Estimation of Tail Risk based on Extreme Expectiles

Abdelaati Daouia, Stéphane Girard, Gilles Stupfler

Estimation of Tail Risk based on Extreme Expectiles

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

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

Abstract

We use tail expectiles to estimate Value at Risk (VaR), Expected Shortfall (ES) and Marginal Expected Shortfall (MES), three instruments of risk protection of utmost importance in actuarial science and statistical finance. The concept of expectiles is a least squares analogue of quantiles. Both expectiles and quantiles were embedded in the more general class of M-quantiles as the minimizers of an asymmetric convex loss function. It has been proved very recently that the only M-quantiles that are coherent risk measures are the expectiles. Moreover, expectiles define the only coherent risk measure that is also elicitable. The elicitability corresponds to the existence of a natural backtesting methodology. The estimation of expectiles did not, however, receive yet any attention from the perspective of extreme values. The first estimation method that we propose enables the usage of advanced high quantile and tail-index estimators. The second method joins together the least asymmetrically weighted squares estimation with the tail restrictions of extreme-value theory. 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. A main tool is to first estimate the intermediate large expectile-based VaR, ES and MES, and then extrapolate these estimates to the very far tails. We show through a detailed simulation study the good performance of the procedures, and also present concrete applications to medical insurance data and three large US investment banks.

Key words : Asymmetric squared loss; Coherency; Expected shortfall; Expectiles; Extrapolation; Extreme values; Heavy tails; Marginal expected shortfall; Value at Risk.

1 Introduction

The concept of expectiles 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 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

$$\xi_\tau = \mathrm{argmin}_{\theta\in\mathbb{R}} \mathbb{E}\left\{\eta_\tau(Y-\theta) - \eta_\tau(Y)\right\}, \tag{1}$$

where $\eta_\tau(y) = |\tau - \mathbb{1}(y \leqslant 0)|\, y^2$, with $\mathbb{1}(\cdot)$ being the indicator function. The first advantage of this asymmetric least squares approach relative to quantiles lies in the computational expedience of sample expectiles using only scoring or iteratively-reweighted least squares. The second advantage is that 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. This benefit in terms of increased efficiency comes at the price of increased sensitivity to the magnitude of extremes. 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 tail expectiles can serve as a more efficient instrument of risk protection than the traditional quantile-based risk measures, namely Value at Risk (VaR) and Expected Shortfall (ES).

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. Furthermore, sample expectiles provide a class of smooth curves as functions of the level τ, which is not the case for sample quantiles. Most importantly, inference on expectiles is much easier than inference on quantiles. Unlike quantiles, the estimation of the asymptotic variance of sample expectiles does not involve the tedious "smoothing" of the values of the density function at quantiles. In terms of interpretability, 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.*,

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

Thus, the τ-expectile shares an intuitive interpretation similar to the τ-quantile, replacing the number of observations by the distance. 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 expectiles and quantiles of the same distribution coincide under the hypothesis of weighted-symmetry. Yao and Tong (1996) showed that quantiles are identical to expectiles, but with different orders τ. Very recently, Zou (2014) has derived the class of 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 an 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. These statistical M-functionals have been extensively investigated especially from the point of view of the axiomatic theory of risk measures. In particular, Bellini (2012) has shown that expectiles with $\tau \geqslant \frac{1}{2}$ are the only M-quantiles that are isotonic with respect to the increasing convex order. 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 robust in the sense of lipschitzianity with respect to the Wasserstein metric. Perhaps most importantly, expectiles benefit from the property of elicitability that corresponds to the existence of a natural backtesting methodology. The relevance of this property in connection with backtesting has been discussed, for instance, by Embrechts and Hofert (2014) and Bellini and Di Bernardino (2015) while its relationship with coherency has been addressed in Ziegel (2014) among others. Actually, expectiles define the only coherent risk measure that is also elicitable. As such, it has been shown by Gneiting (2011) that ES, the most popular coherent risk measure, is not elicitable. It is generally accepted that elicitability is a desirable property for model selection, computational efficiency, forecasting and testing algorithms.

Theoretical and numerical results, obtained very recently by Bellini and Di Bernardino (2015), indicate that expectiles are perfectly reasonable alternatives to classical quantile-based VaR and ES. They also provide a transparent financial meaning of expectiles in terms of their acceptance sets as being the amount of money that should be added to a position in order to have a prespecified, sufficiently high gain-loss ratio. Ehm *et al.* (2015) have shown that expectiles are optimal decision thresholds in binary investment problems with fixed cost basis and differential taxation of profits versus losses. Expectiles are also becoming increasingly popular in the econometric literature as can be seen, for instance, from Kuan *et al.* (2009), De Rossi and Harvey (2009), Embrechts and Hofert (2014) and the references therein. The statistical problem of expectile estimation did not, however, receive yet any attention from the perspective of extreme values.

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 Sobotka and Kneib (2012) and the references therein]. 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. Bellini et al. (2014), Mao *et al.* (2015), Bellini and Di Bernardino (2015) and Mao and Yang (2015) have already initiated and studied the connection of such extreme population expectiles with their quantile analogues when Y belongs to the domain of attraction of a Generalized Extreme Value distribution. They do not enter, however, into the crucial statistical question of how to estimate in practice these unknown tail quantities from available historical data.

There are many important applications in econometrics, environment, finance and insurance, 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. In this article, we focus on high expectiles ξ_{τ_n} in the challenging maximum domain of attraction of Pareto-type distributions, where standard expectile estimates at the tails are often unstable due to data sparsity. It has been found in statistical finance and actuarial science that Pareto-type distributions describe quite well the tail structure of losses. The rival quantile-based VaR and ES are investigated extensively in theoretical statistics and used widely in applied work. Notice that in applications, extreme losses correspond to tail probabilities τ_n at an extremely high level that can be even larger than $(1 - 1/n)$. Therefore, estimating the corresponding quantile-based risk measures is a typical extreme value problem. We refer the reader to the books of Embrechts *et al.* (1997), Beirlant *et al.* (2004), de Haan and Ferreira (2006), and the recent and elegant devices of El Methni *et al.* (2014) and Cai *et al.* (2015).

