WORKING PAPERS

N° TSE-566

April 2015

"Assessing coherent Value-at-Risk and expected shortfall with extreme expectiles"

Abdelaati Daouia, Stéphane Girard and Gilles Stupfler

Toulouse School of Economics

Assessing coherent Value-at-Risk and expected shortfall with extreme expectiles

Abdelaati Daouia[a1], Stéphane Girard[b] and Gilles Stupfler[c]

[a] Toulouse School of Economics, University of Toulouse
[b] INRIA Grenoble Rhône-Alpes / LJK Laboratoire Jean Kuntzmann - MISTIS
[c] GREQAM, Aix Marseille Université

Abstract

The class of quantiles lies at the heart of extreme-value theory and is one of the basic tools in risk management. The alternative family of expectiles is based on squared rather than absolute error loss minimization. The flexibility and virtues of these least squares analogues of quantiles are now well established in actuarial science, econometrics and statistical finance. Both quantiles and expectiles were embedded in the more general class of M-quantiles as the minimizers of a generic asymmetric convex loss function. It has been proved very recently that the only M-quantiles that are coherent risk measures are the expectiles. Also, in contrast to the quantile-based expected shortfall, expectiles benefit from the important property of elicitability that corresponds to the existence of a natural backtesting methodology. Least asymmetrically weighted squares estimation of expectiles did not, however, receive yet as much attention as quantile-based risk measures from the perspective of extreme values. In this article, we develop new methods for estimating the Value-at-Risk and expected shortfall measures via high expectiles. We focus on the challenging domain of attraction of heavy-tailed distributions that better describe the tail structure and sparseness of most actuarial and financial data. We first estimate the intermediate large expectiles and then extrapolate these estimates to the very far tails. We establish the limit distributions of the proposed estimators when they are located in the range of the data or near and even beyond the maximum observed loss. Monte Carlo experiments and a concrete application are given to illustrate the utility of extremal expectiles as an efficient instrument of risk protection.

Key words : Asymmetric squared loss; Coherent Value-at-Risk; Expected shortfall; Expectiles; Extrapolation; Extreme value theory; Heavy tails.

1 Introduction

Growing interest in modern tail analysis has focused on the concept of expectiles. This concept is a least squares analogue of quantiles, which summarizes the underlying distribution of an asset return or a loss variable Y in much the same way that quantiles do. It is a natural generalization of the usual mean $\mathbb{E}(Y)$, which bears the same relationship to

[1]A. Daouia's research was supported by the Toulouse School of Economics Individual Research Fund (IRF/Daouia-20125)

this noncentral moment as the class of quantiles bears to the median. Both expectiles and quantiles are found to be useful descriptors of the higher and lower regions of the data points in the same way as the mean and median are related to their central behavior. Koenker and Bassett (1978) elaborated an absolute error loss minimization framework to define quantiles, which successfully extends the conventional definition of quantiles as left-continuous inverse functions. Instead, Newey and Powell (1987) substituted the "absolute deviations" in the asymmetric loss function of Koenker and Bassett with "squared deviations" to obtain the population expectile of order $\tau \in (0,1)$ as the minimizer

$$\xi_\tau = \mathrm{argmin}_{\theta\in\mathbb{R}}\mathbb{E}\left\{\eta_\tau(Y-\theta) - \eta_\tau(Y)\right\},$$

where $\eta_\tau(y) = |\tau - \mathbb{I}(y \leq 0)|\, y^2$, with $\mathbb{I}(\cdot)$ being the indicator function. A natural estimator of ξ_τ is obtained from a given sample $(Y_1, \ldots, Y_n)$ via minimization of the asymmetrically weighted squared residuals criterion $\sum_{i=1}^n \eta_\tau(Y_i - \theta)$ with respect to θ. The first advantage of this asymmetric least squares approach relative to quantiles lies in the computational expedience of expectiles using only scoring or iteratively-reweighted least squares (though efficient linear programming routines are nowadays available for quantiles). The second advantage is that sample expectiles are more efficient as the weighted least squares rely on the distance to data points, while empirical quantiles only utilize the information on whether an observation is below or above the predictor (because they are based on absolute rather than squared error loss minimization). This benefit in terms of increased efficiency comes at the price of decreased robustness against outliers. Expectiles are indeed more sensitive to the magnitude of extremes than quantiles are. Henceforth, the choice between expectiles and quantiles usually depends on the application at hand, as is the case in the duality between the mean and the median. In this paper, we shall discuss how extreme expectiles can serve as a more efficient instrument of risk protection than the traditional quantile-based measures.

The classical mean being a special case ($\tau = \frac{1}{2}$) of expectiles, this indicates that the latter are closer to the notion of explained variance in least squares estimation. This is of particular relevance in the context of regression analysis. Furthermore, sample expectiles provide a class of smooth curves as functions of the level τ, which is not the case for sample quantiles. Finally, inference on expectiles is much easier than inference on quantiles. Standard results from M-estimation theory show that, under mild moment conditions, the empirical τ-expectiles are $\sqrt{n}$-consistent and asymptotically normal with mean ξ_τ and variance described in Abdous and Remillard (1995). Unlike expectiles, the estimation of the asymptotic variance/covariance of quantile estimators involves the tedious "smoothing" of the values of the density function at those quantiles (*i.e.*, sparsity function). The only drawback of asymmetric least squares relative to quantiles is that expectiles do not have an intuitive interpretation as direct as quantiles. The τ-quantile determines the point below which $100\tau\%$ of the mass of Y lies, while the τ-expectile specifies the position such that the average distance from the data below that position to itself is $100\tau\%$, *i.e.*,

$$\tau = \mathbb{E}\left\{|Y - \xi_\tau| \mathbb{1}(Y \leq \xi_\tau)\right\} / \mathbb{E}\left|Y - \xi_\tau\right|.$$

This reduced interpretability of expectiles should not be viewed as a serious disadvantage however, since Efron (1991) already suggested an elegant device to recover quantiles and their strong intuitive appeal from a set of expectiles. Koenker (1993) derived a first example of a distribution where the τth expectile and quantile coincide for all $\tau \in (0, 1)$. Also, Jones (1994) established that expectiles are precisely the quantiles, not of the original distribution, but of a related transformation. Abdous and Remillard (1995) proved that quantiles and expectiles of the same distribution coincide under the hypothesis of weighted-symmetry. In case of location and scale family of distributions, Yao and Tong (1996) showed that quantiles are identical to expectiles, but with different orders τ. Very recently, Zou (2014) has extended Koenker's argument to more generic distributions for which expectiles and quantiles coincide.

Both families of quantiles and expectiles were embedded in the more general class of M-quantiles defined by Breckling and Chambers (1988) as the minimizers of a generic asymmetric convex loss function. This class is one of the basic tools in statistical applications as has been well reflected by the large amount of recent literature on M-quantiles. The properties of these statistical M-functionals have been extensively investigated during the last decade, especially from the point of view of the axiomatic theory of risk measures. In particular, Bellini (2012) has shown that expectiles with $\tau \geq \frac{1}{2}$ are the only M-quantiles that are isotonic with respect to the increasing convex order, in accordance with the results of Baüerle and Müller (2006). More recently, Bellini *et al.* (2014) have proved that the only M-quantiles that are coherent risk measures are the expectiles. They have also established that expectiles are more conservative than the usual quantiles for extremely heavy-tailed distributions, and are robust in the sense of lipschitzianity with respect to the Wasserstein metric. Perhaps most importantly, expectiles benefit from the prominent property of elicitability that corresponds to the existence of a natural backtesting methodology. The relevance of this property in connection with backtesting has been addressed by Embrechts and Hofert (2013) while its relationship with coherency has been very recently discussed in Ziegel (2014). In contrast to expectiles, the famous expected shortfall, which is the only risk measure that possesses a highly desirable combination of properties such as coherency, suffers from the vexing defect of lack of elicitability [Gneiting (2011)]. Expectiles are becoming increasingly popular also in the econometric literature as can be seen, for instance, from Kuan *et al.* (2009), De Rossi and Harvey (2009), Embrechts and Hofert (2013) and the references therein.

Although least asymmetrically weighted squares estimation of expectiles dates back to Newey and Powell (1987) in case of linear regression, it recently regained growing interest in the context of nonparametric, semiparametric and more complex models [see for example Schulze Waltrup *et al.* (2015)]. Attention has been, however, restricted to ordinary expectiles of fixed order τ staying away from the tails of the underlying distribution. The purpose of this paper is to extend their estimation and asymptotic theory far enough into the tails. This translates into considering the expectile level $\tau = \tau_n \to 0$ or $\tau_n \to 1$ as the sample size n goes to infinity. We focus on high expectiles ξ_{τ_n} in the challenging maximum domain of attraction of Pareto-type (heavy-tailed) distributions, where standard expectile estimates at the tails are often unstable due to data sparsity. Specifically, we first estimate the intermediate tail expectiles of order $\tau_n \to 1$ such that $n(1 - \tau_n) \to \infty$, and then extrapolate these estimates to the very extreme expectile level τ_n which approaches one at an arbitrarily fast rate in the sense that $n(1 - \tau_n) \to c$, for some constant c. Two such estimation methods are considered. One is indirect, based on the use of asymptotic approximations involving intermediate quantiles, and the other relies directly on least asymmetrically weighted squares estimation. Similar considerations evidently apply to the case $\tau_n \to 0$.

There are many important applications in finance, insurance and econometrics, where extending expectile estimation and large sample theory further into the tails is a highly welcome development. Motivating examples include big financial losses, highest bids in auctions, large claims in (re)insurance, and high medical costs, to name a few. To our knowledge, this is the first work actually joining together the expectile perspective on asymmetric least squares with the tail restrictions of modern extreme-value theory.

We organize this paper as follows. Section 2 revisits the basic properties of the coherent expectile-based Value-at-Risk (VaR) and discusses its connection with the conventional quantile-based VaR for high levels of prudentiality. Section 3 presents the two estimation methods of intermediate and extreme expectiles. Section 4 returns to the connection of expectiles with the expected shortfall and discusses new interesting developments. Section 5 compares the proposed estimation methods of the expectile-based VaR and expected shortfall via Monte-Carlo experiments. Section 6 provides a concrete application to the Society of Actuaries Group Medical Insurance Large Claims Database. Section 7 concludes and the Appendix collects the proofs.

2 Setting and basic properties

In this paper, the generic financial position Y is a real-valued random variable, and the available data $\{Y_1, Y_2, \ldots\}$ are the negative of a series of financial returns. As such, a positive value of $-Y$ denotes a profit and a negative value denotes a loss. This implies that the right-tail of the distribution of Y corresponds to the negative of extreme losses. Following Newey and Powell (1987), the expectile ξ_τ of order $\tau \in (0,1)$ of the random variable Y is defined as the minimizer of a piecewise-quadratic loss function. Namely,

$$\xi_\tau = \text{argmin}_{\theta \in \mathbb{R}} \left\{ \tau \mathbb{E}\left[(Y-\theta)_+^2 - Y_+^2 \right] + (1-\tau)\mathbb{E}\left[(Y-\theta)_-^2 - Y_-^2 \right] \right\},$$

where $y_+ := \max(y,0)$ and $y_- := \max(-y,0)$. The first-order necessary condition for optimality related to this problem can be written in several ways such as

$$\tau \mathbb{E}\left[(Y-\xi_\tau)_+ \right] = (1-\tau)\mathbb{E}\left[(Y-\xi_\tau)_- \right]$$

or equivalently

$$\xi_\tau - \mathbb{E}(Y) = \frac{2\tau-1}{1-\tau}\,\mathbb{E}\left[(Y-\xi_\tau)_+ \right]. \tag{1}$$

These equations have a unique solution for all Y such that $\mathbb{E}|Y| < \infty$ [*i.e.* $Y \in L^1$]. Thenceforth expectiles of a distribution function F_Y with finite absolute first moment are well-defined. They summarize the distribution function in much the same way that the quantiles $q_\tau := F_Y^{-1}(\tau) = \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\}$ do. A justification for their use to describe distributions and their tails may be based on the following collection of elementary properties [Newey and Powell (1987), Abdous and Remillard (1995) and Bellini *et al.* (2014)]:

(i) Law invariance: a distribution is uniquely defined by its class of expectiles in the sense that the laws of two integrable random variables $Y \in L^1$ and $\widetilde{Y} \in L^1$ are identical if and only if $\xi_{Y,\tau} = \xi_{\widetilde{Y},\tau}$ for every $\tau \in (0,1)$.

(ii) Location and scale equivariance: the τth expectile of the linear transformation $\widetilde{Y} = a + bY$, where $a, b \in \mathbb{R}$, satisfies

$$\xi_{\widetilde{Y},\tau} = \begin{cases} a + b\,\xi_{Y,\tau} & \text{if} \quad b > 0 \\ a + b\,\xi_{Y,1-\tau} & \text{if} \quad b \leq 0\,. \end{cases}$$

(iii) Constancy: if $Y = c$ with probability 1, for some constant c (*i.e.* Y is degenerate), then $\xi_{Y,\tau} = c$ for any τ.

(iv) Strict monotonicity in τ: if $\tau_1 < \tau_2$, with $\tau_1, \tau_2 \in (0,1)$, then $\xi_{\tau_1} < \xi_{\tau_2}$. Also, the function $\tau \mapsto \xi_\tau$ maps $(0,1)$ onto its range $\{y \in \mathbb{R} : 0 < F_Y(y) < 1\}$.

(v) Preserving of stochastic order: if $Y \leq \tilde{Y}$ with probability 1, then $\xi_{Y,\tau} \leq \xi_{\tilde{Y},\tau}$ for any τ.

(vi) Subadditivity: for any variables $Y, \tilde{Y} \in L^1$, $\xi_{Y+\tilde{Y},\tau} \leq \xi_{Y,\tau} + \xi_{\tilde{Y},\tau}$ for all $\tau \geq \frac{1}{2}$. Also, $\xi_{Y+\tilde{Y},\tau} \geq \xi_{Y,\tau} + \xi_{\tilde{Y},\tau}$ for all $\tau \leq \frac{1}{2}$.

(vii) Lipschitzianity w.r.t. to the Wasserstein distance: for all $Y, \tilde{Y} \in L^1$ and all $\tau \in (0,1)$, it holds that $\left|\xi_{Y,\tau} - \xi_{\tilde{Y},\tau}\right| \leq \tilde{\tau} \cdot d_W(Y, \tilde{Y})$, where $\tilde{\tau} = \max\left\{\frac{\tau}{1-\tau}, \frac{1-\tau}{\tau}\right\}$ and

$$d_W(Y, \tilde{Y}) = \int_{-\infty}^{\infty} |F_Y(y) - F_{\tilde{Y}}(y)| dy = \int_0^1 |F_Y^{-1}(t) - F_{\tilde{Y}}^{-1}(t)| dt.$$

(viii) Sensitivity vs resistance: expectiles are very sensitive to the magnitude of extreme observations since their gross-error-sensitivity and rejection points are infinite. Whereas they are resistant to systematic rounding and grouping since their local-shift-sensitivity is bounded.

