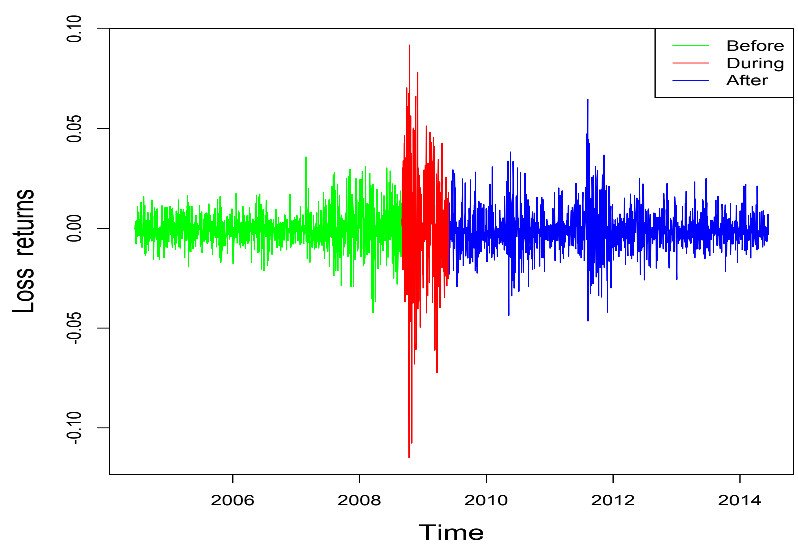

Figure 4: *Daily open-to-close loss returns (i.e. minus returns) of the SPDR S&P 500 ETF (SPY) starting from June 15th, 2004, to June 13th, 2014.*

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. The competing estimates $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}_n'(p_n)} := \widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}_n'(p_n)}(\alpha)$ in (24) and $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}_n'(p_n)} := \overline{\mathrm{XES}}^{\star}_{\widehat{\tau}_n'(p_n)}(\alpha, \beta)$ in (25) of QES_{p_n} are determined in two steps. We first choose the most favorable values of the weighting coefficients α and β, then we select an appropriate intermediate level τ_n for each estimator. A common practice in extreme value analysis is to use the discrete reparameterization $\tau_n = 1 - k/n$, for the selected range of values $1 \leqslant k \leqslant n/\log n$, where the integer to be selected k represents the effective sample size for tail extrapolation.

First, we verify the model assumption of a heavy-tailed distribution with $\gamma < \frac{1}{2}$ that is required for the procedure. This assumption is already confirmed by the plots of the expectHill estimator $\overline{\gamma}_{1-k/n}(\alpha)$ in Figure 1(a) in Section A of the Supplementary Material document, for the special cases $\alpha = 0, \frac{1}{2}, 1$. The estimated values of γ obtained therein in Table 1 (second row) vary between $\overline{\gamma}_{1-k/n}(0) = 0.33$ and $\overline{\gamma}_{1-k/n}(1) = 0.35$, with $\overline{\gamma}_{1-k/n}(\frac{1}{2}) = 0.34$.

The optimal value of the combination parameter α that minimizes the asymptotic variance of $\overline{\gamma}_{1-k/n}(\alpha)$ can be estimated, as described in Section 6, by

$$\overline{\alpha}_{1-k/n} := \alpha\left(\overline{\gamma}_{1-k/n}\,(1/2)\right) \equiv \alpha(0.34) = 0.92.$$

The corresponding expectHill estimator, described in (19), is thus

$$\overline{\gamma}_{1-k/n} := \overline{\gamma}_{1-k/n}(0.92) = 0.92\widehat{\gamma}_{1-k/n} + 0.08\widetilde{\gamma}_{1-k/n}.$$

Its plot against k is depicted on Figure 5(a), along with the plot of the standard Hill estimator $\widehat{\gamma}_{1-k/n}$. The two plots are similar due to the important contribution of the Hill component in the linear combination defining the expectHill estimate. To examine the influence of the crisis period on this contribution, we first divide the full period into three subperiods: Before Crisis, from June 15th, 2004, to August 29th, 2008 (1,053 trading days); During Crisis, from September 2nd, 2008, to May 29th, 2009 (185 trading days), and After Crisis, from June 1st, 2009, to June 13th, 2014 (1,259 trading days). For each subperiod, the model assumption of tail heaviness with $\gamma < \frac{1}{2}$ is confirmed by the resulting expectHill estimates $\overline{\gamma}_{1-k/n}(\alpha)$ in Figure 1(b)-(d) and Table 1 in Section A of the Supplementary Material document, for the particular values $\alpha = 0, \frac{1}{2}, 1$. The estimated optimal values $\overline{\alpha}_{1-k/n}$ of the weight α are displayed below in Table 1 (fifth column) for the three subperiods. The corresponding expectHill estimators $\overline{\gamma}_{1-k/n}$ are graphed below in Figure 5(b)-(d) against k, along with the Hill estimator $\widehat{\gamma}_{1-k/n}$.