Let us point out some conceptual results of this paper. 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 (LAWS) estimation. Our main results establish the asymptotic normality of the thus obtained estimators, which makes statistical inference for both expectile-based VaR and ES feasible. Also, we provide adapted extreme expectile-based tools for the estimation of the Marginal Expected Shortfall (MES), an important factor when measuring the systemic risk of financial institutions. Denoting by X and Y, respectively, the loss of the equity return of a financial firm and that of the entire market, the MES is equal to $\mathbb{E}(X|Y > t)$, where t is a high threshold reflecting a systemic crisis, *i.e.*, a substantial market decline. For an extreme expectile $t = \xi_{\tau_n}$ and for a wide nonparametric class of bivariate distributions of (X, Y), we construct two asymptotically normal estimators of the MES. A rival procedure by Cai *et al.* (2015) is based on extreme quantiles. To our knowledge, this is the first work to actually join together the expectile perspective with the tail restrictions of extreme-value theory. Simulation evidence suggests that the direct LAWS method is more efficient for estimating the expectile-based VaR and ES than the MES.

We organize this paper as follows. Section 2 discusses the basic properties of the expectile-VaR including its connection with the standard quantile-VaR for high levels of prudentiality.

Section 3 presents the two estimation methods of intermediate and extreme expectiles. Section 4 explores the notion of expectile-based ES and discusses interesting axiomatic and asymptotic developments. Section 5 considers the problem of estimating the MES when the related variable is extreme. The good performance of the presented procedures is shown in Section 6 and concrete applications to medical insurance data and three large US investment banks are provided in Section 7.

2 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 can be defined as the minimizer (1) of a piecewise-quadratic loss function or, equivalently, as

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

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, one of them being

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

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 $q_\tau := F_Y^{-1}(\tau) = \inf\{y \in \mathbb{R} : F_Y(y) \geqslant \tau\}$ do. A justification for their use to describe distributions and their tails, as well as to quantify the "riskiness" implied by the return distribution under consideration, may be based on the collection of properties given in Supplement A.

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), Mao *et al.* (2015) and Mao and Yang (2015) have described what happens for large population 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

$$F_Y(y) = 1 - \ell(y)\cdot y^{-1/\gamma} \tag{3}$$

where $\ell(\cdot)$ is a slowly-varying function at infinity, *i.e*, $\ell(\lambda y)/\ell(y) \to 1$ as $y \to \infty$ for all $\lambda > 0$. 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 (3) describes sufficiently well the tail structure of actuarial and financial data. Writing $\overline{F}_Y := 1 - F_Y$, Bellini *et al.* (2014) have shown in the case $\gamma < 1$ that

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

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 (4) between high expectiles and quantiles can actually be refined appreciably by considering the second-order version of the regular variation condition (3). 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 $\gamma > 0$, $\rho \leqslant 0$, 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 extremal value condition can be found in de Haan and Ferreira (2006) along with abundant examples of commonly used 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*

$$\begin{aligned} \frac{\overline{F}_Y(\xi_\tau)}{1-\tau} &= (\gamma^{-1} - 1)(1 + \varepsilon(\tau)) \\ \textit{with } \varepsilon(\tau) &= -\frac{(\gamma^{-1} - 1)^\gamma \mathbb{E}(Y)}{q_\tau}(1 + \mathrm{o}(1)) - \frac{(\gamma^{-1} - 1)^{-\rho}}{\gamma(1 - \rho - \gamma)} A((1-\tau)^{-1})(1 + \mathrm{o}(1)) \ \textit{as } \tau \uparrow 1. \end{aligned}$$

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*

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

Other refinements under similar second order regular variation conditions can also be found in Mao *et al.* (2015) and Mao and Yang (2015). 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 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 $Y_{1,n} \leqslant \cdots \leqslant Y_{n,n}$ 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 (3) and on the asymptotic equivalence (4). 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, (4) entails that

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

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

$$\widehat{\gamma}_H = \frac{1}{k}\sum_{i=1}^{k} \log \frac{Y_{n-i+1,n}}{Y_{n-k,n}}, \tag{6}$$

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*

$$\sqrt{n(1-\tau_n)}\left(\widehat{\gamma} - \gamma, \frac{\widehat{q}_{\tau_n}}{q_{\tau_n}} - 1\right) \xrightarrow{d} (\Gamma, \Theta). \tag{7}$$

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 (6) of γ with $k = n(1-\tau_n)$, sufficient regularity conditions for (7) 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. As an immediate consequence we get the following.

Corollary 2. *If F_Y verifies $\mathcal{C}_2(\gamma, \rho, A)$ with $0 < \gamma < 1$ and $\tau_n \to 1$ is such that $n(1-\tau_n) \to \infty$, $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to 0$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to 0$, then*

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

Yet, a drawback to the resulting 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 $\widehat{\xi}_{\tau_n}$ may inherit the vexing defects of both $\widehat{q}_{\tau_n}$ and $\widehat{\gamma}$. 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 (LAWS) 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

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

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

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) = \arg\min_{u\in\mathbb{R}} \psi_n(u) \tag{8}$$

$$\text{with} \qquad \psi_n(u) := \frac{1}{2\xi_{\tau_n}^2}\sum_{i=1}^{n}\left[\eta_{\tau_n}(Y_i - \xi_{\tau_n} - u\xi_{\tau_n}/\sqrt{n(1-\tau_n)}) - \eta_{\tau_n}(Y_i - \xi_{\tau_n})\right].$$

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 (8), it is enough to investigate the asymptotic properties of the sequence (ψ_n). Built on this idea, we get the asymptotic normality of the LAWS 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*

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

Interestingly, in contrast to Theorem 1 and Corollary 2, 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 $\widetilde{\xi}_{\tau_n}$ does not hinge by construction on any particular type of quantile or tail-index estimators. A comparison of the asymptotic variance $V(\gamma)$ of $\widetilde{\xi}_{\tau_n}$ with the asymptotic variance $v(\gamma)$ of $\widehat{\xi}_{\tau_n}$ is provided in Figure 1. It can be seen that $V(\gamma) < v(\gamma)$ almost overall the domain $(0, 1/2)$, and that both asymptotic variances are extremely stable for values of $\gamma < 0.3$. Also, while $v(\gamma)$ remains lower than the level 2, $V(\gamma)$ explodes in the neighborhood of 1/2.