Of interest are the cases $\tau \uparrow 1$ and $\tau \downarrow 0$, which lead to access, respectively, the upper and lower endpoints of the distribution support. Like quantiles, expectiles are often argued in the statistical and econometric literatures to be useful descriptors of the distribution tails [see, *e.g.*, De Rossi and Harvey (2009), Schulze Waltrup *et al.* (2015) and the references therein]. The estimation of tail quantities is of utmost importance in extreme-value theory, especially when the distribution function of interest F_Y is heavy-tailed. This has been well reflected by the large amount of literature on related actuarial and financial fields of application [see, *e.g.*, Coles (2001) and Embrechts *et al.* (1987)]. A very important problem involves quantifying the "riskiness" implied by the return distribution under consideration. Greater variability of the financial position Y and particularly a heavier right-tail of its distribution F_Y necessitates a higher capital reserve for portfolios or price of the insurance risk. The most common risk measure used in banking and finance is the quantile-based Value-at-Risk (VaR) q_τ with a given confidence level τ, which is defined as the τth quantile of the underlying distribution F_Y. Recently, Kuan *et al.* (2009) have suggested and favored the use of the alternative expectile-based VaR ξ_τ. Similar to the definition of the quantile-VaR, this measure is understood as the maximal possible loss within a given holding period under the prudentiality level $\tau \geq \frac{1}{2}$. It has the important advantage of being coherent as it satisfies the added property of subadditivity. Also, taking ξ_τ as a margin (capital requirement), the level τ can be understood as the relative cost of the expected margin shortfall, as discussed in Kuan *et al.* (2009). In contrast, τ defines the tail probability when taking q_τ as a margin requirement.

The sign convention we have chosen for values of Y as the negative of returns implies that extreme losses correspond to levels τ close to one. Only Bellini *et al.* (2014) have described what happens for large expectiles ξ_τ and how they are linked to extreme quantiles q_τ when F_Y is attracted to the maximum domain of Pareto-type distributions with tail-index $0 < \gamma < 1$. According to Bingham *et al.* (1987), such a heavy-tailed distribution function can be expressed as

$$F_Y(y) = 1 - \ell(y) \cdot y^{-1/\gamma} \tag{2}$$

where $\ell(\cdot)$ is a slowly-varying function at infinity, *i.e*,

$$\lim_{y\to\infty} \frac{\ell(\lambda y)}{\ell(y)} = 1 \quad \text{for all} \quad \lambda > 0.$$

The extreme-value index γ tunes the tail heaviness of the distribution function F_Y. Note also that the moments of F_Y do not exist when $\gamma > 1$. For most applicational purposes in risk management, it has been found in previous studies that assumption (2) describes sufficiently well the tail structure of actuarial and financial data [see Embrechts *et al.* (2009)]. Writing $\overline{F}_Y := 1 - F_Y$, Bellini *et al.* (2014) have shown in the case $\gamma < 1$ that

$$\frac{\overline{F}_Y(\xi_\tau)}{\overline{F}_Y(q_\tau)} \sim \gamma^{-1} - 1 \quad \text{as} \quad \tau \to 1, \tag{3}$$

or equivalently $\frac{\overline{F}_Y(\xi_\tau)}{1-\tau} \sim \gamma^{-1} - 1$ as $\tau \to 1$. It follows that extreme expectiles ξ_τ are more spread than extreme quantiles q_τ when $\gamma > \frac{1}{2}$, whereas $\xi_\tau < q_\tau$ for all large τ when $\gamma < \frac{1}{2}$. The connection (3) between high expectiles and quantiles can actually be refined appreciably by considering the second-order version of the regular variation condition (2). Assume that the tail quantile function U of Y, namely the left-continuous inverse of $1/\overline{F}_Y$, satisfies the second-order condition indexed by (γ, ρ, A), that is, there exist $\gamma > 0$, $\rho \leq 0$, and a function $A(\cdot)$ converging to 0 at infinity and having constant sign such that

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

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

Here and in what follows, $(x^\rho - 1)/\rho$ is to be understood as $\log x$ when $\rho = 0$. The interpretation of this so-called extremal value condition can be found in de Haan and Ferreira (2006) along with abundant examples of commonly used families of continuous distributions satisfying $\mathcal{C}_2(\gamma, \rho, A)$.

Proposition 1. *Assume that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1$. Then*

$$\frac{\overline{F}_Y(\xi_\tau)}{1-\tau} = (\gamma^{-1} - 1)(1 + \varepsilon(\tau))$$

with

$$\varepsilon(\tau) = -\frac{(\gamma^{-1} - 1)^\gamma \mathbb{E}(Y)}{q_\tau}(1 + \mathrm{o}(1)) - \frac{(\gamma^{-1} - 1)^{-\rho}}{\gamma(1 - \rho - \gamma)} A((1-\tau)^{-1})(1 + \mathrm{o}(1)) \quad \textit{as} \quad \tau \uparrow 1.$$

Even more strongly, one can establish the precise bias term in the asymptotic approximation of (ξ_τ/q_τ) itself.

Corollary 1. *Assume that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds, with $0<\gamma<1$. If F_Y is strictly increasing, then*

$$\frac{\xi_\tau}{q_\tau} = (\gamma^{-1}-1)^{-\gamma}(1+r(\tau))$$

with

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

In practice, the tail quantities ξ_τ, q_τ and γ are unknown and only a sample of random copies $(Y_1,\ldots,Y_n)$ of Y is typically available. While extreme-value estimates of high quantiles and of the tail-index γ are used widely in applied work and investigated extensively in theoretical statistics, the problem of estimating ξ_τ, when $\tau=\tau_n\to 1$ at an arbitrary rate as $n\to\infty$, has not been addressed yet. Direct expectile estimates at the tails are incapable of extrapolating outside the data and are often unstable due to data sparseness. This motivated us to construct estimators of large expectiles ξ_{τ_n} and derive their limit distributions when they are located in the range of the data or near and even beyond the sample maximum. We shall assume the extended regular variation condition $\mathcal{C}_2(\gamma,\rho,A)$ to obtain some convergence results.

3 Estimation of the expectile-VaR

Our main objective in this section is to estimate the expectile-based VaR ξ_{τ_n} for high levels of prudentiality τ_n that may approach one at any rate, covering both scenarios of intermediate expectiles with $n(1-\tau_n)\to\infty$ and extreme expectiles with $n(1-\tau_n)\to c$, for some constant c. We assume that the available data consists of an n-tuple $(Y_1,\ldots,Y_n)$ of independent copies of Y, and denote by $Y_{1,n}\le\cdots\le Y_{n,n}$ their ascending order statistics.

3.1 Intermediate expectile estimation

Here, we first use an indirect estimation method based on intermediate quantiles, and then discuss a direct asymmetric least squares estimation method.

3.1.1 Estimation based on intermediate quantiles

The rationale for this first method relies on the regular variation property (2) and on the asymptotic equivalence (3). Given that $\overline{F}_Y$ is regularly varying at infinity with index $-1/\gamma$ [*i.e.* it satisfies, for any $x > 0$, the property $\overline{F}_Y(tx)/\overline{F}_Y(t) \to x^{-1/\gamma}$ as $t \to \infty$], it follows that U is regularly varying as well with index γ. Hence, (3) entails that

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

This is also an immediate consequence of Corollary 1. Therefore, for a suitable estimator $\widehat{\gamma}$ of γ, we may suggest estimating the intermediate expectile ξ_{τ_n} by

$$\widehat{\xi}_{\tau_n} := (\widehat{\gamma}^{-1} - 1)^{-\widehat{\gamma}}\, \widehat{q}_{\tau_n}, \quad \text{where} \quad \widehat{q}_{\tau_n} := Y_{n-\lfloor n(1-\tau_n)\rfloor, n}$$

and $\lfloor \cdot \rfloor$ stands for the floor function. This estimator parallels the intermediate quantile-VaR $\widehat{q}_{\tau_n}$ and crucially hinges on the estimated tail-index $\widehat{\gamma}$. Accordingly, it is more conservative than $\widehat{q}_{\tau_n}$ when $\widehat{\gamma} > \frac{1}{2}$, but more liberal when $\widehat{\gamma} < \frac{1}{2}$. A simple and widely used estimator of γ is given by the popular Hill estimator

$$\widehat{\gamma}_H = \frac{1}{k} \sum_{i=1}^{k} \log \frac{Y_{n-i+1,n}}{Y_{n-k,n}}, \tag{4}$$

where $k = k(n)$ is an intermediate sequence in the sense that $k(n) \to \infty$ such that $k(n)/n \to 0$ as $n \to \infty$. See, *e.g.*, Section 3.2 in de Haan and Ferreira (2006) for a detailed review of the properties of $\widehat{\gamma}_H$.

Next, we formulate conditions that lead to asymptotic normality for $\widehat{\xi}_{\tau_n}$.

Theorem 1. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $0 < \gamma < 1$, that $\tau_n \uparrow 1$ and $n(1-\tau_n) \to \infty$. Assume further that*

$$\sqrt{n(1-\tau_n)} \left(\widehat{\gamma} - \gamma, \frac{\widehat{q}_{\tau_n}}{q_{\tau_n}} - 1 \right) \xrightarrow{d} (\Gamma, \Theta). \tag{5}$$

If $\sqrt{n(1-\tau_n)}\, q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}\, A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$, then

$$\sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 \right) \xrightarrow{d} m(\gamma)\Gamma + \Theta - \lambda$$

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

$$\lambda := \gamma(\gamma^{-1} - 1)^{\gamma} \mathbb{E}(Y) \lambda_1 + \left(\frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \rho - \gamma} + \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho} \right) \lambda_2.$$

When using the Hill estimator (4) of γ with $k = n(1-\tau_n)$, sufficient regularity conditions for (5) to hold can be found in Theorems 2.4.1 and 3.2.5 in de Haan and Ferreira (2006, p.50 and p.74). Under these conditions, the limit distribution Γ is then Gaussian with mean $\lambda_2/(1-\rho)$ and variance γ^2, while Θ is the standard Gaussian distribution. Lemma 3.2.3 in de Haan and Ferreira (2006, p.71) shows that both Gaussian limiting distributions are independent.

Yet, a drawback to the resulting expectile estimator $\widehat{\xi}_{\tau_n}$ lies in its heavy dependency on the estimated quantile $\widehat{q}_{\tau_n}$ and tail-index $\widehat{\gamma}$ in the sense that the former may inherit the vexing defects of the latters. Note also that $\widehat{\xi}_{\tau_n}$ is asymptotically biased, which is not the case for $\widehat{q}_{\tau_n}$. Another efficient way of estimating ξ_{τ_n} is by joining together the least asymmetrically weighted squares estimation with the tail restrictions of modern extreme-value theory.

3.1.2 Asymmetric least squares estimation

Here, we consider estimating the expectile ξ_{τ_n} by its empirical counterpart defined through

$$\widetilde{\xi}_{\tau_n} = \text{argmin}_{u\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n} \eta_{\tau_n}(Y_i - u),$$

where $\eta_\tau(y) = |\tau - \mathbb{1}\{y \leq 0\}|y^2$ is the expectile check function. This minimizer can easily be calculated by applying the function *"expectile"* implemented in the R package *'expectreg'*. It is not hard to verify that

$$\begin{aligned} \sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) &= \text{argmin}_{u\in\mathbb{R}}\psi_n(u) \qquad (6)\\ \text{with } \ \psi_n(u) &:= \frac{1}{2\xi_{\tau_n}^2}\sum_{i=1}^{n} \eta_{\tau_n}(Y_i - \xi_{\tau_n} - u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}) - \eta_{\tau_n}(Y_i - \xi_{\tau_n}). \end{aligned}$$

It follows from the continuity and the convexity of η_τ that (ψ_n) is a sequence of almost surely continuous and convex random functions. A result of Geyer (1996) [see also Theorem 5 in Knight (1999)] then states that to examine the convergence of the left-hand side term of (6), it is enough to investigate the asymptotic properties of the sequence (ψ_n). Built on this idea, we get the asymptotic normality of the least asymmetrically weighted squares estimator $\widetilde{\xi}_{\tau_n}$ by applying standard techniques involving sums of independent and identically distributed random variables.

Theorem 2. *Assume that* $0 < \gamma < 1/2$ *and* $\tau_n \uparrow 1$ *is such that* $n(1-\tau_n) \to \infty$. *Then*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2 V(\gamma)\right) \quad \textit{with} \quad V(\gamma) = \frac{2\gamma}{1-2\gamma}.$$

Interestingly, in contrast to Theorem 1, the limit distribution in Theorem 2 is derived without recourse to either the extended regular variation condition $\mathcal{C}_2(\gamma, \rho, A)$ or any bias condition. The mild assumption $0 < \gamma < 1/2$ suffices. Most importantly, unlike the indirect expectile estimator $\widehat{\xi}_{\tau_n}$, the new estimator $\widetilde{\xi}_{\tau_n}$ is asymptotically unbiased and does not hinge by construction on any particular type of quantile or tail-index estimators. It should also be clear that its asymptotic variance $\gamma^2 V(\gamma)$ can easily be compared with the asymptotic variance γ^2 obtained in the intermediate quantile estimation [see, *e.g.*, Theorem 2.4.1 in de Haan and Ferreira (2006)]. Specifically, $\gamma^2 V(\gamma) > \gamma^2$ for $\gamma > 1/4$, and $\gamma^2 V(\gamma) < \gamma^2$ for $\gamma < 1/4$.

3.2 Extreme expectile estimation

We now discuss the important issue of estimating extreme tail expectiles $\xi_{\tau'_n}$, where $\tau'_n \uparrow 1$ with $n(1-\tau'_n) \to c < \infty$ as $n \to \infty$. The basic idea is to extrapolate intermediate expectile estimates of order $\tau_n \to 1$, such that $n(1-\tau_n) \to \infty$, to the very extreme level τ'_n. This is achieved by transferring the elegant device of Weissman (1978) for estimating an extreme quantile to our expectile setup. Note that, in standard extreme-value theory and related fields of application, the levels τ'_n and τ_n are typically set to be $\tau'_n = 1 - p_n$ for a p_n much smaller than $\frac{1}{n}$, and $\tau_n = 1 - \frac{k(n)}{n}$ for an intermediate sequence of integers $k(n)$.