The final pointwise estimates $\widehat{\gamma}_{1-k/n}$ and $\overline{\gamma}_{1-k/n}$ are shown in Table 1 (third and fourth columns) for all considered periods. These values are chosen according to the same automatic selection procedure described in Section A of the Supplementary Material document: This selection consists first in computing the standard deviations of the estimator over a *moving window* large enough to cover around 5% (20% for the crisis period whose length is only 185 trading days) of the possible values of k in the selected range $1 \leqslant k \leqslant n/\log n$. The first window over which the standard deviation has a local minimum, and is less than the average standard deviation across all windows, is then selected as the first stable region of the plot. Finally, the value of k which corresponds to the median estimate within this window defines the desired sample fraction.

Interestingly, the difference between the obtained Hill and expectHill estimates becomes more pronounced during the crisis period. Also, the linear combination coefficient $\overline{\alpha}_{1-k/n}$ decreases during and after the period of crisis, which indicates that the contribution of the asymmetric least squares (expectile-based) component to the estimation procedure increases appreciably with the crisis. We arrive at this same tentative conclusion regarding the evolution of the estimated values $\overline{\beta}_{1-k/n}$ in (20) of the second combination parameter β, which are displayed in the sixth column of Table 1.

Using the resulting weights $\alpha = \overline{\alpha}_{1-k/n}$ and $\beta = \overline{\beta}_{1-k/n}$ in (24) and (25), we can apply our two-step method to obtain the QES_{p_n} estimates $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)} := \widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)$ and $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)} := \overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha, \beta)$, studied in Theorem 6. The plots of these estimates against k are depicted on Figure 6, for all considered periods, as rainbow curve and dashed black curve, respectively. The effect of the expectHill estimate $\overline{\gamma}_{1-k/n}$ on $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ is highlighted by a colour-scheme, ranging from dark red (low $\overline{\gamma}_{1-k/n}$) to dark violet (high $\overline{\gamma}_{1-k/n}$). By Theorem 6, under the bias condition $\lambda_1 = \lambda_2 = 0$, we have

$$\frac{\sqrt{k}}{\log[k/n(1-p_n)]}\left(\frac{\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)}{\mathrm{QES}_{p_n}} - 1\right) \xrightarrow{d} \mathcal{N}(0, v_\alpha(\gamma)),$$

where $v_\alpha(\gamma) := v_\alpha$ is described in (10). The (symmetric) expectile-based asymptotic confidence interval with confidence level $100\vartheta\%$ then has the form $\widetilde{\mathrm{CI}}_\vartheta(k) = \widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha) \times \mathcal{I}$, where $\mathcal{I}$ stands for the interval

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

with $z_{(1+\vartheta)/2}$ being the $(1+\vartheta)/2-$quantile of the standard Gaussian distribution. Likewise,