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

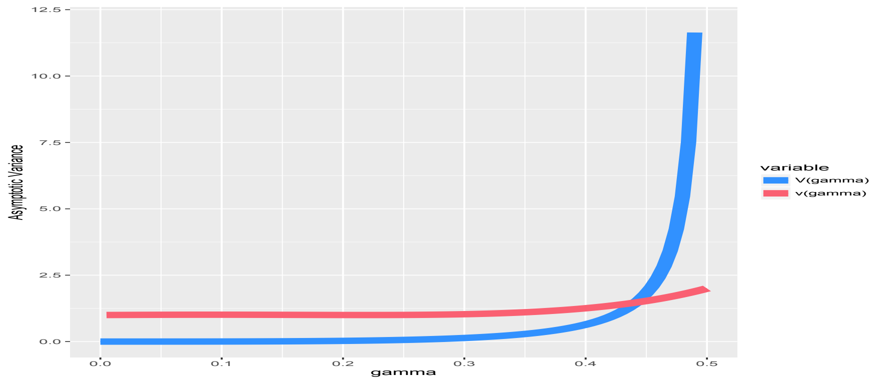

Figure 1: *Asymptotic variances $V(\gamma)$ of $\widetilde{\xi}_{\tau_n}$ in blue and $v(\gamma)$ of $\widehat{\xi}_{\tau_n}$ in red, with $\gamma \in (0, 1/2)$.*

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 (3) means that $U(tx)/U(t) \to x^\gamma$ as $t \to \infty$, which in turn suggests that

$$\frac{q_{\tau_n'}}{q_{\tau_n}} = \frac{U((1-\tau_n')^{-1})}{U((1-\tau_n)^{-1})} \approx \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma} \quad \text{and thus} \quad \frac{\xi_{\tau_n'}}{\xi_{\tau_n}} \approx \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}$$

by (5), for τ_n, τ_n' satisfying suitable conditions. This approximation motivates the following class of $\xi_{\tau_n'}$ plug-in estimators

$$\overline{\xi}^{\star}_{\tau_n'} \equiv \overline{\xi}^{\star}_{\tau_n'}(\tau_n) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}} \overline{\xi}_{\tau_n} \tag{9}$$

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 $\overline{\xi}^{\star}_{\tau_n'}/\overline{\xi}_{\tau_n} = \hat{q}^{\star}_{\tau_n'}/\widehat{q}_{\tau_n}$ 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

$$\hat{q}^{\star}_{\tau_n'} \equiv \hat{q}^{\star}_{\tau_n'}(\tau_n) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{q}_{\tau_n}. \tag{10}$$

Next we show that $(\frac{\overline{\xi}^{\star}_{\tau_n'}}{\xi_{\tau_n'}} - 1)$ has the same limit distribution as $(\widehat{\gamma} - \gamma)$ 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*

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

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, we can choose $\overline{\xi}_{\tau_n}$ in (9) to be either the indirect intermediate expectile estimator $\widehat{\xi}_{\tau_n}$, the resulting extreme expectile estimator $\hat{\xi}^{\star}_{\tau_n'} := \overline{\xi}^{\star}_{\tau_n'}$ being

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

or we may choose $\overline{\xi}_{\tau_n}$ to be the LAWS estimator $\widetilde{\xi}_{\tau_n}$, yielding the extreme expectile estimator

$$\widetilde{\xi}^{\star}_{\tau_n'} = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}}\widetilde{\xi}_{\tau_n}, \tag{12}$$

Their respective asymptotic properties are given in the next two corollaries of Theorem 3.

Corollary 3. *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*

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

Corollary 4. *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

The conventional quantile-based VaR was often criticized for being too optimistic since it only depends on the frequency of tail losses and not on their values. Acerbi (2002), Rockafellar and Uryasev (2002) proposed to change the measurement method for calculating losses from the usual quantile-VaR to an alternative coherent quantile-based method known as Expected Shortfall (ES). This proposal was criticized though for being too pessimistic because the ES only depends on the tail event. This motivated Kuan *et al.* (2009) to introduce the expectile-based VaR which depends on both the tail realizations of the loss variable and their probability. When estimating these three risk measures from a sample of historical data of size n, it is customary to choose "ordinary" tail probability levels τ substantially smaller than $(1-1/n)$. However, with the recent crisis in the financial industry, the vast majority of market participants (investors, risk managers, clearing houses), academics and 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, τ should be at an extremely high level that can be even larger than $(1-1/n)$. Unfortunately, with this required extreme perspective, both quantile-based VaR and ES tend, by construction, to break down and hence to change drastically the order of magnitude of the capital requirements. In contrast, since the realized values of the tail-index γ were found to be smaller than $\frac{1}{2}$ in most studies on actuarial and financial data, the expectile-VaR ξ_τ becomes in view of (5) definitely more liberal than the quantile-VaR q_τ, as the level $\tau \to 1$. Next, we introduce a new concept of expectile-based ES which steers asymptotically an advantageous middle course between the optimism of the expectile-VaR and the pessimism of the quantile-ES.

4.1 Basic properties

The standard ES, also known under the names Conditional Value at Risk or Average Value at Risk, is defined as the average of the quantile function above a given confidence level τ. It is traditionally expressed at the $100(1-\tau)\%$ security level as

$$\mathrm{QES}(\tau) := \frac{1}{1-\tau}\int_\tau^1 q_\alpha d\alpha.$$

When the financial position Y is continuous, $\mathrm{QES}(\tau)$ is just the conditional expectation of Y given that it exceeds the VaR q_τ. In this sense, it is referred to as Tail Conditional Expectation, with $-\mathrm{QES}(\tau)$ being interpreted as the expected return on the portfolio in the worst $100(1-\tau)\%$ of cases. Similarly, one may define an alternative expectile-based ES as

$$\mathrm{XES}(\tau) := \frac{1}{1-\tau}\int_\tau^1 \xi_\alpha d\alpha.$$

Both $\mathrm{XES}(\tau)$ and $\mathrm{QES}(\tau)$ obey the reasonable rule of assigning bigger weights to worse cases. Before moving to a deep study of the presented 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 a Monte Carlo experiment. Similar to 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 $P = 0.005$ 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 2, where $\tau = 0.99, 0.995 < 1 - \frac{1}{n}$ in top panels and $\tau = 0.999, 0.9995 \geqslant 1 - \frac{1}{n}$ in bottom panels.