The model assumption of Pareto-type tails (2) means that $U(tx)/U(t) \to x^\gamma$ as $t \to \infty$, which in turn suggests that

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

for τ_n, τ'_n satisfying suitable conditions. By (A.3), we arrive at

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

This approximation motivates the following class of $\xi_{\tau'_n}$ plug-in estimators

$$\overline{\xi}^{\star}_{\tau'_n} \equiv \overline{\xi}^{\star}_{\tau'_n}(\tau_n) := \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}} \overline{\xi}_{\tau_n} \tag{7}$$

where $\widehat{\gamma}$ is an estimator of γ, and $\overline{\xi}_{\tau_n}$ stands for either the estimator $\widehat{\xi}_{\tau_n}$ or $\widetilde{\xi}_{\tau_n}$ of the intermediate expectile ξ_{τ_n}. As a matter of fact, we have

$$\frac{\overline{\xi}^{\star}_{\tau'_n}}{\overline{\xi}_{\tau_n}} = \frac{\hat{q}^{\star}_{\tau'_n}}{\widehat{q}_{\tau_n}} \tag{8}$$

where $\widehat{q}_{\tau_n} = Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$ is the intermediate quantile estimator introduced above, and $\hat{q}^{\star}_{\tau'_n}$ is the extreme Weissman quantile estimator defined as

$$\hat{q}^{\star}_{\tau'_n} \equiv \hat{q}^{\star}_{\tau'_n}(\tau_n) := \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{q}_{\tau_n}. \tag{9}$$

By making use of the identity (8), we show that $(\frac{\overline{\xi}^{\star}_{\tau'_n}}{\xi_{\tau'_n}} - 1)$ has the same limit distribution as $(\widehat{\gamma} - \gamma)$, but with a different scaling.

Theorem 3. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $\rho < 0$, that τ_n, $\tau'_n \uparrow 1$, with $n(1-\tau_n) \to \infty$ and $n(1-\tau'_n) \to c < \infty$. If moreover*

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \Delta \quad \text{and} \quad \sqrt{n(1-\tau_n)}(\widehat{\gamma} - \gamma) \xrightarrow{d} \Gamma,$$

with $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$, then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\overline{\xi}^{\star}_{\tau'_n}}{\xi_{\tau'_n}} - 1\right) \xrightarrow{d} \Gamma.$$

More specifically, when $\overline{\xi}_{\tau_n}$ in (7) is chosen to be the indirect intermediate expectile estimator $\widehat{\xi}_{\tau_n}$, the resulting extreme expectile estimator $\hat{\xi}^{\star}_{\tau'_n} := \overline{\xi}^{\star}_{\tau'_n}$, or equivalently,

$$\begin{aligned} \hat{\xi}^{\star}_{\tau'_n} &= \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{\xi}_{\tau_n} \\ &= \left(\widehat{\gamma}^{-1} - 1\right)^{-\widehat{\gamma}} \hat{q}^{\star}_{\tau'_n} \end{aligned} \tag{10}$$

satisfies the following general convergence result in view of Theorem 1.

Corollary 2. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $0 < \gamma < 1$ and $\rho < 0$, and that τ_n, $\tau'_n \uparrow 1$ with $n(1-\tau_n) \to \infty$ and $n(1-\tau'_n) \to c < \infty$. Assume further that*

$$\sqrt{n(1-\tau_n)}\left(\widehat{\gamma} - \gamma, \frac{\widehat{q}_{\tau_n}}{q_{\tau_n}} - 1\right) \xrightarrow{d} (\Gamma, \Theta).$$

If $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$, then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\hat{\xi}^{\star}_{\tau'_n}}{\xi_{\tau'_n}} - 1\right) \xrightarrow{d} \Gamma.$$

Likewise, Theorem 2 yields the following corollary for the alternative extreme expectile estimator

$$\widetilde{\xi}^{\star}_{\tau'_n} = \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}} \widetilde{\xi}_{\tau_n}, \tag{11}$$

obtained by substituting the least asymmetrically weighted squares estimator $\widetilde{\xi}_{\tau_n}$ in place of $\overline{\xi}_{\tau_n}$ in (7).

Corollary 3. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds with $0<\gamma<1/2$ and $\rho<0$, and that τ_n, $\tau_n'\uparrow 1$ with $n(1-\tau_n)\to\infty$ and $n(1-\tau_n')\to c<\infty$. If in addition*

$$\sqrt{n(1-\tau_n)}(\widehat{\gamma}-\gamma)\xrightarrow{d}\Gamma$$

and $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1}\to\lambda_1\in\mathbb{R}$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})\to\lambda_2\in\mathbb{R}$, then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{\xi}^{\star}_{\tau_n'}}{\xi_{\tau_n'}}-1\right)\xrightarrow{d}\Gamma.$$

4 Expectile-based expected shortfall

Due to systemic risk nowadays, the vast majority of market participants (Investors, Risk managers, Clearing houses, Regulators) are more concerned with the risk exposure to a catastrophic event that might wipe out an investment in terms of the size of potential losses. In this respect, the standard quantile-based VaR is often criticized for being too optimistic since it only depends on the frequency of tail losses and not on their values. Also, in most studies on actuarial and financial data, the realized values of the tail-index γ were found to be smaller than $\frac{1}{2}$, indicating thereby that the expectile-based VaR would be even more liberal than the quantile-VaR in these studies. In contrast, the pessimist market participants can expect the worst by resorting to the alternative popular expected shortfall measure.

4.1 Basic properties

The expected shortfall (ES) is defined as the conditional expectation of the financial position Y given that it exceeds the VaR. It is traditionally expressed at the $100(1-\omega)\%$ security level in terms of the ωth quantile as

$$\mathrm{QES}(\omega):=\mathbb{E}[Y|Y>q_\omega].$$

The standard interpretation of $-\mathrm{QES}(\omega)$ is as the expected return on the portfolio in the worst $100(1-\omega)\%$ of the cases. As pointed out earlier by Taylor (2008), this quantile-based ES is closely related to expectiles as well. Indeed, the solution

$$\xi_\tau=\mathrm{argmin}_{\theta\in\mathbb{R}}\mathbb{E}\left\{\eta_\tau(Y-\theta)-\eta_\tau(Y)\right\}$$

satisfies (see (1)):

$$\left(\frac{2\tau-1}{1-\tau}\right)\mathbb{E}[(Y-\xi_\tau)\mathbb{1}(Y>\xi_\tau)]=\xi_\tau-\mathbb{E}(Y).$$

The expectile ξ_τ is then determined by the properties of the expectation of the random variable Y conditional on Y exceeding ξ_τ. This suggests the alternative expectile-based ES

$$\begin{aligned}\mathrm{XES}(\tau) &:= \mathbb{E}[Y|Y>\xi_\tau]\\ &= \left(1+\frac{1-\tau}{(2\tau-1)\overline{F}_Y(\xi_\tau)}\right)\xi_\tau - \frac{1-\tau}{(2\tau-1)\overline{F}_Y(\xi_\tau)}\mathbb{E}(Y).\end{aligned}$$

For a continuous distribution, if the τ-expectile ξ_τ coincides with a ω-quantile q_ω, we have $\overline{F}_Y(\xi_\tau) = 1-\omega$ and the classical quantile-based ES can be rewritten as

$$\mathrm{QES}(\omega) = \left(1+\frac{1-\tau}{(2\tau-1)(1-\omega)}\right)\xi_\tau - \frac{1-\tau}{(2\tau-1)(1-\omega)}\mathbb{E}(Y).$$

Before moving to a deeper study of the expectile-ES, we first illustrate its sensitiveness to tail events by comparing its relative performance with the quantile-VaR, expectile-VaR and quantile-ES in the presence of catastrophic loss via Monte Carlo experiments. Similar to Duffie and Pan (1997) and Kuan *et al.* (2009), the data are independently drawn from $\mathcal{N}(0, 1/\sqrt{1-P})$ with probability $1-P$ or from $\mathcal{N}(c, 1/\sqrt{P})$ with probability P, where $P \in \{0.01, 0.005\}$ and $c \in [1, 50]$. Hence the observations shall be often taken from $\mathcal{N}(0, 1/\sqrt{1-P})$, but there may be infrequent catastrophic losses drawn from the more disperse scenario $\mathcal{N}(c, 1/\sqrt{P})$. For each c, we simulate 1000 samples of size $n = 1000$ and compute the Monte Carlo averages of the empirical versions of the four risk measures. The results are graphed in Figure 1, where $\tau = 0.95$, 0.99, from left to right for $P = 0.01$ (top panels), and $\tau = 0.99$, 0.995, for $P = 0.005$ (bottom panels). As expected, the expectile-VaR, the quantile-ES and the expectile-ES are affected by the extreme values from $\mathcal{N}(c, 1/\sqrt{P})$ for all c, whereas the quantile-VaR may not respond properly to such catastrophic losses. In particular, the expectile-ES is clearly more alert to infrequent disasters as its magnitude is overall larger than that of all the other risk measures. That eternal maxim of the cautious aunt and misanthropic uncle, *"expect the worst, and you won't be disappointed"* [Bassett *et al.* (2004)] can thus be transformed here into a precise computation via the expectile-based ES, with τ being a natural measure of the degree of pessimism.

As a matter of fact, by considering a Pareto-type distribution $F_Y(\cdot)$ with tail-index $\gamma < 1$ as above, we show that the choice between the expectile-ES and quantile-ES depends on the value at hand of $\gamma \lesseqgtr \frac{1}{2}$ as is the case in the duality between the expectile-VaR and quantile-VaR. More precisely, the theoretical expectile-ES defined earlier as $\mathrm{XES}(\tau) := \mathbb{E}[Y|Y>\xi_\tau]$ is more conservative (respectively, liberal) than the quantile-ES $\mathrm{QES}(\tau) := \mathbb{E}[Y|Y>q_\tau]$ for all τ large enough when $\gamma > \frac{1}{2}$ (respectively, $\gamma < \frac{1}{2}$).

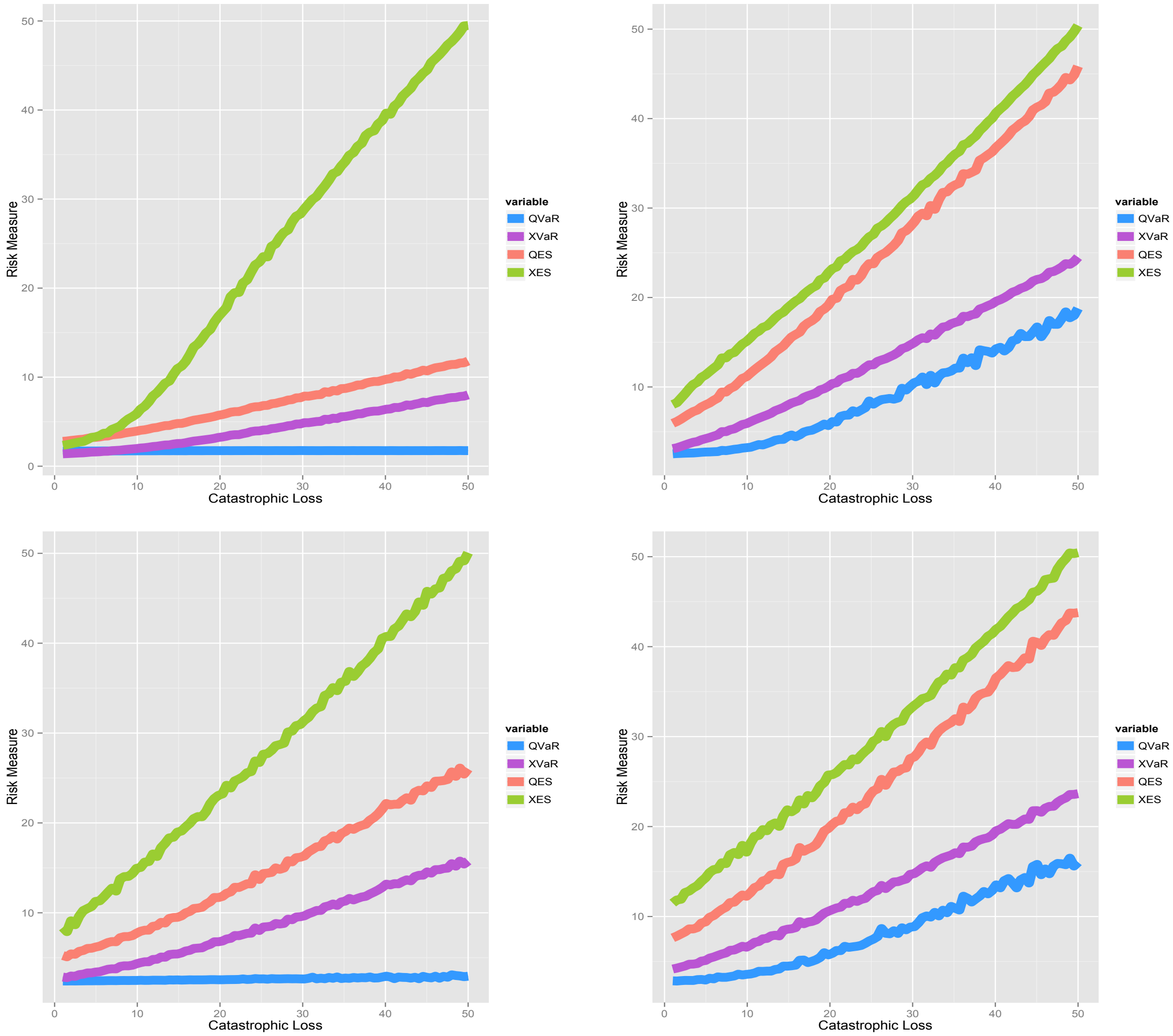


Figure 1: *The catastrophic loss sensitivity of empirical quantile-VaR (QVaR), expectile-VaR (XVaR), quantile-ES (QES) and expectile-ES (XES). From left to right and from top to bottom, we have* $(P, \tau) = (0.01, 0.95)$, $(0.01, 0.99)$, $(0.005, 0.99)$, $(0.005, 0.995)$.

Proposition 2. *Assume that the distribution of Y belongs to the Fréchet maximum domain of attraction with tail-index $\gamma < 1$, or equivalently, that condition* (2) *holds. Then, as $\tau \to 1$,*

$$\frac{XES(\tau)}{QES(\tau)} \sim \frac{\xi_\tau}{q_\tau} \quad \textit{and} \quad \frac{XES(\tau)}{\xi_\tau} \sim \frac{1}{1-\gamma}.$$

One may also establish, in the spirit of Proposition 1, a precise control of the remainder term which arises when using Proposition 2. This will prove to be quite useful when examining the asymptotic properties of the extreme expectile-ES estimators.

Proposition 3. *Assume that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1$. Then*

$$\frac{XES(\tau)}{\xi_\tau} = \frac{1}{1-\gamma}\left[1 + \frac{(\gamma^{-1}-1)^{-\rho}}{1-\rho-\gamma} A((1-\tau)^{-1})(1+\mathrm{o}(1)) + \mathrm{o}(q_\tau^{-1})\right] \quad \textit{as} \quad \tau \uparrow 1.$$

An influential paper in the literature by Artzner *et al.* (1999) provides an axiomatic foundation for coherent risk measures. Like the expectile-VaR, the ES satisfies all of their requirements (Translation invariance, Monotonicity, Subadditivity, and Positive homogeneity). However, unlike the expectile-VaR, the ES is not elicitable [Gneiting (2011), Ziegel (2014)]. The elicitability corresponds to the existence of a natural backtesting methodology, which allows to validate a given estimation procedure for a risk measure on historical data. In spite of the debate about the financial relevance of the elicitability property, a remarkable property which follows from Propositions 2 and 3 is that the pessimistic ES risk measure $\mathrm{XES}(\tau)$ is asymptotically proportional to the elicitable expectile-VaR ξ_τ. This connection can potentially be used to achieve the elicitability property for the desired large values of τ.