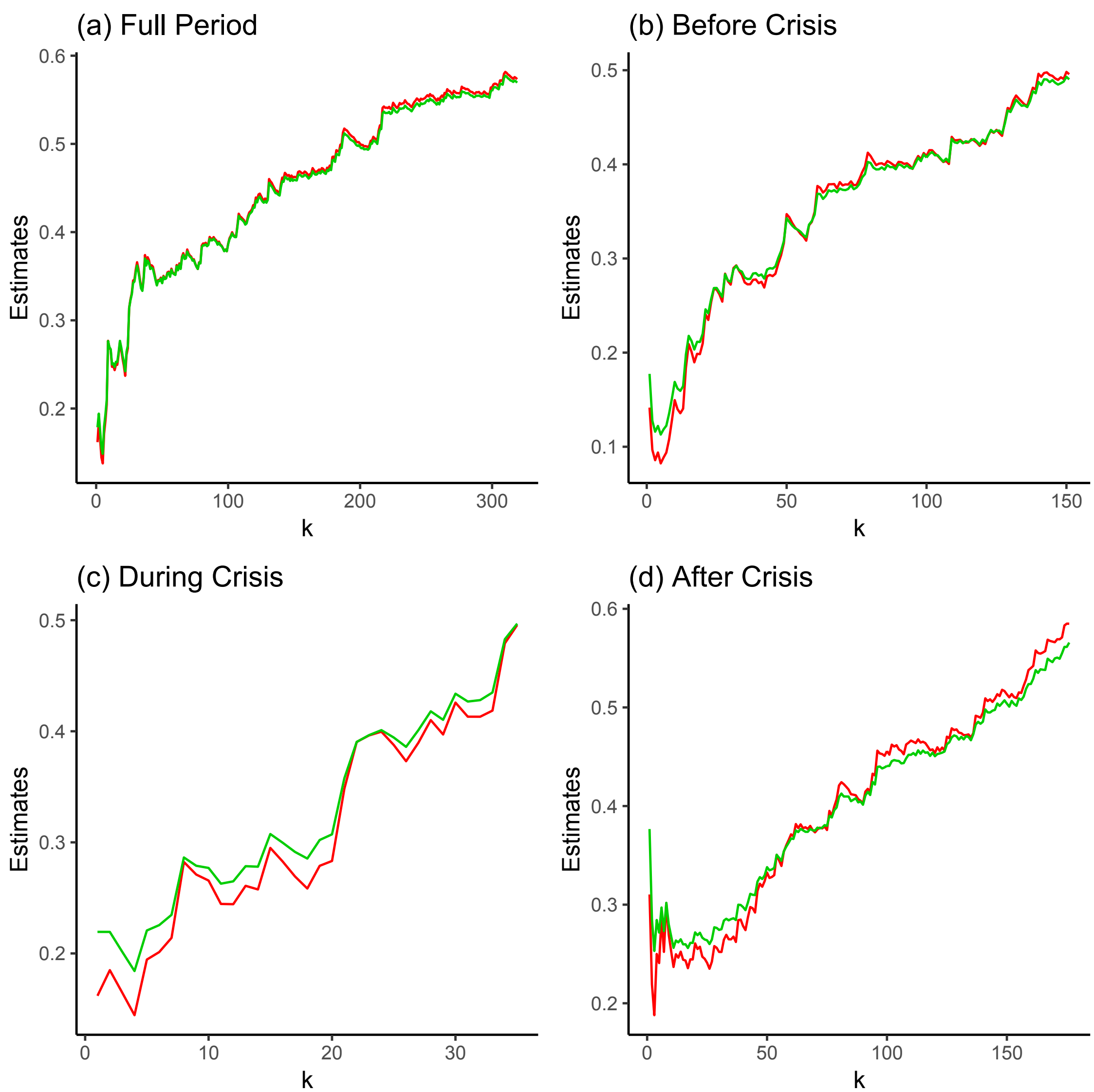

Figure 5: *Plots of the Hill and expectHill estimates* $\widehat{\gamma}_{1-k/n}$ *and* $\overline{\gamma}_{1-k/n}$ *against various values of* k*, based on daily loss returns of the SPDR S&P 500 ETF (SPY). The estimates depicted on (a)-(d) correspond, respectively, to the full 10-years period (2004-2014) and the three sub-periods: Before Crisis (2004-2008), During Crisis (2008-2009) and After Crisis (2009-2014).*

the confidence interval derived from the asymptotic normality of $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha)$, in Theorem 6, can be expressed as $\overline{\mathrm{CI}}_{\vartheta}(k) = \overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}(\alpha,\beta) \times \mathcal{I}$.

The plots of the asymptotic 95% confidence intervals $\widetilde{\mathrm{CI}}_{0.95}(k)$ and $\overline{\mathrm{CI}}_{0.95}(k)$ against k are superimposed in Figure 6, respectively, in dotted blue lines and solid grey lines. It can be seen that the (rainbow) paths of the estimates $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ and their associated (dotted blue) confidence bands are less volatile and less pessimistic than, respectively, their corresponding (dashed black) paths of the estimates $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ and their associated (solid grey) confidence bands. In this situation of real-valued profit-loss distributions, we have already provided some Monte Carlo evidence that the estimates $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ are more efficient and accurate relative to their competitors $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$.

The final selected pointwise levels $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ and $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$, based on minimizing the standard deviations of the estimates over a moving window, are displayed in the second and fourth columns of Table 2, along with their corresponding confidence intervals $\widetilde{\mathrm{CI}}_{0.95}(k)$ and $\overline{\mathrm{CI}}_{0.95}(k)$ in the third and fifth columns. The last column indicates the sample maximum loss $Y_{n,n}$ for each period. The messages yielded by the two competing methods are broadly similar, indicating particularly that the expected shortfall (ES) levels differ appreciably before, during and after the crisis period. Clearly, the crisis period exhibits ES levels (around -11.7% to -12.7%) three times higher than the pre-crisis period (around -3.6% to -3.8%) and about twice and a half higher than the post-crisis period (around -4.8% to -4.9%). Also, the ES levels during the crisis period are more conservative than the most catastrophic recorded loss (around -9.2%), extrapolating thus outside the sample maximum $Y_{n,n}$.