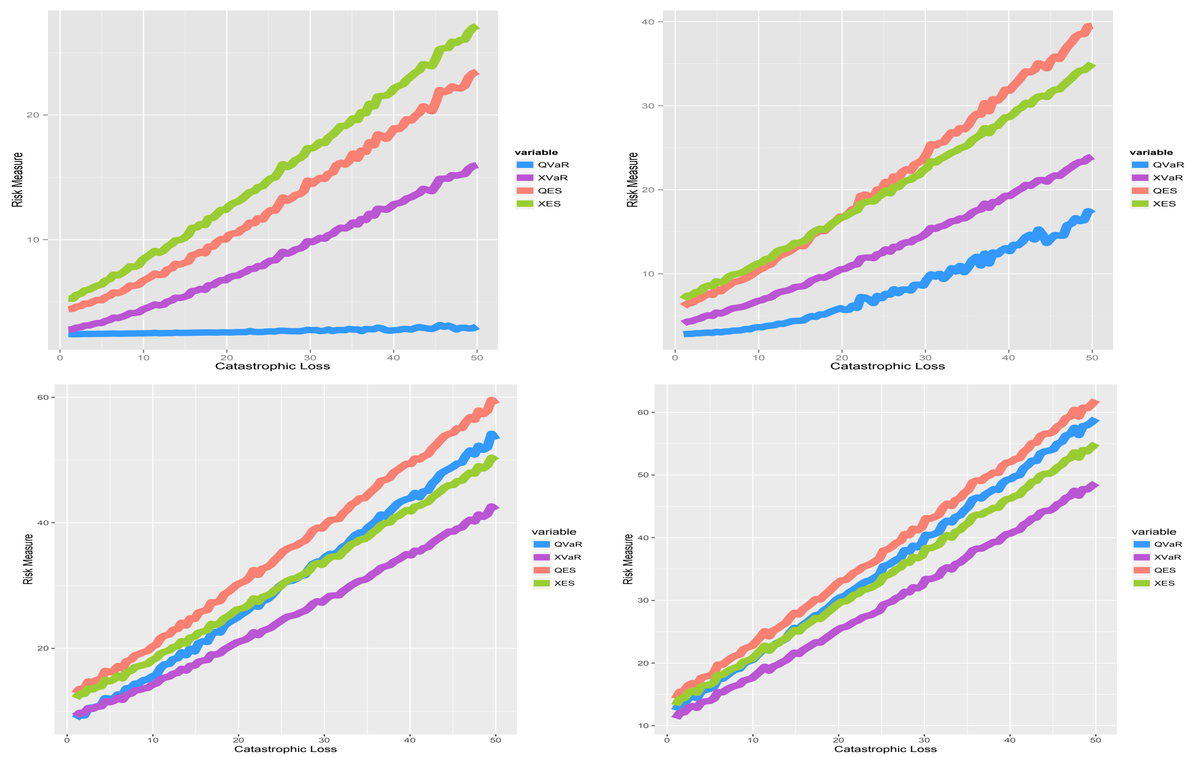

Figure 2: *The catastrophic loss sensitivity of empirical quantile-VaR (QVaR), expectile-VaR (XVaR), quantile-ES (QES) and expectile-ES (XES). From left to right and from top to bottom, we have* $\tau = 0.99,\ 0.995,\ 0.999,\ 0.9995$.

As expected, when $\tau < 1 - \frac{1}{n}$, the expectile-VaR (XVaR), the quantile-ES (QES) and the expectile-ES (XES) are affected by the extreme values from $\mathcal{N}(c, 1/\sqrt{P})$ for all c and in all scenarios, whereas the quantile-VaR (QVaR) may not respond properly to such catastrophic losses. It should be also clear that, in the case of QVaR, it is customary to choose in applications the level $\tau = 0.99$. Interestingly, in this case where the QVaR is not affected by any infrequent disaster, the XES is clearly more alert to all catastrophic losses as its magnitude is overall larger than that of all the other risk measures. When the QVaR becomes sensitive to the magnitude of extreme losses (*i.e.* $\tau = 0.995$), it remains too optimistic and even the conservative XES and QES still underestimate the infrequent catastrophic losses for $c \geqslant 10$. This advocates the use of extremely higher levels $\tau \geqslant 1 - \frac{1}{n}$. But in this case, it may be seen from the bottom panels that both quantile-based VaR and ES become excessively pessimistic, whereas their expectile-based analogues tend to be more realistic.

As a matter of fact, by considering the challenging maximum domain of attraction of Pareto-type distributions $F_Y(\cdot)$ with tail-index $\gamma < 1$, 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 $\mathrm{XES}(\tau) := (1-\tau)^{-1}\int_\tau^1 \xi_\alpha d\alpha$, is more conservative (respectively, liberal) than the quantile-ES $\mathrm{QES}(\tau) := (1-\tau)^{-1}\int_\tau^1 q_\alpha d\alpha$, for all τ large enough, when $\gamma > \frac{1}{2}$ (respectively, $\gamma < \frac{1}{2}$).

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* (3) *holds. Then*

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

These connections are very useful when it comes to proposing estimators for $\mathrm{XES}(\tau)$. 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, as $\tau \to 1$,*

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

From the point of view of the axiomatic theory, an influential paper in the literature by Artzner *et al.* (1999) provides an axiomatic foundation for coherent risk measures. Like the quantile-ES and the expectile-VaR, the expectile-ES satisfies all of their requirements (Translation invariance, Monotonicity, Subadditivity, and Positive homogeneity). However, in contrast to the quantile-ES, the coherence of the expectile-ES is actually a straightforward consequence of the coherence of the expectile-VaR above the median in conjunction with the fact that the expectile-ES is an increasing linear functional of the expectile-VaR above some high level.

Proposition 4. *For all $\tau \geqslant 1/2$, the expectile-based expected shortfall $XES(\tau)$ induces a coherent risk measure.*

This result does not seem to have been appreciated in the literature before. It affords an additional convincing reason that the use of both expectile-based VaR and ES may be preferred over the classical quantile-based versions.