4.2 Estimation and asymptotics

According to Proposition 2, in the challenging maximum domain of attraction of Pareto-type distributions with tail-index $\gamma < 1$, the expectile-based ES is definitely more sensitive to the magnitude of the right heavy tails than is the expectile-based VaR, as $\tau \to 1$. Typically, financial institutions and insurance companies are interested in the extreme region $\tau = \tau_n' \uparrow 1$, with τ_n' being much larger than $(1-\frac{1}{n})$. The asymptotic equivalence $\mathrm{XES}(\tau_n') \sim (1-\gamma)^{-1}\xi_{\tau_n'}$, established in Proposition 2, suggests the following estimators of the expectile-ES:

$$\widehat{\mathrm{XES}}^{\star}(\tau_n') = (1-\widehat{\gamma})^{-1} \cdot \hat{\xi}^{\star}_{\tau_n'} \quad \text{and} \quad \widetilde{\mathrm{XES}}^{\star}(\tau_n') = (1-\widehat{\gamma})^{-1} \cdot \widetilde{\xi}^{\star}_{\tau_n'} \tag{12}$$

where $\hat{\xi}^{\star}_{\tau_n'}$ and $\widetilde{\xi}^{\star}_{\tau_n'}$ are the extreme expectile estimators defined above in (10)-(11), and $\widehat{\gamma}$ is an estimator of γ. Another option motivated by the second asymptotic equivalence $\mathrm{XES}(\tau_n') \sim \frac{\xi_{\tau_n'}}{q_{\tau_n'}} \cdot \mathrm{QES}(\tau_n')$ would be to estimate $\mathrm{XES}(\tau_n')$ by

$$\widehat{\mathrm{XES}}^{\dagger}(\tau_n') = \hat{\xi}^{\star}_{\tau_n'} \cdot \frac{\widehat{\mathrm{QES}}^{\star}(\tau_n')}{\hat{q}^{\star}_{\tau_n'}} \quad \text{or} \quad \widetilde{\mathrm{XES}}^{\dagger}(\tau_n') = \widetilde{\xi}^{\star}_{\tau_n'} \cdot \frac{\widehat{\mathrm{QES}}^{\star}(\tau_n')}{\hat{q}^{\star}_{\tau_n'}} \tag{13}$$

for a suitable estimator $\widehat{\mathrm{QES}}^{\star}(\tau_n')$ of $\mathrm{QES}(\tau_n')$ [see, *e.g.*, El Methni *et al.* (2014)], with $\hat{q}^{\star}_{\tau_n'}$ being the extreme Weissman quantile estimator defined in (9). Our experience with real and simulated data indicates, however, that the estimates $\widehat{\mathrm{XES}}^{\star}(\tau_n')$ and $\widehat{\mathrm{XES}}^{\dagger}(\tau_n')$ [respectively, $\widetilde{\mathrm{XES}}^{\star}(\tau_n')$ and $\widetilde{\mathrm{XES}}^{\dagger}(\tau_n')$] point toward very similar results. We therefore restrict our theoretical treatment to the first versions given in (12). Our first asymptotic result is for the extreme XES estimator $\widehat{\mathrm{XES}}^{\star}(\tau_n')$:

Corollary 4. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds with $0<\gamma<1$ and $\rho<0$, and that τ_n, $\tau_n' \uparrow 1$ with $n(1-\tau_n)\to\infty$ and $n(1-\tau_n')\to c<\infty$.*

Assume further that

$$\sqrt{n(1-\tau_n)}\left(\widehat{\gamma}-\gamma, \frac{\widehat{q}_{\tau_n}}{q_{\tau_n}}-1\right) \xrightarrow{d} (\Gamma, \Theta).$$

If $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$ *and* $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$, *then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{XES}^{\star}(\tau_n')}{XES(\tau_n')}-1\right) \xrightarrow{d} \Gamma.$$

In what concerns the asymmetric least squares type of estimator $\widetilde{\text{XES}}^{\star}(\tau_n')$, we have the following result.

Corollary 5. *Assume that F_Y is strictly increasing, that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $0 < \gamma < 1/2$ and $\rho < 0$, and that τ_n, $\tau_n' \uparrow 1$ with $n(1-\tau_n) \to \infty$ and $n(1-\tau_n') \to c < \infty$. If in addition*

$$\sqrt{n(1-\tau_n)}(\widehat{\gamma}-\gamma) \xrightarrow{d} \Gamma$$

and $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$, *then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{XES}^{\star}(\tau_n')}{XES(\tau_n')}-1\right) \xrightarrow{d} \Gamma.$$

Both results are derived by noticing that, on the one hand, the extreme expectile estimators $\hat{\xi}^{\star}_{\tau_n'}$ and $\widetilde{\xi}^{\star}_{\tau_n'}$ converge at a slower rate than the estimator $\widehat{\gamma}$ in view of Corollaries 2 and 3. On the other hand, the nonrandom remainder term coming from the use of Proposition 2 can be controlled by applying Proposition 3, so detailed proofs are omitted.

5 Some simulation evidence

This section provides Monte-Carlo evidence that the direct estimation method is more efficient relative to the indirect method. Recall that the direct type estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ is obtained via least asymmetrically weighted squares (LAWS) estimation, while the indirect type estimator $\hat{\xi}^{\star}_{\tau_n'}$ results from a full plug-in procedure based on an asymptotic equivalence with intermediate quantiles.

To evaluate finite-sample performance of the presented extreme expectile estimators, we have considered $10,000$ replications for samples of size 100 and 1000 simulated from various Student's t-scenarios: t_3, t_5, t_7 and t_9. The t_5, t_7 and t_9-distributions fit heavy-tailed returns from financial market variables rather well in the case of independent and identically distributed data. However, the t_3-distribution may be viewed as a violent model of the empirical distribution of returns since its tails are too heavy [see, *e.g.*, Tsay (2002)].

We used in all our simulations the Hill estimator of the tail-index γ, the extreme level $\tau_n' = 0.995$ for $n = 100$ and $\tau_n' = 0.9994$ for $n = 1000$. The corresponding true extreme expectiles $\xi_{\tau_n'}$ can be calculated by the existing function "et(τ_n', df)" in the R package *'expectreg'*. In what concerns the intermediate levels τ_n involved in both estimators $\widetilde{\xi}^{\star}_{\tau_n'} \equiv \widetilde{\xi}^{\star}_{\tau_n'}(\tau_n)$ and $\hat{\xi}^{\star}_{\tau_n'} \equiv \hat{\xi}^{\star}_{\tau_n'}(\tau_n)$, we used the same considerations as in Ferreira et al. (2003). Namely, they always considered $\tau_n = 1 - \frac{k}{n}$ with the range of intermediate integers k, say, from $\log(n^{1-\varepsilon})$ to $n/\log(n^{1-\varepsilon})$, where $\varepsilon = 0.1$ [this restriction allows to reject too small values or those very near $n^{1-\varepsilon}$]. The value k can actually be viewed as the effective sample size for tail extrapolation. A larger k leads to estimators with more bias, while smaller k results in higher variance.

Figure 2 gives the root-MSE estimates for the Student's t-models, while Figure 3 gives the bias estimates for the same models. Each figure displays the evolution of the obtained Monte-Carlo results, for the two normalized estimators $\widetilde{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$ and $\hat{\xi}^{\star}_{\tau_n'}(k)/\xi_{\tau_n'}$, as functions of the sample fraction k. Table 1 reports the root-MSE and bias estimates obtained by using for each estimator the optimal value of k minimizing its MSE.

Our tentative conclusion from this exercise is that the indirect estimator $\hat{\xi}^{\star}_{\tau_n'}$ has a harder time with small samples, and this can be compensated by taking larger samples. Indeed, for $n = 100$, the direct estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ performs better than $\hat{\xi}^{\star}_{\tau_n'}$ in terms of both MSE and bias, whatever the thickness of the tails. Also, in contrast to the direct estimator's plot, the indirect one exhibits more volatility. In what concerns $n = 1000$, it seems that $\hat{\xi}^{\star}_{\tau_n'}$ is superior to $\widetilde{\xi}^{\star}_{\tau_n'}$ only in terms of MSE for slightly heavy tails (*i.e.* $df = 7, 9$), whereas the accuracy of $\widetilde{\xi}^{\star}_{\tau_n'}$ is more respectable for heavier tails (*i.e.* $df = 3, 5$), as can be seen from Table 1. It should be, however, clear that even in the favorable case to $\hat{\xi}^{\star}_{\tau_n'}$, where $n = 1000$ and $df \in \{7, 9\}$, the estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ has actually almost overall a smaller MSE except for a very small zone of values of k, as can be seen from Figure 2 (bottom-right panels). Due to the tightness of that zone, the detection of the optimal k which minimizes the MSE of $\hat{\xi}^{\star}_{\tau_n'}$ is hard to manage in practice. It may also be seen that most of the error is due to variance, the squared bias being much smaller in all cases. It is interesting that in almost all cases the bias was positive. This may be explained by the sensitivity of high expectiles to the magnitude of heavy tails, since they are based on "squared" error loss minimization.

We have also undertaken simulation experiments to evaluate the finite-sample performance of the presented expectile-ES estimators $\widehat{\text{XES}}^{\star}(\tau_n')$, $\widetilde{\text{XES}}^{\star}(\tau_n')$, $\widehat{\text{XES}}^{\dagger}(\tau_n')$ and $\widetilde{\text{XES}}^{\dagger}(\tau_n')$. The experiments all employed the same family of Student's t-distributions as before. The lessons were similar to those from the expectile-VaR setting, hence the results are not reported here. It may also be noticed that the Monte-Carlo estimates corresponding to $\widehat{\text{XES}}^{\star}(\tau_n')$ and $\widehat{\text{XES}}^{\dagger}(\tau_n')$ [respectively, $\widetilde{\text{XES}}^{\star}(\tau_n')$ and $\widetilde{\text{XES}}^{\dagger}(\tau_n')$] are very similar.

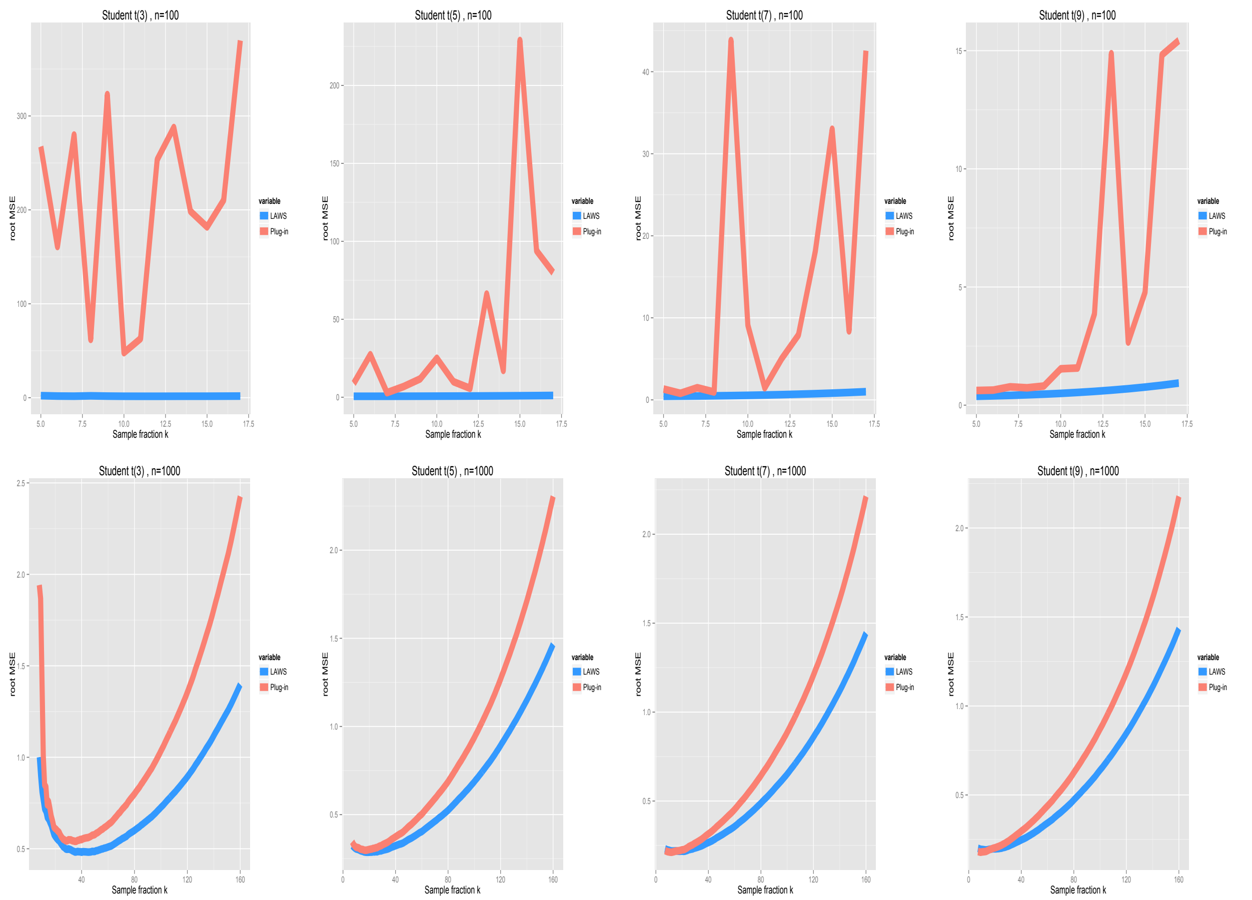


Figure 2: *Root MSE estimates for the t_3, t_5, t_7 and t_9-distributions, respectively, from left to right. The sample size $n = 100$ (top panels) and $n = 1000$ (bottom panels).*

$n = 100$

	RMSE		BIAS	
df	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\hat{\xi}^{\star}_{\tau'_n}$	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\hat{\xi}^{\star}_{\tau'_n}$
3	1.5010	47.9486	0.4888	1.7107
5	0.5963	2.9132	0.1253	0.4139
7	0.4385	0.8001	0.0797	0.2486
9	0.3753	0.6200	0.0579	0.1685