The theory for our composite expectHill-based estimators $\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ and $\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ is derived for independent and identically distributed random variables $Y_1, \ldots, Y_n$. In this

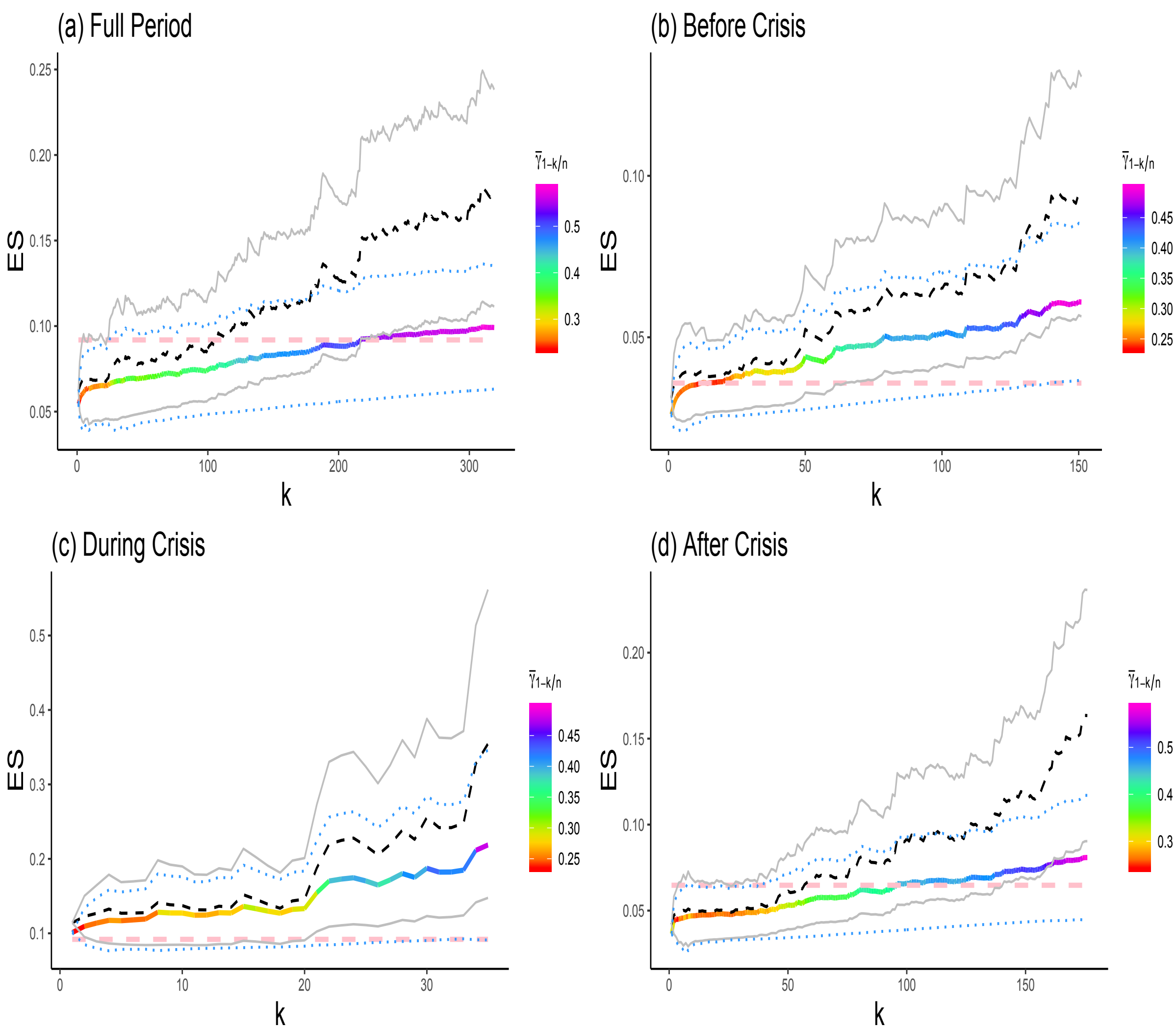

Figure 6: *Plots of the ES estimates based on daily loss returns of the SPDR S&P 500 ETF (SPY). The estimates* $\widetilde{XES}^{\star}_{\widehat{\tau}'_n(p_n)}$ *as rainbow curve and* $\overline{XES}^{\star}_{\widehat{\tau}'_n(p_n)}$ *as dashed black curve, along with the asymptotic* 95% *confidence intervals* $\widetilde{CI}_{0.95}(k)$ *in dotted blue lines and* $\overline{CI}_{0.95}(k)$ *in solid grey lines. The sample maximum* $Y_{n,n}$ *indicated in horizontal dashed pink line.*