4.2 Estimation and asymptotics

Typically, financial institutions and insurance companies are interested in the region $\tau = \tau_n' \uparrow 1$, as the sample size $n \to \infty$, which is particularly required to handle extreme events. 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:

$$\widehat{\mathrm{XES}}^{\star}(\tau_n') = (1-\widehat{\gamma})^{-1} \cdot \hat{\xi}^{\star}_{\tau_n'} \quad \text{and} \quad \widetilde{\mathrm{XES}}^{\star}(\tau_n') = (1-\widehat{\gamma})^{-1} \cdot \widetilde{\xi}^{\star}_{\tau_n'} \tag{13}$$

where $\hat{\xi}^{\star}_{\tau_n'}$ and $\widetilde{\xi}^{\star}_{\tau_n'}$ are the extreme expectile estimators defined above in (11)-(12), 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

$$\widehat{\mathrm{XES}}^{\dagger}(\tau_n') = \hat{\xi}^{\star}_{\tau_n'} \cdot \frac{\widehat{\mathrm{QES}}^{\star}(\tau_n')}{\hat{q}^{\star}_{\tau_n'}} \quad \text{or} \quad \widetilde{\mathrm{XES}}^{\dagger}(\tau_n') = \widetilde{\xi}^{\star}_{\tau_n'} \cdot \frac{\widehat{\mathrm{QES}}^{\star}(\tau_n')}{\hat{q}^{\star}_{\tau_n'}} \tag{14}$$

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 (10). 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 initial versions given in (13). Our first asymptotic result is for the extreme XES estimator $\widehat{\mathrm{XES}}^{\star}(\tau_n')$:

Corollary 5. *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

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

Regarding the LAWS-type estimator $\widetilde{\mathrm{XES}}^{\star}(\tau_n')$, we have the following result.

Corollary 6. *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 to the same distribution as the estimator $\widehat{\gamma}$ but with a slower rate in view of Corollaries 3 and 4. 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 Marginal expected shortfall

5.1 Setting and objective

With the recent financial crisis and the rising interconnection between financial institutions, interest in the concept of systemic risk has grown. Acharya *et al.* (2012), Brownlees and Engle (2012) and Engle *et al.* (2014) define systemic risk as the propensity of a financial institution to be undercapitalized when the financial system as a whole is undercapitalized. They have proposed econometric and statistical approaches to measure the systemic risk of financial institutions. An important step in constructing a systemic risk measure for a financial firm is to measure the contribution of the firm to a systemic crisis. A systemic event or crisis is specified as a major stock market decline that happens once or twice a decade. The total risk measured by the expected capital shortfall in the financial system during a systemic crisis is typically decomposed into firm level contributions. Each financial firm's contribution to systemic risk can then be measured as its marginal expected shortfall (MES), *i.e.*, the expected loss on its equity return conditional on the occurrence of an extreme loss in the aggregated return of the financial market. More specifically, denote the loss return on the equity of a financial firm as X and that of the entire market as Y. Then the MES at probability level $(1-\tau)$ is defined as

$$\mathrm{QMES}(\tau) = \mathbb{E}\{X|Y > q_{Y,\tau}\}, \quad \tau \in (0,1),$$

where $q_{Y,\tau}$ is the τth quantile of the distribution of Y. Typically, a systemic crisis defined as an extreme tail event corresponds to a probability τ at an extremely high level that can be even larger than $(1-1/n)$, where n is the sample size of historical data that are used for estimating $\mathrm{QMES}(\tau)$. The estimation procedure in Acharya *et al.* (2012) relies on daily data from only 1 year and assumes a specific linear relationship between X and Y. A nonparametric kernel estimation method has been performed in Brownlees and Engle (2012) and Engle *et al.* (2014), but cannot handle extreme events required for systemic risk measures (*i.e.* $1-\tau = O(1/n)$). Very recently, Cai *et al.* (2015) have proposed adapted extreme-value tools for the estimation of $\mathrm{QMES}(\tau)$ without recourse to any parametric structure on (X,Y). Here, instead of the extreme τth quantile $q_{Y,\tau}$, we will explore the use of the τth expectile analogue $\xi_{Y,\tau}$ in the MES at least for the following two respects: (i) The first advantage is that expectile estimation is more efficient as the weighted least squares rely on the distance to data points, while quantile estimation only knows whether an observation is below or above the predictor. It would be awkward to measure extreme risk based only on the frequency of tail losses and not on their values. Perhaps most importantly, (ii) since loss distributions typically belong to the maximum domain of attraction of Pareto-type distributions with tail-index $\gamma < 1/2$, the extreme quantile $q_{Y,\tau}$ is more spread (conservative) than the expectile $\xi_{Y,\tau}$ as the level $\tau \to 1$. Accordingly, the use of the more liberal τth expectile as an extremely high threshold in the marginal expected shortfall

$$\mathrm{XMES}(\tau) = \mathbb{E}\{X|Y > \xi_{Y,\tau}\}$$

would result in less excessive amounts of required capital reserve, which might be good news to financial institutions (we expect that high values of Y correspond to high values of X). A formal asymptotic connection between $\mathrm{XMES}(\tau)$ and $\mathrm{QMES}(\tau)$ is provided below in Proposition 5. It is the goal of the next section to establish estimators of the tail expectile-based MES and to unravel their asymptotic behavior. The asymptotic normality is derived for a large class of bivariate distributions of (X, Y), which makes statistical inference for $\mathrm{XMES}(\tau)$ feasible.

5.2 Tail dependence model

Suppose the random vector (X, Y) has a continuous bivariate distribution function $F_{(X,Y)}$ and denote by F_X and F_Y the marginal distribution functions of X and Y. Given that our goal is to estimate $\mathrm{XMES}(\tau)$ at an extreme level τ, we adopt the same conditions as Cai *et al.* (2015) on the right-hand tail of X and on the right-hand upper tail dependence of (X, Y). Here, the right-hand upper tail dependence between X and Y is described by the following joint convergence condition:

$\boldsymbol{\mathcal{JC}(R)}$ For all $(x, y) \in [0, \infty]^2$ such that at least x or y is finite, the limit

$$\lim_{t\to\infty} t\mathbb{P}(\overline{F}_X(X) \leqslant x/t, \overline{F}_Y(Y) \leqslant y/t) := R(x, y)$$

exists, with $\overline{F}_X = 1 - F_X$ and $\overline{F}_Y = 1 - F_Y$. The limit function R completely determines the so-called tail dependence function ℓ [Drees and Huang (1998)] via the identity $\ell(x, y) = x + y - R(x, y)$ for all $x, y \geqslant 0$ [see also Beirlant *et al.* (2004), Section 8.2]. Regarding the marginal distributions, we assume that X and Y are heavy-tailed with respective tail indices $\gamma_X, \gamma_Y > 0$, or equivalently, for all $z > 0$,

$$\frac{U_X(tz)}{U_X(t)} \to z^{\gamma_X} \quad \text{and} \quad \frac{U_Y(tz)}{U_Y(t)} \to z^{\gamma_Y} \quad \text{as} \quad t \to \infty,$$

with U_X and U_Y being, respectively, the left-continuous inverse functions of $1/\overline{F}_X$ and $1/\overline{F}_Y$. Compared with the quantile-based MES framework in Cai *et al.* (2015), we need the extra condition of heavy-tailedness of Y which is quite natural in the financial setting. Under these regularity conditions, we get the following asymptotic approximations for XMES(τ).