$n = 1000$

	RMSE		BIAS	
df	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\hat{\xi}^{\star}_{\tau'_n}$	$\widetilde{\xi}^{\star}_{\tau'_n}$	$\hat{\xi}^{\star}_{\tau'_n}$
3	0.4809	0.5403	0.2080	0.2599
5	0.2867	0.2981	0.0816	0.1088
7	0.2172	0.2119	0.0666	0.0629
9	0.1908	0.1781	0.0271	0.0440

Table 1: *Monte-Carlo results obtained for the optimal sample fraction k minimizing the MSE of each estimator.*

6 Application: SOA Group Medical Insurance data

The Society of Actuaries (SOA) Group Medical Insurance Large Claims Database records all the claim amounts exceeding 25,000 USD over the period 1991-92. As in Beirlant *et al.* (2004), we only deal here with the 75,789 claims for 1991. The histogram shown in Figure 4 (top) gives evidence of an important right-skewness. Accordingly, nothing guarantees that the future does not hold some unexpected higher claim amounts. Insurance companies

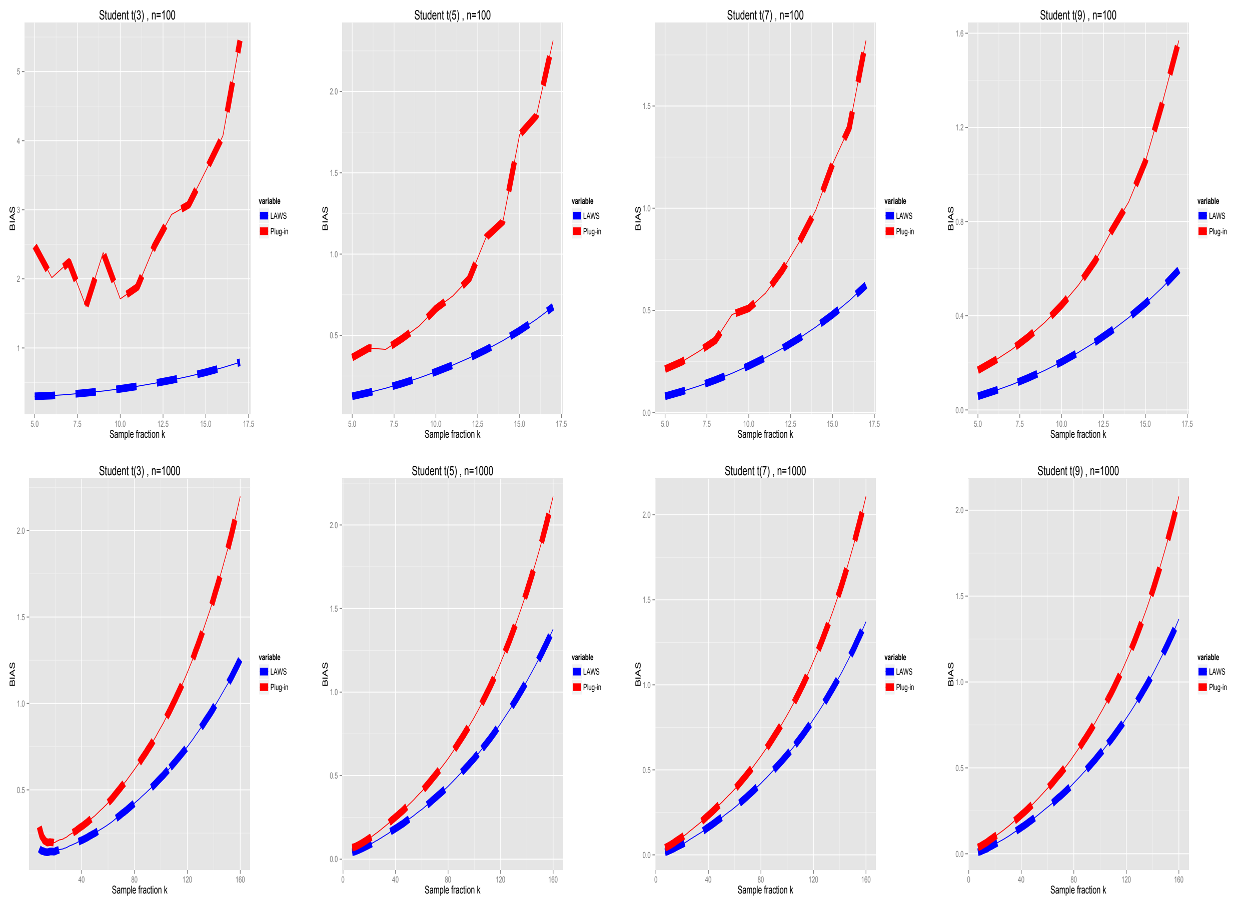


Figure 3: *Bias estimates for the* t_3*,* t_5*,* t_7 *and* t_9*-distributions, respectively, from left to right. The sample size* $n = 100$ *(top panels) and* $n = 1000$ *(bottom panels).*

are then interested in estimating the worst tail value of the corresponding loss severity distribution. One way of measuring this value at risk is by considering the Weissman quantile estimate $\hat{q}^{\star}_{1-p_n} = Y_{n-k,n}\left(\frac{k}{np_n}\right)^{\widehat{\gamma}_H}$ as described in (9), where $\widehat{\gamma}_H$ is the Hill estimator defined in (4), with $\tau'_n = 1 - p_n$ and $\tau_n = 1 - \frac{k}{n}$. Insurers typically are interested in $p_n = \frac{1}{100{,}000} < \frac{1}{n}$, that is, in an estimate of the claim amount that will be exceeded (on average) only once in 100,000 cases. Figure 4 (bottom) shows the quantile-VaR estimates $\hat{q}^{\star}_{1-p_n}$ against the sample fraction k (rainbow curve). A commonly used heuristic approach for selecting a pointwise estimate is to pick out a value of k corresponding to the first stable part of the plot [see, *e.g.*, Section 3 in de Haan and Ferreira (2006)]. Here, a stable region appears for k from 150 up to 500, leading to an estimate between 3.73 and 4.12 million. This result is consistent with the earlier analysis of Beirlant *et al.* (2004, p.125 and p.159), but it does not succeed in exceeding the sample maximum $Y_{n,n} = 4,518,420$ (indicated by the horizontal pink line).

Also, we would like to comment on the effect of $\widehat{\gamma}_H$ on $\hat{q}^{\star}_{1-p_n}$. This Hill estimate of the extreme-value index γ seems to mainly vary within the interval $[0.27, 0.43]$. Its effect on

$\hat{q}^{\star}_{1-p_n}$ is highlighted by a colour-scheme, ranging from dark red (low $\widehat{\gamma}_H$) to dark violet (high $\widehat{\gamma}_H$).

Given that $\widehat{\gamma}_H < \frac{1}{2}$, the proposed "indirect" estimate $\hat{\xi}^{\star}_{1-p_n}$ of the alternative expectile-based VaR, described in (10), is by construction more liberal than the quantile-VaR $\hat{q}^{\star}_{1-p_n}$. Its plot graphed in Figure 4 (bottom) in yellow indicates a more optimistic VaR between 3.02 and 3.40 million, for k ranging from 150 up to 500.

The "direct" asymmetric-least-squares based estimator $\widetilde{\xi}^{\star}_{1-p_n}$ of the expectile-VaR, defined in (11), is also displayed in the same figure in orange. It is more liberal than the quantile-VaR $\hat{q}^{\star}_{1-p_n}$ as well, but is more conservative than the indirect version $\hat{\xi}^{\star}_{1-p_n}$. It varies between 3.18 and 3.57 million over $k \in \{150, \ldots, 500\}$.

Another alternative option for measuring risk, which is more capable of extrapolating outside the range of the available observations, is by using the estimated quantile-ES

$$\widehat{\mathrm{QES}}^{\star}(1-p_n) = \frac{1}{k}\sum_{i=1}^{n} Y_i \mathbb{1}(Y_i > Y_{n-k,n}) \cdot \left(\frac{k}{np_n}\right)^{\widehat{\gamma}_H}$$

[see El Methni *et al.* (2014)]. Its graph shown in Figure 4 (bottom) in black line indicates a stable region for k ranging from 150 up to 500 with an averaged estimate of around 6.13 million, which is successfully extrapolated beyond the data but seems unrealistically high for the SOA.

To summarize, both estimates $\hat{\xi}^{\star}_{1-p_n}$ and $\widetilde{\xi}^{\star}_{1-p_n}$ of the expectile-VaR are too liberal, while the quantile-ES $\widehat{\mathrm{QES}}^{\star}(1-p_n)$ is too conservative. Although the quantile-VaR $\hat{q}^{\star}_{1-p_n}$ is less liberal, it remains too optimistic as it does not even succeed in exceeding the sample maximum. Our proposed plug-in estimates $\widehat{\mathrm{XES}}^{\star}(1-p_n)$, $\widetilde{\mathrm{XES}}^{\star}(1-p_n)$, $\widehat{\mathrm{XES}}^{\dagger}(1-p_n)$ and $\widetilde{\mathrm{XES}}^{\dagger}(1-p_n)$ of the alternative expectile-based expected shortfall, described in (12) and (13), steer an advantageous middle course between the optimism of the $\hat{\xi}^{\star}_{1-p_n}$, $\widetilde{\xi}^{\star}_{1-p_n}$ and $\hat{q}^{\star}_{1-p_n}$ values at risk and the excessive pessimism of the quantile-based expected shortfall $\widehat{\mathrm{QES}}^{\star}(1-p_n)$. The two estimates $\widehat{\mathrm{XES}}^{\star}(1-p_n)$ and $\widehat{\mathrm{XES}}^{\dagger}(1-p_n)$, based on the indirect expectile-VaR $\hat{\xi}^{\star}_{1-p_n}$ and graphed in Figure 4 (bottom) in gray and red lines, indicate a more realistic averaged risk estimate of around 5 million, for k from 150 up to 500, which might be good news to both insurers and pessimist regulators. The remaining two estimates $\widetilde{\mathrm{XES}}^{\star}(1-p_n)$ and $\widetilde{\mathrm{XES}}^{\dagger}(1-p_n)$, based on the direct expectile-VaR $\widetilde{\xi}^{\star}_{1-p_n}$ and shown in Figure 4 (bottom) in cyan and magenta lines, indicate a slightly higher averaged risk estimate of around 5.30 million, for $k \in \{150, \ldots, 500\}$.

A popular approach to the estimation of the optimal sample fraction k needed to apply the Hill extreme-value index and Weissman quantile estimators is by minimizing the asymptotic mean squared error of $\widehat{\gamma}_H$ and $\hat{q}^{\star}_{1-p_n}$. We refer to Beirlant *et al.* (2004, p.125) for a thorough discussion of the rationale for this adaptive selection of the tail sample fraction. They arrive in this way at the value $\hat{k} = 486$ which minimizes the estimated asymptotic mean squared error of $\hat{q}^{\star}_{1-p_n}$. The corresponding Weissman quantile-based VaR and ES estimates are

$$\hat{q}^{\star}_{1-p_n} = 3,807,575 \quad , \quad \widehat{\text{QES}}^{\star}(1-p_n) = 5,946,019.$$

For our comparison purposes, we find by using the same optimal sample fraction $\hat{k}$ that

$$\begin{aligned} \hat{\xi}^{\star}_{1-p_n} &= 3,092,991 \quad , \quad \widetilde{\xi}^{\star}_{1-p_n} = 3,294,602 \\ \widehat{\text{XES}}^{\star}(1-p_n) &= 4,827,261 \quad , \quad \widehat{\text{XES}}^{\dagger}(1-p_n) = 4,830,104 \\ \widetilde{\text{XES}}^{\star}(1-p_n) &= 5,141,918 \quad , \quad \widetilde{\text{XES}}^{\dagger}(1-p_n) = 5,144,946. \end{aligned}$$

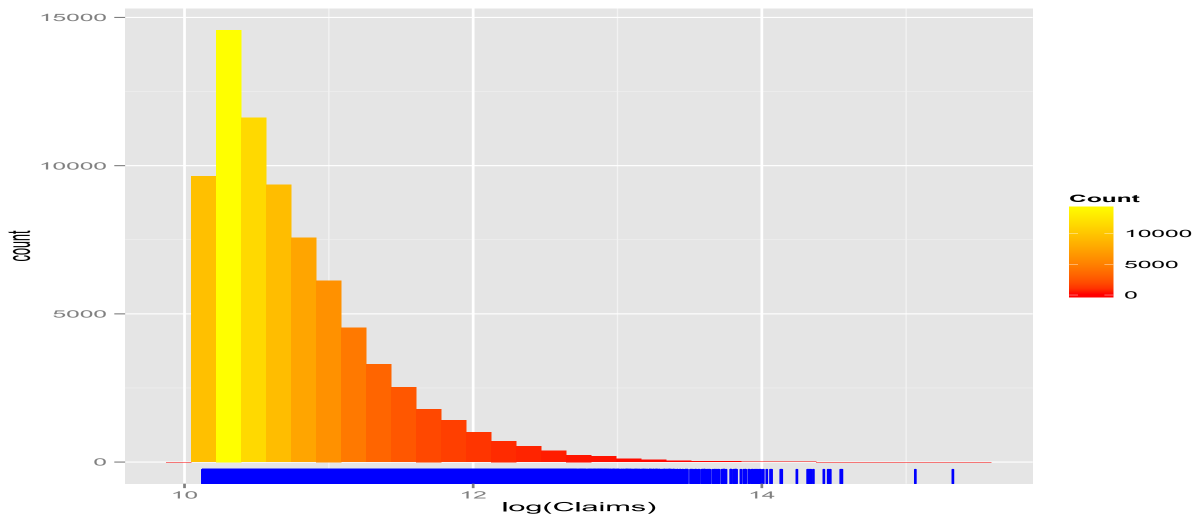
We also get the tail-index estimate $\widehat{\gamma}_H(\hat{k}) = 0.3593$.

If the political decision is to use the quantile-ES to determine the capital reserve, insurance companies would be motivated to merge in order to diminish the amount of required capital that is of the order of $5,946,019$ USD. This incentive to merge may create non-competitive effects and increase the risk. This may not occur, however, if the less pessimistic expectile-ES were favored, since it only requires the amount of $4,830,104$ USD or at most $5,144,946$ USD as a hedge against extreme risks. This exceeds the sample maximum $Y_{n,n} = 4,518,420$ USD, but not by much compared to the quantile-ES.

In contrast, if the political decision is to favor the use of a VaR in order to avoid changing severely the order of magnitude of the capital requirements, then the expectile-based VaR is the winner in terms of coherency, but also *a priori* psychologically in terms of its optimism or, say, realism in certain sectors of activity of the financial industry. Extreme expectile estimators are more liberal than their quantile analogues, since they are by construction less spread in the usual encountered practical settings where the tail-index estimate $\widehat{\gamma} < \frac{1}{2}$.