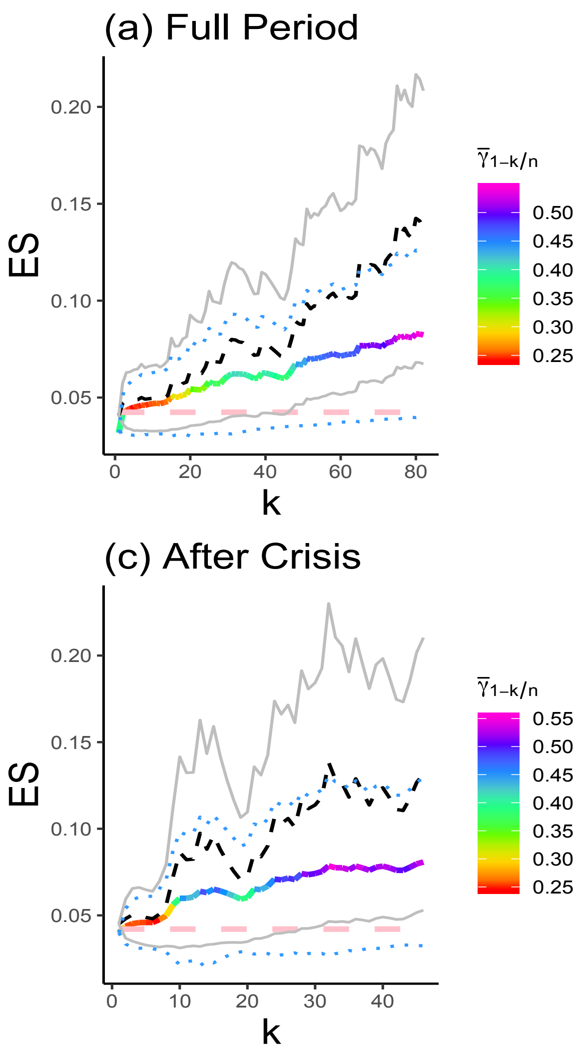

Figure 7: *Plots of the ES estimates based on weekly loss returns of the SPDR S&P 500 ETF (SPY). The estimates* $\widetilde{XES}^{\star}_{\widehat{\tau}'_n(p_n)}$ *as rainbow curve and* $\overline{XES}^{\star}_{\widehat{\tau}'_n(p_n)}$ *as dashed black curve, along with the asymptotic* 95% *confidence intervals* $\widetilde{CI}_{0.95}(k)$ *in dotted blue lines and* $\overline{CI}_{0.95}(k)$ *in solid grey lines. The sample maximum* $Y_{n,n}$ *indicated in horizontal dashed pink line.*

application to financial returns, the potential serial dependence may then affect the resulting asymptotic confidence intervals. A practical solution to reduce substantially the potential serial dependence in this particular dataset is by using weekly loss returns (corresponding to Wednesdays) in the same sample period. Given the length of the crisis period (38 trading weeks), we perform our extreme value estimation here only for the full period ($n = 516$), the pre-crisis period ($n = 219$) and the post-crisis period ($n = 259$). For each considered period, the final estimates of the tail index γ and the weights α and β are reported in Table 3. The plots of the ES estimates $\widetilde{\text{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ and $\overline{\text{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$ against k are graphed in Figure 7, and the final ES levels along with their corresponding confidence bands are displayed in Table 4. By comparing the obtained estimates before the crisis period (third rows in Tables 2 and 4), it may be seen that the results are quantitatively robust to the change from daily to weekly data. However, both the full period and the post-crisis period suggest fatter tails when moving to weekly data, as indicated by the new expectHill estimates in Table 3.