Proposition 5. *Suppose that condition $\mathcal{JC}(R)$ holds and that X and Y are heavy-tailed with respective indices $\gamma_X, \gamma_Y \in (0,1)$. Then*

$$\lim_{\tau \uparrow 1} \frac{XMES(\tau)}{U_X(1/\overline{F}_Y(\xi_{Y,\tau}))} = \int_0^\infty R(x^{-1/\gamma_X}, 1)dx, \tag{15}$$

$$\lim_{\tau \uparrow 1} \frac{XMES(\tau)}{QMES(\tau)} = \left(\gamma_Y^{-1} - 1\right)^{-\gamma_X}. \tag{16}$$

The first convergence result indicates that XMES(τ) is asymptotically equivalent to the small exceedance probability $U_X(1/\overline{F}_Y(\xi_{Y,\tau}))$ up to a multiplicative constant. Since as usual in the financial setting $0 < \gamma_X, \gamma_Y < 1/2$, the second result shows that XMES(τ) is more liberal than QMES(τ) as $\tau \to 1$. This is visualised in Figure 3 in the case of a standard bivariate Student t_ν-distribution on $(0,\infty)^2$ with density

$$f_\nu(x,y) = \frac{2}{\pi}\left(1 + \frac{x^2+y^2}{\nu}\right)^{-(\nu+2)/2}, \quad x, y > 0, \tag{17}$$

where $\nu = 3, 5, 7, 9$, respectively from left to right. It can be seen that QMES(τ) becomes overall much more conservative than XMES(τ) as τ approaches 1.

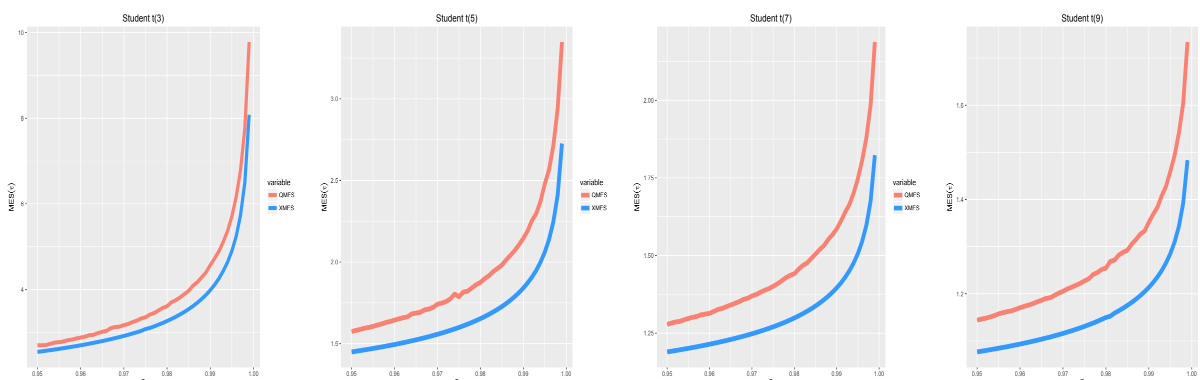

Figure 3: *QMES(τ) in red and XMES(τ) in blue, with $\tau \in [0.95, 1)$.*

5.3 Estimation and results

The asymptotic equivalences in Proposition 5 are of particular interest when it comes to proposing estimators for tail expectile-based MES. Two approaches will be distinguished. We consider first asymmetric least squares estimation by making use of the asymptotic equivalence (15). Subsequently we shall deal with a nonparametric estimator derived from the asymptotic connection (16) with the tail quantile-based MES.

5.3.1 Asymmetric least squares estimation

On the basis of the limit (15) and then of the heavy-tailedness assumption on X, we have for $\tau < \tau' < 1$ that, as $\tau \to 1$,

$$\text{XMES}(\tau') \approx \frac{U_X(1/\overline{F}_Y(\xi_{Y,\tau'}))}{U_X(1/\overline{F}_Y(\xi_{Y,\tau}))}\text{XMES}(\tau) \approx \left(\frac{\overline{F}_Y(\xi_{Y,\tau})}{\overline{F}_Y(\xi_{Y,\tau'})}\right)^{\gamma_X} \text{XMES}(\tau).$$

It follows then from Proposition 1 that

$$\text{XMES}(\tau') \approx \left(\frac{1-\tau'}{1-\tau}\right)^{-\gamma_X} \text{XMES}(\tau). \tag{18}$$

Hence, to estimate XMES(τ') at an arbitrary extreme level $\tau' = \tau_n'$, we first consider the estimation of XMES(τ) at an intermediate level $\tau = \tau_n$, and then we use the extrapolation technique of Weissman (1978). For estimating $\text{XMES}(\tau_n) = \mathbb{E}\{X|Y > \xi_{Y,\tau_n}\}$ at an intermediate level $\tau_n \to 1$ such that $n(1-\tau_n) \to \infty$, as $n \to \infty$, we use the empirical version

$$\widetilde{\text{XMES}}(\tau_n) := \frac{\sum_{i=1}^n X_i \mathbb{1}\{X_i > 0, Y_i > \widetilde{\xi}_{Y,\tau_n}\}}{\sum_{i=1}^n \mathbb{1}\{Y_i > \widetilde{\xi}_{Y,\tau_n}\}},$$

where $\widetilde{\xi}_{Y,\tau_n}$ is the LAWS estimator of ξ_{Y,τ_n}. As a matter of fact, in actuarial settings, we typically have a positive loss variable X, and hence $\mathbb{1}\{X_i > 0\} = 1$. When considering a real-valued profit-loss variable X, the MES is mainly determined by high, and hence positive, values of X as shown in Cai *et al.* (2015).