7 Conclusion

The search of efficient instruments of risk protection is of utmost importance in actuarial and portfolio allocation problems. Expectiles are used here to estimate the underlying concepts of Value-at-Risk and expected shortfall from the perspective of modern extreme-value theory. The first estimation method enables the usage of advanced high quantile and tail-index estimators. The second proposed method joins together the least asymmetrically weighted squares (LAWS) estimation with the tail restrictions of extreme-value theory. Simulation evidence suggests that the LAWS estimation is more efficient relative to the first method. The presented methodology is successfully applied in practice to the Society of Actuaries Group Medical Insurance data. Compared with the conventional quantile-based VaR and



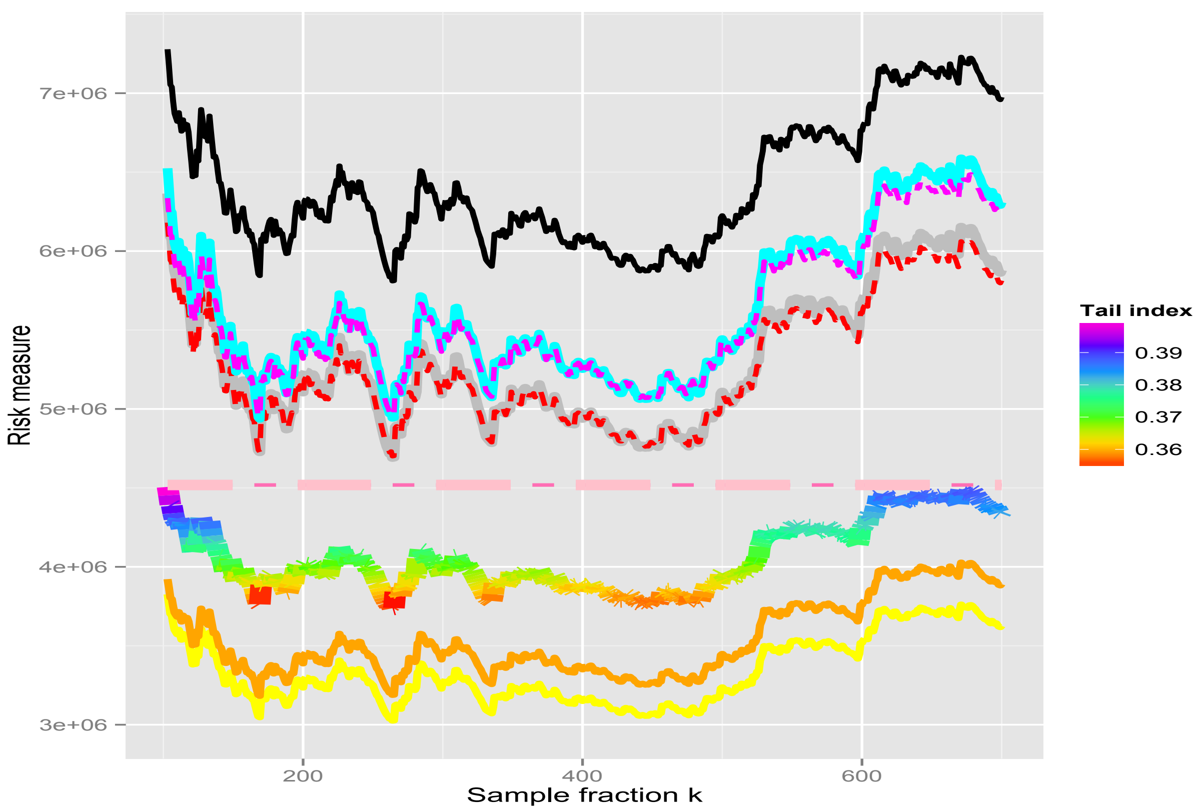


Figure 4: *SOA Group Medical Insurance data. (top) Histogram and scatterplot of the log-claim amounts; (bottom) The expectile-based VaR and ES plots* $\{(k, \hat{\xi}^{\star}_{1-p_n}(k))\}_k$ *in yellow,* $\{(k, \widetilde{\xi}^{\star}_{1-p_n}(k))\}_k$ *in orange,* $\{(k, \widehat{XES}^{\star}_k(1-p_n))\}_k$ *in gray,* $\{(k, \widehat{XES}^{\dagger}_k(1-p_n))\}_k$ *in red,* $\{(k, \widetilde{XES}^{\star}_k(1-p_n))\}_k$ *in cyan and* $\{(k, \widetilde{XES}^{\dagger}_k(1-p_n))\}_k$ *in magenta, along with the quantile-based VaR and ES plots* $\{(k, \hat{q}^{\star}_{1-p_n}(k))\}_k$ *as rainbow curve and* $\{(k, \widehat{QES}^{\star}_k(1-p_n))\}_k$ *in black. The sample maximum* $Y_{n,n}$ *is indicated by the horizontal pink line.*

expected shortfall, we favor the use of the expectile-based versions that afford more realistic risk estimates in terms of both liberalism and conservatism, respectively.

The proposed estimation methods all employ an estimator of the tail-index γ. Typically, the existing estimators of γ only rely on the use of quantiles. One way to extend our results is by developing a new estimator of γ completely based on expectiles themselves rather than quantiles. Another topic of interest for future research is to explore the estimation of extreme conditional expectiles for understanding the implications of regression data on their related risk measures. Also, we do not discuss the extensions of our theorems to a time dynamic setting such as, for instance, forecasting the expected shortfall in the conditional autoregressive expectile model introduced by Taylor (2008), but they are of genuine interest.

Appendix

For notational simplicity, let $\overline{F} = \overline{F}_Y$ be the survival function of Y. It is a consequence of Theorem 2.3.9 in de Haan and Ferreira (2006, p.48) that condition $\mathcal{C}_2(\gamma, \rho, A)$ entails the following second-order condition for the related survival function $\overline{F}$:

$$\forall x > 0,\ \lim_{t\to\infty} \frac{1}{A(1/\overline{F}(t))}\left[\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-1/\gamma}\right] = x^{-1/\gamma}\frac{x^{\rho/\gamma}-1}{\gamma\rho}. \tag{A.1}$$

Proof of Proposition 1. We start by noticing that equation (1) entails, for τ sufficiently large so that $\xi_\tau > 0$,

$$1 - \frac{\mathbb{E}(Y)}{\xi_\tau} = \frac{2\tau-1}{1-\tau}\mathbb{E}\left(\left[\frac{Y}{\xi_\tau} - 1\right] 1\!\!1\{Y/\xi_\tau \geq 1\}\right). \tag{A.2}$$

An integration by parts yields

$$\begin{aligned}\mathbb{E}\left(\left[\frac{Y}{\xi_\tau} - 1\right] 1\!\!1\{Y/\xi_\tau \geq 1\}\right) =& \int_1^\infty \overline{F}(\xi_\tau x)dx \\ =& \overline{F}(\xi_\tau)\left(\frac{\gamma}{1-\gamma} + A\left(\frac{1}{\overline{F}(\xi_\tau)}\right)\int_1^{+\infty}\frac{1}{A(1/\overline{F}(\xi_\tau))}\left[\frac{\overline{F}(\xi_\tau x)}{\overline{F}(\xi_\tau)} - x^{-1/\gamma}\right]dx\right).\end{aligned}$$

Recall that since Y has an infinite right endpoint, $\xi_\tau \to \infty$ as $\tau \uparrow 1$; using together equation (A.1), Theorem 2.3.9 in de Haan and Ferreira (2006) and a uniform inequality such as Theorem B.3.10 in de Haan and Ferreira (2006) applied to the function $\overline{F}$, we get after some easy computations

$$\mathbb{E}\left(\left[\frac{Y}{\xi_\tau} - 1\right] 1\!\!1\{Y/\xi_\tau \geq 1\}\right) = \overline{F}(\xi_\tau)\left(\frac{\gamma}{1-\gamma} + A\left(\frac{1}{\overline{F}(\xi_\tau)}\right)\frac{1}{(1-\gamma)(1-\rho-\gamma)}(1+\mathrm{o}(1))\right).$$

Plugging this equality into (A.2), we thus get

$$\frac{\overline{F}(\xi_\tau)}{1-\tau} = (\gamma^{-1}-1)\left(1 - \frac{\mathbb{E}(Y)}{\xi_\tau}\right)\frac{1}{2\tau-1}\left(1 + A\left(\frac{1}{\overline{F}(\xi_\tau)}\right)\frac{1}{\gamma(1-\rho-\gamma)}(1+\mathrm{o}(1))\right)^{-1}$$

and therefore

$$\frac{\overline{F}(\xi_\tau)}{1-\tau} = (\gamma^{-1}-1)\left(1 - \frac{\mathbb{E}(Y)}{\xi_\tau}(1+\mathrm{o}(1)) + 2(1-\tau)(1+\mathrm{o}(1)) - A\left(\frac{1}{\overline{F}(\xi_\tau)}\right)\frac{1}{\gamma(1-\rho-\gamma)}(1+\mathrm{o}(1))\right).$$

In particular, as noted in Bellini *et al.* (2014):

$$\frac{\overline{F}(\xi_\tau)}{1-\tau} \to (\gamma^{-1}-1) \text{ and thus } \xi_\tau = \left(\gamma^{-1}-1\right)^{-\gamma} q_\tau(1+\mathrm{o}(1)) \tag{A.3}$$

as $\tau \uparrow 1$. A consequence of this is that $(1-\tau)\xi_\tau = \mathrm{O}((1-\tau)q_\tau) \to 0$ as $\tau \uparrow 1$ and so

$$\frac{\overline{F}(\xi_\tau)}{1-\tau} = (\gamma^{-1}-1)\left(1 - \frac{(\gamma^{-1}-1)^\gamma \mathbb{E}(Y)}{q_\tau}(1+\mathrm{o}(1)) - \frac{(\gamma^{-1}-1)^{-\rho}}{\gamma(1-\rho-\gamma)}A((1-\tau)^{-1})(1+\mathrm{o}(1))\right)$$

where the regular variation property of $|A|$ was used. This completes the proof. ■

The key element in the proof of Corollary 1 is to apply Proposition 1 in conjunction with the following generic result.

Lemma 1. *Assume that v, V are such that $v(\tau) \uparrow \infty$ and $V(\tau) \downarrow 0$, as $\tau \uparrow 1$, and there exists $B > 0$ such that*

$$\frac{V(\tau)}{\overline{F}(v(\tau))} = B(1+e(\tau))$$

where $e(\tau) \to 0$ as $\tau \uparrow 1$. If condition $\mathcal{C}_2(\gamma, \rho, A)$ holds, with $\gamma > 0$ and F strictly increasing, then

$$\frac{v(\tau)}{U(1/V(\tau))} = B^\gamma\left(1 + \gamma e(\tau)(1+\mathrm{o}(1)) + \frac{B^\rho - 1}{\rho}A(1/V(\tau))(1+\mathrm{o}(1))\right) \quad \textit{as} \quad \tau \uparrow 1.$$

Proof of Lemma 1. Apply the function U to get

$$\frac{v(\tau)}{U(1/V(\tau))} - B^\gamma = \frac{U(B[1+e(\tau)]/V(\tau))}{U(1/V(\tau))} - B^\gamma.$$

By Theorem 2.3.9 in de Haan and Ferreira (2006), we may find a function A_0, equivalent to A at infinity, such that for any $\varepsilon > 0$, there is $t_0(\varepsilon) > 1$ such that for t, $tx \geq t_0(\varepsilon)$,

$$\left|\frac{1}{A_0(t)}\left(\frac{U(tx)}{U(t)} - x^\gamma\right) - x^\gamma\frac{x^\rho - 1}{\rho}\right| \leq \frac{\varepsilon}{2B^{\gamma+\varepsilon}}x^{\gamma+\rho}\max(x^\varepsilon, x^{-\varepsilon}).$$

Thus, for τ sufficiently close to 1, using this inequality with $t = 1/V(\tau)$ and $x = B[1+e(\tau)]$ gives

$$\left|\frac{1}{A_0(1/V(\tau))}\left(\frac{U(B[1+e(\tau)]/V(\tau))}{U(1/V(\tau))} - B^\gamma(1+e(\tau))^\gamma\right) - B^\gamma(1+e(\tau))^\gamma\frac{B^\rho(1+e(\tau))^\rho - 1}{\rho}\right| \leq \varepsilon$$

and therefore

$$\frac{1}{A_0(1/V(\tau))}\left(\frac{U(B[1+e(\tau)]/V(\tau))}{U(1/V(\tau))} - B^\gamma(1+e(\tau))^\gamma\right) \to B^\gamma\frac{B^\rho - 1}{\rho} \quad \text{as} \quad \tau \uparrow 1.$$

The desired result follows by a simple first-order Taylor expansion. ■

Proof of Corollary 1. We have in view of Proposition 1 that

$$\frac{1-\tau}{\overline{F}(\xi_\tau)} = (\gamma^{-1}-1)^{-1}(1+e(\tau))$$

with

$$e(\tau) = \frac{(\gamma^{-1}-1)^{\gamma}\mathbb{E}(Y)}{q_\tau}(1+\mathrm{o}(1)) + \frac{(\gamma^{-1}-1)^{-\rho}}{\gamma(1-\rho-\gamma)}A((1-\tau)^{-1})(1+\mathrm{o}(1)) \quad \text{as} \quad \tau \uparrow 1.$$

Using Lemma 1 and recalling that $U(1/(1-\tau)) = q_\tau$ gives the result. ■

Proof of Theorem 1. The consistency statement is an immediate consequence of the convergence

$$\frac{Y_{n-\lfloor n(1-\tau_n)\rfloor,n}}{q_{\tau_n}} = \frac{Y_{n-\lfloor n(1-\tau_n)\rfloor,n}}{U((1-\tau_n)^{-1})} = \frac{Y_{n-\lfloor n(1-\tau_n)\rfloor,n}}{U(n/\lfloor n(1-\tau_n)\rfloor)}(1+\mathrm{o}(1)) \xrightarrow{\mathbb{P}} 1$$

which follows from the regular variation of U and Corollary 2.2.2 in de Haan and Ferreira (2006, p.41). The asymptotic distribution is obtained by writing

$$\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 = \left(\frac{(\widehat{\gamma}^{-1}-1)^{-\widehat{\gamma}}}{(\gamma^{-1}-1)^{-\gamma}} - 1\right) + \left(\frac{\widehat{q}_{\tau_n}}{q_{\tau_n}} - 1\right)(1+\mathrm{o}_{\mathbb{P}}(1)) - r(\tau_n)(1+\mathrm{o}_{\mathbb{P}}(1)),$$

where $\sqrt{n(1-\tau_n)}r(\tau_n) \to \lambda$ in view of Corollary 1. Since

$$\forall x \in (0,1),\ \frac{d}{dx}\left((x^{-1}-1)^{-x}\right) = (x^{-1}-1)^{-x}\left\{(1-x)^{-1} - \log(x^{-1}-1)\right\},$$

the delta-method entails

$$\sqrt{n(1-\tau_n)}\left(\frac{(\widehat{\gamma}^{-1}-1)^{-\widehat{\gamma}}}{(\gamma^{-1}-1)^{-\gamma}} - 1\right) \xrightarrow{d} [(1-\gamma)^{-1} - \log(\gamma^{-1}-1)]\Gamma = m(\gamma)\Gamma, \tag{A.4}$$

from which the result easily follows. ■

Before moving to the proof of Theorem 2, we shall show a couple of useful preliminary results. The next two lemmas are entirely based on non-probabilistic arguments. In the first one, we use the fact that $\eta_\tau(y)/2$ is continuously differentiable with derivative

$$\varphi_\tau(y) := |\tau - \mathbb{1}\{y \le 0\}|y.$$

Lemma 2. *For all x, $y \in \mathbb{R}$ and $\tau \in (0,1)$,*

$$\frac{1}{2}(\eta_\tau(x-y) - \eta_\tau(x)) = -y\varphi_\tau(x) - \int_0^y (\varphi_\tau(x-t) - \varphi_\tau(x))dt.$$

Proof of Lemma 2. The result is a simple consequence of the equality

$$\frac{1}{2}(\eta_\tau(x-y) - \eta_\tau(x)) = \int_x^{x-y} \varphi_\tau(s)ds = -\int_0^y \varphi_\tau(x-t)dt$$

obtained by the change of variables $s = x - t$. ■

The next result gives a Lipschitz property for the derivative φ_τ.