9 Final comments and perspectives for future research

Let us point out the main conceptual results of this paper that provide a novel take on extreme value analysis using asymmetric least squares estimation. Under the model assumption (2) of heavy-tailed distributions with tail index $\gamma < 1/2$, what first distinguishes our contribution is that it introduces a pure expectile-based estimator $\widetilde{\gamma}_{\tau_n}$ of γ in (6), where τ_n is the tuning parameter to be selected in practice. This new estimator has the same form as the traditional quantile-based Hill estimator $\widehat{\gamma}_{\tau_n}$ in (7), with the tail empirical quantile process in (7) replaced by its expectile analogue. While the asymmetric least squares estima-

Period considered	n	$\widehat{\gamma}_{1-k/n}$	$\overline{\gamma}_{1-k/n}$	$\overline{\alpha}_{1-k/n}$	$\overline{\beta}_{1-k/n}$
Full period	2,497	0.3585	0.3530	0.9235	1.1677
Before Crisis	1,053	0.2722	0.2844	0.7695	0.6892
During Crisis	185	0.2445	0.2780	0.7427	0.6118
After Crisis	1,259	0.2523	0.2617	0.6273	0.4392

Table 1: *Final estimates of the tail index γ and the combination parameters α and β, based on daily loss returns of the SPDR S&P 500 ETF (SPY) over the full 10-years period (2004-2014) and three sub-periods: Before Crisis (2004-2008), During Crisis (2008-2009) and After Crisis (2009-2014).*

Period	$\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$	$\widetilde{\mathrm{CI}}_{0.95}$	$\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$	$\overline{\mathrm{CI}}_{0.95}$	$Y_{n,n}$
Full period	0.0652	(0.0446, 0.0874)	0.0690	(0.0448, 0.0922)	0.0919
Before Crisis	0.0359	(0.0259, 0.0464)	0.0383	(0.0271, 0.0505)	0.0358
During Crisis	0.1169	(0.0784, 0.1597)	0.1277	(0.0842, 0.1712)	0.0919
After Crisis	0.0485	(0.0334, 0.0664)	0.0496	(0.0333, 0.0665)	0.0647

Table 2: *Final ES levels with the 95% confidence intervals and the sample maxima. Results based on daily loss returns, with $p_n = 1 - \frac{1}{n}$.*

Period considered	n	$\widehat{\gamma}_{1-k/n}$	$\overline{\gamma}_{1-k/n}$	$\overline{\alpha}_{1-k/n}$	$\overline{\beta}_{1-k/n}$
Full period	516	0.39094	0.39091	0.9770	1.0625
Before Crisis	219	0.2182	0.2547	0.6506	0.3936
After Crisis	259	0.4316	0.4313	0.9940	1.0158

Table 3: *Final estimates of γ, α and β, based on weekly loss returns over the full period (2004-2014) and the two sub-periods: Before (2004-2008) and After (2009-2014) Crisis.*

Period	$\widetilde{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$	$\widetilde{\mathrm{CI}}_{0.95}$	$\overline{\mathrm{XES}}^{\star}_{\widehat{\tau}'_n(p_n)}$	$\overline{\mathrm{CI}}_{0.95}$	$Y_{n,n}$
Full period	0.0609	(0.0305, 0.0872)	0.0748	(0.0328, 0.0925)	0.0424
Before Crisis	0.0364	(0.0232, 0.0490)	0.0378	(0.0248, 0.0507)	0.0292
After Crisis	0.0651	(0.0238, 0.1076)	0.0810	(0.0330, 0.1278)	0.0420

Table 4: *Final ES levels with the 95% confidence intervals and the sample maxima. Results based on weekly loss returns, with $p_n = 1 - \frac{1}{n}$.*

tor $\widetilde{\gamma}_{\tau_n}$ provides smoother and more stable plots against τ_n, and hence may be less sensitive to the choice of τ_n, the Hill estimator $\widehat{\gamma}_{\tau_n}$ may have lower bias in certain situations. In order to obtain the best of both $\widetilde{\gamma}_{\tau_n}$ and $\widehat{\gamma}_{\tau_n}$, we consider their weighted combination $\overline{\gamma}_{\tau_n}(\alpha)$ in (8) that we call *expectHill estimator*. This is the first work to actually implement the idea of tail index estimation, in Section 3, based either on pure asymmetric least squares estimates or their combination with top order statistics. Our expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ of γ is then used as the basis for estimating extreme expectiles and expected shortfall (ES) in Sections 4 and 5. We first estimate tail expectiles ξ_{τ_n} with the same intermediate level τ_n as in $\overline{\gamma}_{\tau_n}(\alpha)$. The proposed intermediate expectile estimator $\overline{\xi}_{\tau_n}(\alpha,\beta)$ in (11) is itself a weighted combination of the two competing nonparametric and semiparametric estimators $\widetilde{\xi}_{\tau_n}$ and $\widehat{\xi}_{\tau_n}(\alpha)$. The estimator $\overline{\xi}_{\tau_n}(\alpha,\beta)$ is then extrapolated to the very far tail in (14). Built on the resulting extrapolated expectile estimator in (14) and on the expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$, we construct two estimators for the tail expectile-based form of ES in (16) and (17). Finally, we develop composite versions of the latter estimators to estimate the conventional quantile-based ES itself.