We shall show under general conditions that the estimator $\widetilde{\text{XMES}}(\tau_n)$ is $\sqrt{n(1-\tau_n)}$-relatively consistent. By plugging this estimator into approximation (18) together with a $\sqrt{n(1-\tau_n)}$-consistent estimator $\widehat{\gamma}_X$ of γ_X, we obtain the following estimator of XMES(τ_n'):

$$\widetilde{\text{XMES}}^*(\tau_n') \equiv \widetilde{\text{XMES}}^*(\tau_n'; \tau_n) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_X} \widetilde{\text{XMES}}(\tau_n).$$

To determine the limit distribution of this estimator, we need to quantify the rate of convergence in condition $\mathcal{JC}(R)$ as follows:

$\boldsymbol{\mathcal{JC}_2(R, \beta, \kappa)}$ Condition $\mathcal{JC}(R)$ holds and there exist $\beta > \gamma_X$ and $\kappa < 0$ such that

$$\sup_{\substack{x\in(0,\infty)\\ y\in[1/2,2]}} \left|\frac{t\mathbb{P}(\overline{F}_X(X) \leqslant x/t, \overline{F}_Y(Y) \leqslant y/t) - R(x,y)}{\min(x^\beta, 1)}\right| = \mathrm{O}(t^\kappa) \quad \text{as} \quad t \to \infty.$$

This is exactly condition (a) in Cai *et al.* (2015) under which an extrapolated estimator of QMES(τ_n') converges to a normal distribution. See also condition (7.2.8) in de Haan and Ferreira (2006). We also need to assume that the tail quantile function U_X (resp. U_Y) satisfies the second-order condition $\mathcal{C}_2(\gamma_X, \rho_X, A_X)$ (resp. $\mathcal{C}_2(\gamma_Y, \rho_Y, A_Y)$). The following generic theorem gives the asymptotic distribution of XMES(τ_n'). The asymptotic normality follows by using for example the Hill estimator $\widehat{\gamma}_X$ of the tail-index γ_X.

Theorem 4. *Suppose that condition $\mathcal{JC}_2(R, \beta, \kappa)$ holds, and U_X and U_Y satisfy conditions $\mathcal{C}_2(\gamma_X, \rho_X, A_X)$ and $\mathcal{C}_2(\gamma_Y, \rho_Y, A_Y)$ with $\gamma_X, \gamma_Y \in (0, 1/2)$ and $\rho_X < 0$. Assume further that*

- τ_n, $\tau_n' \uparrow 1$, *with* $n(1-\tau_n) \to \infty$ *and* $n(1-\tau_n') \to c < \infty$ *as* $n \to \infty$;
- $1 - \tau_n = \mathrm{O}(n^{\alpha-1})$ *for some* $\alpha < \min\left(\dfrac{-2\kappa}{-2\kappa+1}, \dfrac{2\gamma_X\rho_X}{2\gamma_X\rho_X + \rho_X - 1}\right)$;
- *The bias conditions* $\sqrt{n(1-\tau_n)}q_{Y,\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$, $\sqrt{n(1-\tau_n)}A_X((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$ *and* $\sqrt{n(1-\tau_n)}A_Y((1-\tau_n)^{-1}) \to \lambda_3 \in \mathbb{R}$ *hold;*
- $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_X - \gamma_X) \xrightarrow{d} \Gamma$.

Then, if $X > 0$ almost surely, we have that

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

This convergence remains still valid if $X \in \mathbb{R}$ provided

- $\mathbb{E}|X_-|^{1/\gamma_X} < \infty$; (19)
- $n(1-\tau_n) = o\left((1-\tau_n')^{-2\kappa(1-\gamma_X)}\right) \quad as \quad n \to \infty$, (20)

where $X_- = X - X_+$ with $X_+ = X \vee 0$.

5.3.2 Estimation based on tail QMES

On the basis of the limit (16), we consider the alternative estimator

$$\widehat{\mathrm{XMES}}^*(\tau_n') := \left(\widehat{\gamma}_Y^{-1} - 1\right)^{-\widehat{\gamma}_X} \widehat{\mathrm{QMES}}^*(\tau_n'),$$

where $\widehat{\gamma}_X$, $\widehat{\gamma}_Y$ and $\widehat{\mathrm{QMES}}^*(\tau_n')$ are suitable estimators of γ_X, γ_Y and $\mathrm{QMES}(\tau_n')$, respectively. Here, we use the Weissman-type device

$$\widehat{\mathrm{QMES}}^*(\tau_n') = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_X} \widehat{\mathrm{QMES}}(\tau_n) \tag{21}$$

of Cai *et al.* (2015) to estimate $\mathrm{QMES}(\tau_n')$, where

$$\widehat{\mathrm{QMES}}(\tau_n) = \frac{1}{\lfloor n(1-\tau_n)\rfloor}\sum_{i=1}^{n} X_i \mathbb{1}\{X_i > 0, Y_i > \widehat{q}_{Y,\tau_n}\},$$

with $\widehat{q}_{Y,\tau_n} := Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$ being an intermediate quantile-VaR. As a matter of fact, Cai *et al.* (2015) have suggested the use of two intermediate sequences in $\widehat{\gamma}_X$ and $\widehat{\mathrm{QMES}}(\tau_n)$ to be chosen in two steps in practice. To ease the presentation, we restrict to the same intermediate sequence τ_n in both $\widehat{\gamma}_X$ and $\widehat{\mathrm{QMES}}(\tau_n)$. Next, we derive the asymptotic distribution of the new estimator $\widehat{\mathrm{XMES}}^*(\tau_n')$.

Theorem 5. *Suppose that condition $\mathcal{JC}_2(R, \beta, \kappa)$ holds, and U_X and U_Y satisfy conditions $\mathcal{C}_2(\gamma_X, \rho_X, A_X)$ and $\mathcal{C}_2(\gamma_Y, \rho_Y, A_Y)$ with $\gamma_X \in (0, 1/2)$ and $\rho_X < 0$. Assume further that*

- τ_n, $\tau_n' \uparrow 1$, *with* $n(1-\tau_n) \to \infty$ *and* $n(1-\tau_n') \to c < \infty$ *as* $n \to \infty$;
- $1 - \tau_n = \mathrm{O}(n^{\alpha-1})$ *for some* $\alpha < \min\left(\dfrac{-2\kappa}{-2\kappa+1}, \dfrac{2\gamma_X\rho_X}{2\gamma_X\rho_X + \rho_X - 1}\right)$;
- *The bias conditions* $\sqrt{n(1-\tau_n)}q_{Y,\tau_n}^{-1} \to \lambda \in \mathbb{R}$ *and* $\sqrt{n(1-\tau_n)}A_X((1-\tau_n)^{-1}) \to 0$ *hold;*
- $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_X - \gamma_X) \xrightarrow{d} \Gamma$ *and* $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_Y - \gamma_Y) = \mathrm{O}_{\mathbb{P}}(1)$.