Lemma 3. *For all x, $h \in \mathbb{R}$ and $\tau \in (0,1)$, we have*

$$\varphi_\tau(x-h) - \varphi_\tau(x) = -h|\tau - \mathbb{1}\{x \le 0\}| + (2\tau - 1)(x-h)(\mathbb{1}\{x \le h\} - \mathbb{1}\{x \le 0\}),$$

and in particular $|\varphi_\tau(x-h) - \varphi_\tau(x)| \le |h|(1 - \tau + 2\mathbb{1}\{x > \min(h,0)\})$.

Proof of Lemma 3. Write

$$\varphi_\tau(x-h) - \varphi_\tau(x) = -h|\tau - \mathbb{1}\{x \le 0\}| + (x-h)(|\tau - \mathbb{1}\{x \le h\}| - |\tau - \mathbb{1}\{x \le 0\}|).$$

Besides,

$$\begin{aligned} |\tau - \mathbb{1}\{x \le h\}| - |\tau - \mathbb{1}\{x \le 0\}| &= \tau(\mathbb{1}\{x \le h\} - \mathbb{1}\{x \le 0\}) + (1-\tau)(\mathbb{1}\{x > h\} - \mathbb{1}\{x > 0\}) \\ &= (2\tau - 1)(\mathbb{1}\{x \le h\} - \mathbb{1}\{x \le 0\}), \end{aligned}$$

from which the desired equality follows. The required bound on $|\varphi_\tau(x-h) - \varphi_\tau(x)|$ is then obtained by noting that

$$|\tau - \mathbb{1}\{x \le 0\}| = \tau \mathbb{1}\{x > 0\} + (1-\tau)\mathbb{1}\{x \le 0\} \le 1 - \tau + \mathbb{1}\{x > 0\} \tag{A.5}$$

and

$$|x-h||\mathbb{1}\{x \le h\} - \mathbb{1}\{x \le 0\}| \le |h||\mathbb{1}\{x \le h\} - \mathbb{1}\{x \le 0\}| \le |h|\mathbb{1}\{x > \min(h,0)\}. \tag{A.6}$$

Combining (A.5) and (A.6) completes the proof. ∎

The last result will be useful to derive the limit distribution of the objective function $\psi_n(u)$ described in (6).

Lemma 4. *Pick $a > 1$ and assume that $0 < \gamma < 1/a$. Then*

$$\mathbb{E}(|\varphi_\tau(Y - \xi_\tau)|^a) = a\xi_\tau^a(1-\tau)(\gamma^{-1} - 1)B(a, \gamma^{-1} - a)(1 + \mathrm{o}(1)) \quad \text{as} \quad \tau \uparrow 1,$$

where $B(s,t) = \int_0^1 u^{s-1}(1-u)^{t-1}du$ is the Beta function evaluated at (s,t).

Proof of Lemma 4. As a first step, write

$$\mathbb{E}(|\varphi_\tau(Y - \xi_\tau)|^a) = (1-\tau)^a\mathbb{E}([\xi_\tau - Y]^a \mathbb{1}\{Y \le \xi_\tau\}) + \tau^a\mathbb{E}([Y - \xi_\tau]^a \mathbb{1}\{Y > \xi_\tau\}). \tag{A.7}$$

Furthermore, for any x, y such that $x < y$, $(y-x)^a \le 2^{a-1}(|x|^a + |y|^a)$ by Hölder's inequality, so that

$$\mathbb{E}([\xi_\tau - Y]^a \mathbb{1}\{Y \le \xi_\tau\}) \le 2^{a-1}\mathbb{E}([|\xi_\tau|^a + |Y|^a]\mathbb{1}\{Y \le \xi_\tau\}).$$

The condition $\gamma < 1/a$ ensures that $\mathbb{E}|Y|^a < \infty$. Recall that $\xi_\tau \uparrow \infty$ as $\tau \uparrow 1$ and use the dominated convergence theorem to get

$$\mathbb{E}([\xi_\tau - Y]^a \mathbb{1}\{Y \le \xi_\tau\}) = \mathrm{O}(\xi_\tau^a) \quad \text{as} \quad \tau \uparrow 1. \tag{A.8}$$

Besides, an integration by parts and a change of variables entail

$$\mathbb{E}([Y-\xi_\tau]^a \mathbb{1}\{Y > \xi_\tau\}) = a\xi_\tau^{a-1} \int_{\xi_\tau}^{\infty} \left(\frac{x}{\xi_\tau} - 1\right)^{a-1} \overline{F}(x)dx = a\xi_\tau^a \overline{F}(\xi_\tau) \int_1^{\infty} (v-1)^{a-1} \frac{\overline{F}(\xi_\tau v)}{\overline{F}(\xi_\tau)} dv.$$

Using a uniform convergence theorem such as Proposition B.1.10 in de Haan and Ferreira (2006, p.360) gives

$$\mathbb{E}([Y - \xi_\tau]^a \mathbb{1}\{Y > \xi_\tau\}) = a\xi_\tau^a \overline{F}(\xi_\tau) \int_1^{\infty} (v - 1)^{a-1} v^{-1/\gamma} dv (1 + \mathrm{o}(1)) \quad \text{as} \quad \tau \uparrow 1.$$

Combining this equality with (A.3) yields

$$\mathbb{E}([Y-\xi_\tau]^a \mathbb{1}\{Y > \xi_\tau\}) = a\xi_\tau^a (1-\tau)(\gamma^{-1}-1) \int_1^{\infty} (v-1)^{a-1} v^{-1/\gamma} dv (1+\mathrm{o}(1)) \quad \text{as} \quad \tau \uparrow 1. \tag{A.9}$$

Combining (A.7), (A.8), (A.9) and using the change of variables $u = 1 - v^{-1}$ gives the desired result. ∎

Proof of Theorem 2. Use Lemma 2 to write, for any u,

$$\begin{aligned} \psi_n(u) &= -uT_{1,n} + T_{2,n}(u) \qquad (A.10) \\ \text{with } \ T_{1,n} &:= \frac{1}{\sqrt{n(1-\tau_n)}} \sum_{i=1}^n \frac{1}{\xi_{\tau_n}} \varphi_{\tau_n}(Y_i - \xi_{\tau_n}) =: \sum_{i=1}^n S_{n,i} \\ \text{and } \ T_{2,n}(u) &:= -\frac{1}{\xi_{\tau_n}^2} \sum_{i=1}^n \int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} (\varphi_{\tau_n}(Y_i - \xi_{\tau_n} - t) - \varphi_{\tau_n}(Y_i - \xi_{\tau_n}))dt. \end{aligned}$$

The random variables $S_{n,i}$ are independent, identically distributed, and centered since

$$\xi_{\tau_n} = \mathrm{argmin}_{u\in\mathbb{R}} \mathbb{E}(\eta_{\tau_n}(Y_i - u) - \eta_{\tau_n}(Y_i)) \Rightarrow \mathbb{E}(\varphi_{\tau_n}(Y_i - \xi_{\tau_n})) = 0$$

(where a differentiation under the expectation sign was used). We shall prove that

$$\frac{T_{1,n}}{\sqrt{\mathrm{Var}(T_{1,n})}} \xrightarrow{d} \mathcal{N}(0,1) \tag{A.11}$$

for which it is sufficient to show that for some $\delta > 0$,

$$\frac{n\mathbb{E}|S_{n,1}|^{2+\delta}}{[n\,\mathrm{Var}(S_{n,1})]^{1+\delta/2}} \to 0 \quad \text{as} \quad n \to \infty$$

and use Lyapunov's criterion. Choose $\delta > 0$ small enough so that $\gamma < 1/(2+\delta)$ and apply Lemma 4 to get

$$\frac{n\mathbb{E}|S_{n,1}|^{2+\delta}}{[n\operatorname{Var}(S_{n,1})]^{1+\delta/2}} = \mathrm{O}([n(1-\tau_n)]^{-\delta/2}) \to 0 \quad \text{as} \quad n \to \infty.$$

Convergence (A.11) follows and, especially, Lemma 4 entails

$$T_{1,n} \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma}{1-2\gamma}\right). \tag{A.12}$$

We now turn to the control of the second term $T_{2,n}(u)$. Write

$$T_{2,n}(u) = T_{3,n}(u) - \frac{n}{\xi_{\tau_n}^2}\int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} [\mathbb{E}(\varphi_{\tau_n}(Y-\xi_{\tau_n}-t)) - \mathbb{E}(\varphi_{\tau_n}(Y-\xi_{\tau_n}))]dt. \tag{A.13}$$

The random term $T_{3,n}(u)$ is a sum of independent, identically distributed and centered random variables, which we shall examine after having controlled first the nonrandom term on the right-hand side of (A.13). By Lemma 3, we obtain

$$\begin{aligned}\mathbb{E}(\varphi_{\tau_n}(Y-\xi_{\tau_n}-t)) - \mathbb{E}(\varphi_{\tau_n}(Y-\xi_{\tau_n})) = (2\tau_n-1)\mathbb{E}((Y-\xi_{\tau_n}-t)(1\!\!1\{Y \le \xi_{\tau_n}+t\} - 1\!\!1\{Y \le \xi_{\tau_n}\})) \\ - t\mathbb{E}(|\tau_n - 1\!\!1\{Y \le \xi_{\tau_n}\}|).\end{aligned} \tag{A.14}$$

Clearly

$$\mathbb{E}(|\tau_n - 1\!\!1\{Y \le \xi_{\tau_n}\}|) = \tau_n\overline{F}(\xi_{\tau_n}) + (1-\tau_n)F(\xi_{\tau_n}).$$

It therefore follows from (A.3) that

$$\mathbb{E}(|\tau_n - 1\!\!1\{Y \le \xi_{\tau_n}\}|) = \gamma^{-1}(1-\tau_n)(1+\mathrm{o}(1)) \tag{A.15}$$

as $n \to \infty$. Let further $\psi(t) := \mathbb{E}((Y-t)1\!\!1\{Y > t\})$ and observe that

$$\begin{aligned}\mathbb{E}((Y-\xi_{\tau_n}-t)(1\!\!1\{Y \le \xi_{\tau_n}+t\} - 1\!\!1\{Y \le \xi_{\tau_n}\})) &= \mathbb{E}((Y-\xi_{\tau_n}-t)(1\!\!1\{Y > \xi_{\tau_n}\} - 1\!\!1\{Y > \xi_{\tau_n}+t\})) \\ &= \psi(\xi_{\tau_n}) - \psi(\xi_{\tau_n}+t) - t\overline{F}(\xi_{\tau_n}).\end{aligned}$$

Integrating by parts entails

$$\psi(\xi_{\tau_n}) - \psi(\xi_{\tau_n}+t) = \int_{\xi_{\tau_n}}^{\xi_{\tau_n}+t} \overline{F}(x)dx = \xi_{\tau_n}\overline{F}(\xi_{\tau_n})\int_1^{1+t/\xi_{\tau_n}} \frac{\overline{F}(\xi_{\tau_n}v)}{\overline{F}(\xi_{\tau_n})}dv$$

from which we deduce that

$$\mathbb{E}((Y-\xi_{\tau_n}-t)(1\!\!1\{Y \le \xi_{\tau_n}+t\} - 1\!\!1\{Y \le \xi_{\tau_n}\})) = t\overline{F}(\xi_{\tau_n})\left(\frac{\xi_{\tau_n}}{t}\int_1^{1+t/\xi_{\tau_n}} \frac{\overline{F}(\xi_{\tau_n}v)}{\overline{F}(\xi_{\tau_n})}dv - 1\right).$$

We now bound the term into brackets as follows: let $I_n(u) = [0, |u|\xi_{\tau_n}/\sqrt{n(1-\tau_n)}]$ and write

$$\sup_{|t|\in I_n(u)} \left| \frac{\xi_{\tau_n}}{t} \int_1^{1+t/\xi_{\tau_n}} \frac{\overline{F}(\xi_{\tau_n} v)}{\overline{F}(\xi_{\tau_n})} dv - 1 \right| \leq \sup_{|t|\in I_n(u)} \frac{\xi_{\tau_n}}{|t|} \left| \int_1^{1+t/\xi_{\tau_n}} \left[\frac{\overline{F}(\xi_{\tau_n} v)}{\overline{F}(\xi_{\tau_n})} - v^{-1/\gamma} \right] dv \right| + \mathrm{o}(1) = \mathrm{o}(1)$$

by the uniform convergence theorem for regularly varying functions [see Theorem 1.5.2 in Bingham *et al.* (1987), p.22], the continuity of $v \mapsto v^{-1/\gamma}$ at 1 and the convergence $n(1-\tau_n) \to \infty$. As a consequence, by (A.3), the equality

$$\mathbb{E}((Y - \xi_{\tau_n} - t)(\mathbb{1}\{Y \leq \xi_{\tau_n} + t\} - \mathbb{1}\{Y \leq \xi_{\tau_n}\})) = t(1-\tau_n) r_n(t) \tag{A.16}$$

holds with $r_n(t) \to 0$ uniformly in t such that $|t| \in I_n(u)$. Combine (A.13), (A.14), (A.15) and (A.16) to get

$$\begin{aligned} T_{2,n}(u) &= \frac{u^2}{2\gamma}(1 + \mathrm{o}(1)) + T_{3,n}(u), \qquad \text{(A.17)} \\ \text{with } \ T_{3,n}(u) &:= -\frac{1}{\xi_{\tau_n}^2} \sum_{i=1}^n \int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} [\mathcal{S}_{n,i}(\xi_{\tau_n} + t) - \mathcal{S}_{n,i}(\xi_{\tau_n})] dt \end{aligned}$$

where the $\mathcal{S}_{n,i}(v) := \varphi_{\tau_n}(Y - v) - \mathbb{E}(\varphi_{\tau_n}(Y - v))$ are independent copies of $\mathcal{S}_n(v) := \varphi_{\tau_n}(Y - v) - \mathbb{E}(\varphi_{\tau_n}(Y - v))$. Thus

$$\mathrm{Var}(T_{3,n}(u)) = \frac{n}{\xi_{\tau_n}^4} \mathrm{Var}\left(\int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} [\mathcal{S}_n(\xi_{\tau_n} + t) - \mathcal{S}_n(\xi_{\tau_n})] dt \right).$$