Our contribution in Sections 4 and 5 extends and substantially improves on the ideas of Daouia *et al.* [15] in two directions. First, in contrast to [15] where tail extrapolation is restricted to the Hill estimator (*i.e.* $\alpha = 1$), we use in our setup the generalized expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ to derive extreme expectile and ES estimators, with variance-optimal weights α that can be even negative for low values of γ. The development of the asymptotic theory of our extrapolated estimators requires a different treatment based on the joint convergence of the expectHill estimator with intermediate sample quantiles and expectiles, as established in Theorem 2. Second and most importantly, unlike [15] where $\alpha = 1$ and only an *a priori* pre-specified set of weights β is used in practice, we suggest in the present paper a simple practical choice of the combination weights α in the expectHill estimator $\overline{\gamma}_{\tau_n}(\alpha)$ and β in the intermediate expectile estimator $\overline{\xi}_{\tau_n}(\alpha,\beta)$. This choice is based on the minimization of the asymptotic variance of $\overline{\gamma}_{\tau_n}(\alpha)$ over α and of $\overline{\xi}_{\tau_n}(\alpha,\beta)$ over β. The obtained variance-optimal weights $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$ are consistently estimated by $\overline{\alpha}_{\tau_n}$ and $\overline{\beta}_{\tau_n}$, respectively, as described in Section 6. Interestingly, the adaptive estimators $\overline{\gamma}_{\tau_n}(\overline{\alpha}_{\tau_n})$ and $\overline{\xi}_{\tau_n}(\overline{\alpha}_{\tau_n},\overline{\beta}_{\tau_n})$, obtained by substituting the estimated values $\overline{\alpha}_{\tau_n}$ and $\overline{\beta}_{\tau_n}$ in place of α and β, inherit the same asymptotic distributions as their analogues $\overline{\gamma}_{\tau_n}(\alpha(\gamma))$ and $\overline{\xi}_{\tau_n}(\alpha(\gamma),\beta_{\alpha(\gamma)}(\gamma))$ that are obtained by substituting in the true optimal weights $\alpha(\gamma)$ and $\beta_{\alpha(\gamma)}(\gamma)$, see Remark 6. This choice of (α,β) guarantees that the resulting adaptive estimators are low-variance and hence have good stability properties. The bias of these estimators might, however, be high. To further investigate this adaptive estimation problem, a subsequent step is to minimize an estimate of the Asymptotic Mean-Squared Error (AMSE) of the proposed combinations instead of their asymptotic variances. As we argued below Theorems 2 and 3, this necessitates the estimation of the second-order extreme value parameters ρ and λ_1 which is notoriously difficult: in addition to the Introduction of [8], we refer to the review in Section 5 of [28]. It should be noted here that ρ estimators typically have a lower rate of convergence than tail index estimators, see *e.g.* p.2638 in [27] and p.298 in [29]. This suggests that ρ estimators are in general quite volatile. In particular, the choice of the effective sample size in second-order parameter estimation is known to be hard. Constructing adaptive expectHill estimators having an optimal AMSE that also perform well in practice is thus a difficult question which is worthy of future research.

We closed Section 8 by an application to financial data, where the potential serial dependence may affect the asymptotic confidence intervals derived in our theorems for independent and identically distributed random variables. Similarly to our extreme value analysis in Daouia *et al.* [14], our convergence results may work under serial dependence with enlarged asymptotic variances. A theoretical question to be solved before adapting our results to a time series framework is, of course, to first prove Gaussian approximations of the tail empirical expectile process similar to those of Proposition 1 in such a framework. Following the method of proof of [15], it is reasonable to look for a Gaussian approximation of the tail empirical quantile process as a starting point. Such an approximation is proven in [17, 18, 19] in a framework of $\beta-$mixing observations. As [19] shows, this assumption covers, among others, processes obtained by solving certain stochastic recurrence equations (including ARCH processes) and ARMA models under reasonably general conditions. Proving that this result on the tail empirical quantile process alone can be strengthened to a joint Gaussian approximation of the tail empirical quantile and expectile processes, similarly to Proposition 1, is the key to be able to write analogs of our Theorems 1–6 in the $\beta-$mixing framework. We leave this theoretically challenging endeavor to future research as well.