We now notice that for any v, $\mathcal{S}_n(v)$ is centered and thus

$$\mathrm{Var}(T_{3,n}(u)) = \frac{n}{\xi_{\tau_n}^4} \int_{[0,\, u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}]^2} \mathbb{E}([\mathcal{S}_n(\xi_{\tau_n} + s) - \mathcal{S}_n(\xi_{\tau_n})][\mathcal{S}_n(\xi_{\tau_n} + t) - \mathcal{S}_n(\xi_{\tau_n})]) ds\, dt$$

(where the integrability properties of Y were used to switch integrals). By the Cauchy-Schwarz inequality,

$$\mathrm{Var}(T_{3,n}(u)) \leq \frac{n}{\xi_{\tau_n}^4} \left(\int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} \sqrt{\mathbb{E}(|\mathcal{S}_n(\xi_{\tau_n} + t) - \mathcal{S}_n(\xi_{\tau_n})|^2)}\, dt \right)^2. \tag{A.18}$$

Applying Lemma 3, we get for any t

$$|\mathcal{S}_n(\xi_{\tau_n} + t) - \mathcal{S}_n(\xi_{\tau_n})| \leq 2|t|[1 - \tau_n + \mathbb{1}\{Y > \xi_{\tau_n} + \min(t, 0)\} + \overline{F}(\xi_{\tau_n} + \min(t, 0))].$$

Using the inequality $|a + b + c|^2 \leq 3(a^2 + b^2 + c^2)$ yields

$$\mathbb{E}(|\mathcal{S}_n(\xi_{\tau_n} + t) - \mathcal{S}_n(\xi_{\tau_n})|^2) \leq 12t^2[(1-\tau_n)^2 + \overline{F}(\xi_{\tau_n} + \min(t, 0))(1 + \overline{F}(\xi_{\tau_n} + \min(t, 0)))]. \tag{A.19}$$

Finally, using again the regular variation property of $\overline{F}$ and the convergence $n(1-\tau_n) \to \infty$,

$$\sup_{|s| \in I_n(u)} |\overline{F}(\xi_{\tau_n} + s) - \overline{F}(\xi_{\tau_n})| = \overline{F}(\xi_{\tau_n}) \sup_{|s| \in I_n(u)} \left| \frac{\overline{F}(\xi_{\tau_n} + s)}{\overline{F}(\xi_{\tau_n})} - 1 \right| = \mathrm{o}(\overline{F}(\xi_{\tau_n})) = \mathrm{o}(1-\tau_n) \tag{A.20}$$

in view of (A.3). Using (A.3) once again and combining (A.18), (A.19) and (A.20) yields

$$\mathrm{Var}(T_{3,n}(u)) = \mathrm{O}\left(\frac{n}{\xi_{\tau_n}^4}(1-\tau_n) \left| \int_0^{u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}} |t| \, dt \right|^2 \right) = \mathrm{O}\left(\frac{1}{n(1-\tau_n)} \right) \to 0$$

as $n \to \infty$. Whence the convergence $T_{3,n}(u) \xrightarrow{\mathbb{P}} 0$; combining (A.10), (A.12) and (A.17) entails

$$\psi_n(u) \xrightarrow{d} -uZ\sqrt{\frac{2\gamma}{1-2\gamma}} + \frac{u^2}{2\gamma} \quad \text{as} \quad n \to \infty$$

(with Z being standard Gaussian) in the sense of finite-dimensional convergence. As a function of u, this limit is almost surely finite and defines a convex function which has a unique minimum at

$$u^* = \gamma\sqrt{\frac{2\gamma}{1-2\gamma}} Z \stackrel{d}{=} \mathcal{N}\left(0, \gamma^2 \frac{2\gamma}{1-2\gamma}\right).$$

Applying the convexity lemma of Geyer (1996) completes the proof. ■

Proof of Theorem 3. By the equality (8), we have

$$\log\left(\frac{\overline{\xi}^{\star}_{\tau_n'}}{\xi_{\tau_n'}}\right) = \log\left(\frac{\hat{q}^{\star}_{\tau_n'}}{q_{\tau_n'}}\right) + \log\left(\frac{\overline{\xi}_{\tau_n}}{\xi_{\tau_n}}\right) - \log\left(\frac{\widehat{q}_{\tau_n}}{q_{\tau_n}}\right) + \log\left(\frac{\xi_{\tau_n}}{q_{\tau_n}}\right) - \log\left(\frac{\xi_{\tau_n'}}{q_{\tau_n'}}\right).$$

Furthermore, the convergence $\log[(1-\tau_n)/(1-\tau_n')] \to \infty$ entails

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\hat{q}^{\star}_{\tau_n'}}{q_{\tau_n'}}\right) \xrightarrow{d} \Gamma, \tag{A.21}$$

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\overline{\xi}_{\tau_n}}{\xi_{\tau_n}}\right) = \mathrm{O}_{\mathbb{P}}\left(1/\log[(1-\tau_n)/(1-\tau_n')]\right) = \mathrm{o}_{\mathbb{P}}(1), \tag{A.22}$$

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\widehat{q}_{\tau_n}}{q_{\tau_n}}\right) = \mathrm{O}_{\mathbb{P}}\left(1/\log[(1-\tau_n)/(1-\tau_n')]\right) = \mathrm{o}_{\mathbb{P}}(1), \tag{A.23}$$

$$\begin{aligned}\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\xi_{\tau_n}}{q_{\tau_n}}\right) &= \mathrm{O}\left(\sqrt{n(1-\tau_n)}r(\tau_n)/\log[(1-\tau_n)/(1-\tau_n')]\right) \\ &= \mathrm{o}(1),\end{aligned} \tag{A.24}$$

and
$$\begin{aligned}\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\xi_{\tau_n'}}{q_{\tau_n'}}\right) &= \mathrm{O}\left(\sqrt{n(1-\tau_n)}r(\tau_n')/\log[(1-\tau_n)/(1-\tau_n')]\right) \\ &= \mathrm{o}\left(\sqrt{n(1-\tau_n)}r(\tau_n)/\log[(1-\tau_n)/(1-\tau_n')]\right) \\ &= \mathrm{o}(1).\end{aligned} \tag{A.25}$$

Here, Theorem 4.3.8 in de Haan and Ferreira (2006, p.138) was used to show (A.21), while (A.22) and (A.23) follow from Theorem 2 above and from Theorem 2.4.1 in de Haan and Ferreira (2006, p.50), respectively. Convergences (A.24) and (A.25) are consequences of Corollary 1 and, in what concerns (A.25), the regular variation of $s \mapsto q_{1-s^{-1}}$ and $|A|$. Combining these convergence results and using the delta-method gives the desired conclusion. ■

Proof of Proposition 2. On the one hand, we have

$$\mathrm{XES}(\tau) = \frac{\mathbb{E}\left[Y \mathbb{1}(Y > \xi_\tau)\right]}{\overline{F}(\xi_\tau)} = \frac{\mathbb{E}\left[(Y-\xi_\tau)_+\right]}{\overline{F}(\xi_\tau)} + \xi_\tau,$$

where $y_+ = \max(y, 0)$. On the other hand, it follows from the proof of Theorem 11 in Bellini *et al.* (2014) that

$$\frac{\mathbb{E}\left[(Y-\xi_\tau)_+\right]}{\overline{F}(\xi_\tau)} \sim \frac{\xi_\tau}{\gamma^{-1}-1} \quad \text{as} \quad \tau \to 1.$$

Therefore $\frac{\mathrm{XES}(\tau)}{\xi_\tau} \sim \frac{1}{1-\gamma}$ as $\tau \to 1$. Likewise, we have

$$\mathrm{QES}(\tau) = \frac{\mathbb{E}\left[Y \mathbb{1}(Y > q_\tau)\right]}{\overline{F}(q_\tau)} = \frac{\mathbb{E}\left[(Y-q_\tau)_+\right]}{\overline{F}(q_\tau)} + q_\tau,$$

with

$$\frac{\mathbb{E}\left[(Y-q_\tau)_+\right]}{\overline{F}(q_\tau)} \sim \frac{q_\tau}{\gamma^{-1}-1} \quad \text{as} \quad \tau \to 1.$$

Then $\frac{\mathrm{QES}(\tau)}{q_\tau} \sim \frac{1}{1-\gamma}$ as $\tau \to 1$. Whence $\frac{\mathrm{XES}(\tau)}{\mathrm{QES}(\tau)} \sim \frac{\xi_\tau}{q_\tau}$ as $\tau \to 1$. ■

Proof of Proposition 3. The starting point is equation (1), which is equivalent to

$$\frac{\mathrm{XES}(\tau)}{\xi_\tau} = 1 + \frac{1-\tau}{\overline{F}(\xi_\tau)}\frac{1}{2\tau-1}\left(1 - \frac{\mathbb{E}(Y)}{\xi_\tau}\right).$$

We have by Proposition 1 and (A.3), with the notation therein, that

$$\frac{1-\tau}{\overline{F}(\xi_\tau)} = \frac{\gamma}{1-\gamma}[1-\varepsilon(\tau)(1+\mathrm{o}(1))] \quad \text{and} \quad \frac{1}{\xi_\tau} = \frac{(\gamma^{-1}-1)^\gamma}{q_\tau}(1+\mathrm{o}(1)),$$

where the $\mathrm{o}(\cdot)$ terms have to be understood in the asymptotic sense as $\tau \uparrow 1$. Using a Taylor expansion thus yields:

$$\frac{\mathrm{XES}(\tau)}{\xi_\tau} = \frac{1}{1-\gamma} + \frac{\gamma}{1-\gamma}\left[2(1-\tau)(1+\mathrm{o}(1)) - \varepsilon(\tau)(1+\mathrm{o}(1)) - \frac{(\gamma^{-1}-1)^\gamma\mathbb{E}(Y)}{q_\tau}(1+\mathrm{o}(1))\right].$$

The condition $\gamma < 1$ entails $(1-\tau)q_\tau \to 0$ as $\tau \uparrow 1$, so that

$$\frac{\mathrm{XES}(\tau)}{\xi_\tau} = \frac{1}{1-\gamma} - \frac{\gamma}{1-\gamma}\left[\varepsilon(\tau)(1+\mathrm{o}(1)) + \frac{(\gamma^{-1}-1)^\gamma\mathbb{E}(Y)}{q_\tau}(1+\mathrm{o}(1))\right].$$

Using once again Proposition 1 gives

$$\varepsilon(\tau) + \frac{(\gamma^{-1}-1)^\gamma\mathbb{E}(Y)}{q_\tau} = -\frac{(\gamma^{-1}-1)^{-\rho}}{\gamma(1-\rho-\gamma)}A((1-\tau)^{-1})(1+\mathrm{o}(1)) + \mathrm{o}(q_\tau^{-1}),$$

whence

$$\frac{\mathrm{XES}(\tau)}{\xi_\tau} = \frac{1}{1-\gamma}\left[1 + \frac{(\gamma^{-1}-1)^{-\rho}}{1-\rho-\gamma}A((1-\tau)^{-1})(1+\mathrm{o}(1)) + \mathrm{o}(q_\tau^{-1})\right].$$

This completes the proof. ■

References

[1] Abdous, B. and Remillard, B. (1995). Relating quantiles and expectiles under weighted-symmetry, *Annals of the Institute of Statistical Mathematics*, **47**, 371–384.

[2] Baüerle, N. and Müller, A. (2006). Stochastic Orders and Risk Measures: Consistency and Bounds, *Insurance: Mathematics and Economics*, **38**, 132–148.

[3] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of extremes: Theory and applications*, Wiley.

[4] Bellini, F. (2012). Isotonicity results for generalized quantiles, *Statistics and Probability Letters*, **82**, 2017–2024.

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

[6] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

[7] Breckling, J. and Chambers, R. (1988). M-quantiles, *Biometrika*, **75**, 761–772.

[8] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics.

[9] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves, *Test*, **20** (2), 311–333.

[10] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, **19** (5B), 2557–2589.

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

[12] De Rossi, G. and Harvey, H. (2009). Quantiles, expectiles and splines, *Journal of Econometrics*, **152**, 179–185.

[13] Duffie, D. and Pan, J. (1997). An overview of value at risk, *Journal of Derivatives*, **4**, 7–49.

[14] Efron, B. (1991). Regression percentiles using asymmetric squared error loss, *Statistica Sinica*, **1**, 93–125.

[15] El Methni, J., Gardes, L. and Girard, S. (2014). Nonparametric estimation of extreme risks from conditional heavy-tailed distributions, *Scandinavian Journal of Statistics*, **41** (4), 988–1012.

[16] Embrechts, P. and Hofert, M. (2013). Statistics and Quantitative Risk Management for Banking and Insurance, ETH Zurich.

[17] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1987). *Modelling Extremal Events for Insurance and Finance*, Springer.

[18] Embrechts, P., Lambrigger, D. and Wüthrich, M. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples, *Extremes*, **12**.

[19] Ferreira, A., de Haan, L. and Peng, L. (2003). On optimizing the estimation of high quantiles of a probability distribution, *Statistics*, **37**, 401–434.

[20] Geyer, C.J. (1996). On the asymptotics of convex stochastic optimization, unpublished manuscript.

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

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

[23] Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity, technical report, University of Toronto.

[24] Koenker, R. (1993). When Are Expectiles Percentiles? *Econometric Theory*, **9**, 526–527.

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

[26] 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*, **2**, 261–270.

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

[28] Schulze Waltrup, L., Sobotka, F., Kneib, T. and Kauermann, G. (2015). Expectile and Quantile Regression - David and Goliath? *Statistical Modelling* (to appear).

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

[30] Tsay, R.S. (2002). Analysis of Financial Time Series. Hoboken, NJ: Wiley.

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

[32] Yao, Q. and Tong, H. (1996). Asymmetric least squares regression and estimation: A nonparametric approach, *Nonparametric statistics*, **6**, 273–292.

[33] Ziegel, J.F. (2015). Coherence and elicitability, *Mathematical Finance*, DOI: 10.1111/mafi.12080 (to appear).

[34] Zou, H. (2014). Generalizing Koenker's distribution, *Journal of Statistical Planning and Inference*, **148**, 123–127.