Supplementary Material

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

Acknowledgements

The authors would like to thank three referees, the Associate Editor and the Co-Managing Editor for their very valuable comments, which led to an improved presentation of the paper. We are also grateful to Nour Meddahi who kindly sent us the empirical data. This research was supported by the French National Research Agency under the grant ANR-19-CE40-0013-01/ExtremReg project. S. Girard gratefully acknowledges the support of the Chair Stress Test, Risk Management and Financial Steering, led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas, as well as the support of the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02).

References

[1] Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking & Finance*, **26**, 1505–1518.

[2] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent Measures of Risk, *Mathematical Finance*, **9**, 203–228.

[3] Bates, J.M. and Granger, C.W.J. (1969). The combination of forecasts, *Operations Research Quarterly*, **20**, 451–468.

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

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

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

[7] Breckling, J. and Chambers, R. (1988). M-quantiles, *Biometrika*, **75**, 761–771.

[8] Cai, J.-J., de Haan, L. and Zhou, C. (2013). Bias correction in extreme value statistics with index around zero, *Extremes*, **16**, 173–201.

[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, *Journal of the Royal Statistical Society: Series B*, **77**, 417–442.

[10] Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography, *International Journal of Forecasting*, **5**, 559–583.

[11] Crane, D.B. and Crotty, J.R. (1967). A two-stage forecasting model: Exponential smoothing and multiple regression, *Management Science*, **13**, B501–B507.

[12] 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.

[13] Daouia, A., Girard, S. and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles, *Journal of the Royal Statistical Society: Series B*, **80**, 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] Daouia, A., Girard, S. and Stupfler, G. (2020). Tail expectile process and risk assessment, *Bernoulli*, **26**, 531–556.

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

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

[18] Drees, H. (2002). Tail empirical processes under mixing conditions, in Dehling, H.G., Mikosch, T., Sørensen, M. (eds.), *Empirical Processes Techniques for Dependent Data*, 325–342.

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

[20] Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot, *Annals of Statistics*, **28**, 254–274.

[21] Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings, *Journal of the Royal Statistical Society: Series B*, **78**, 505–562.

[22] El Methni, J., Gardes, L. and Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions, *Scandinavian Journal of Statistics*, **41**, 988–1012.

[23] El Methni, J. and Stupfler, G. (2017). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions, *Statistica 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-Verlag, Berlin.

[26] Gabaix, X. (2009). Power Laws in Economics and Finance, *Annual Review of Economics*, **1**, 255–294.

[27] Goegebeur, Y., Beirlant, J. and de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics, *Journal of Statistical Planning and Inference*, **140**, 2632–2652.

[28] Gomes, M.I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review, *International Statistical Review*, **83**, 263–292.

[29] Gomes, M.I., Pestana, D. and Caeiro, F. (2009). A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator, *Statistics & Probability Letters*, **79**, 295–303.

[30] Hall, P. and Welsh, A.W. (1985). Adaptive estimates of parameters of regular variation, *Annals of Statistics*, **13**, 331–341.

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

[32] Holzmann, H. and Klar, B. (2016). Expectile asymptotics, *Electronic Journal of Statistics*, **10**, 2355–2371.

[33] Hua, L. and Joe, H. (2011). Second order regular variation and conditional tail expectation of multiple risks, *Insurance: Mathematics and Economics*, **49**, 537–546.

[34] Kim, J. and Meddahi, N. (2019). Volatility Regressions with Fat Tails, preprint, available at https://nourmeddahi.github.io/KM_1.pdf.

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

[36] Krätschmer, V. and Zähle, H. (2017). Statistical inference for expectile-based risk measures, *Scandinavian Journal of Statistics*, **44**, 425–454.

[37] Kuan, C.-M., Yeh, J.-H. and Hsu, Y.-C. (2009). Assessing value at risk with CARE, the Conditional Autoregressive Expectile models, *Journal of Econometrics*, **150**, 261–270.

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

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

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

[41] Resnick, S. and Stărică, C. (1997). Smoothing the Hill estimator, *Advances in Applied Probability*, **29**, 271–293.

[42] Rockafellar, R.T. and Uryasev, S. (2000). Optimization of conditional value-at-risk, *Journal of Risk*, **2**, 21–41.

[43] Rockafellar, R.T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance*, **26**, 1443–1471.

[44] Sobotka, F. and Kneib, T. (2012). Geoadditive expectile regression, *Computational Statistics & Data Analysis*, **56**, 755–767.

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

[46] Timmermann, A. (2006). Forecast combinations, *Handbook of Economic Forecasting* (Chapter 4), **1**, 135–196.

[47] Weiss, C.E., Raviv, E. and Roetzer, G. (2018). Forecast Combinations in R using the ForecastComb Package, *The R Journal*, **10**, 262–281.

[48] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, **73**, 812–815.

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