Then, if $X > 0$ almost surely, we have that

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

This convergence remains still valid if $X \in \mathbb{R}$ provided that (19) *and* (20) *hold.*

6 Simulation study

The aim of this section is to highlight some of the theoretical findings with numerical simulations. We will briefly touch on the presented tail XVaR and XES estimators in section 6.1 and tail XMES estimators in section 6.2.

6.1 Expectile-based VaR and ES

This section provides Monte-Carlo evidence that the direct estimation method is more efficient relative to the indirect method in terms of Mean-Squared Error (MSE), and is also the winner when it comes to making and evaluating point forecasts. Recall that the direct type estimator $\widetilde{\xi}^{\star}_{\tau_n'} \equiv \widetilde{\xi}^{\star}_{\tau_n'}(\tau_n)$, described in (12) and utilized in (13), is obtained via LAWS estimation. The indirect type estimator $\hat{\xi}^{\star}_{\tau_n'} \equiv \hat{\xi}^{\star}_{\tau_n'}(\tau_n)$, introduced in (11) and used in (13), results from a plug-in procedure based on an asymptotic equivalence with intermediate quantiles.

To evaluate finite-sample performance of these estimators, we have considered simulated samples from various Student's t-scenarios: t_3, t_5, t_7 and t_9. We used in all our simulations the Hill estimator of γ, the extreme level $\tau_n' = 0.995$ for $n = 100$ and $\tau_n' = 0.9994$ for $n = 1000$, and the intermediate levels $\tau_n = 1 - \frac{k}{n}$, where the integer k can actually be viewed as the effective sample size for tail extrapolation. We only present the results for $n = 1000$ here, a full comparison including additional results for optimal k is given in Supplement B.1.

Figure 4 gives the root-MSE and bias estimates computed over $10,000$ replications for samples of size 1000. 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. Our tentative conclusion is that the accuracy of the direct estimator $\widetilde{\xi}^{\star}_{\tau'_n}$ is quite respectable, especially for heavier tails (*i.e.* $df = 3, 5$). When $n = 100$, as can be clearly seen from Supplement B.1, $\widetilde{\xi}^{\star}_{\tau'_n}$ performs better than $\hat{\xi}^{\star}_{\tau'_n}$ in terms of both MSE and bias, whatever the thickness of the tails. 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.

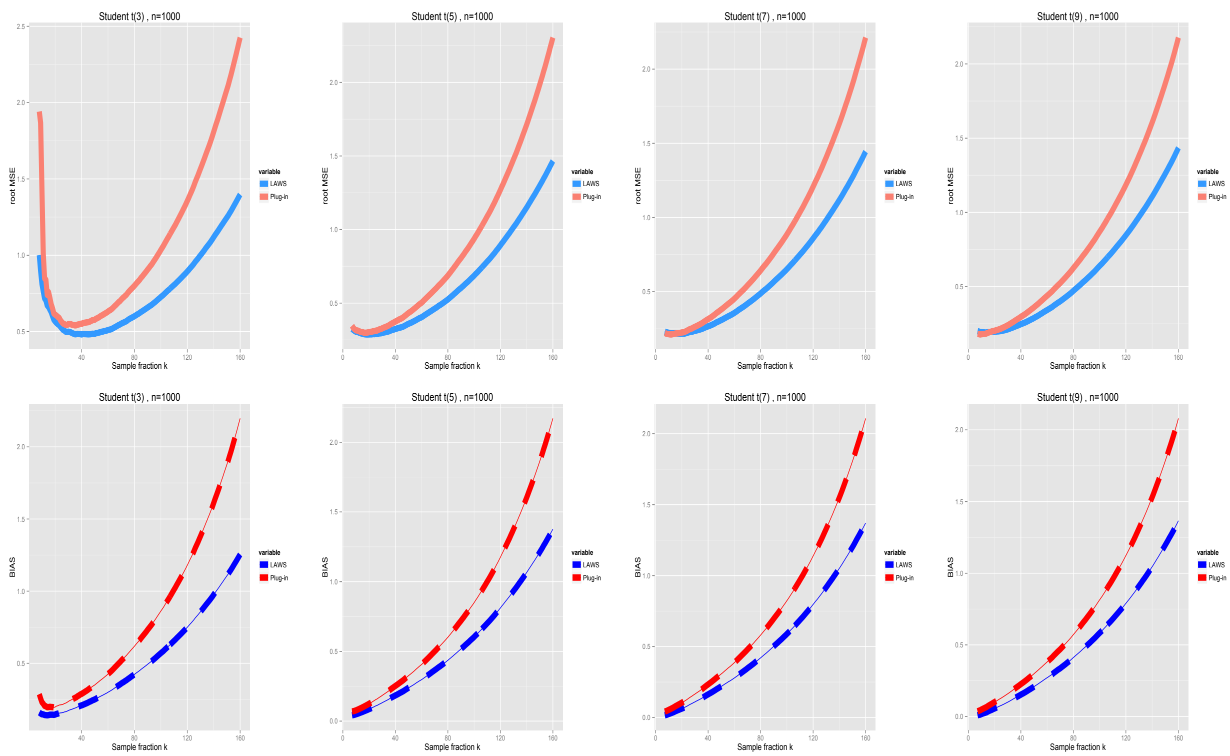

Figure 4: *Root MSE estimates (top) and Bias estimates (bottom) of* $\widetilde{\xi}^{\star}_{\tau'_n}(k)/\xi_{\tau'_n}$ *(blue) and* $\hat{\xi}^{\star}_{\tau'_n}(k)/\xi_{\tau'_n}$ *(red), for the* t_3, t_5, t_7, t_9*-distributions, respectively, from left to right.*

Another way of validating the presented estimation procedures for $\xi_{\tau'_n}$ on historical data is by using the elicitability property of expectiles as pointed out in Section 1. Following the ideas of Gneiting (2011), the competing estimates $\hat{\xi}^{\star}_{\tau'_n}$ and $\widetilde{\xi}^{\star}_{\tau'_n}$ can be compared from a forecasting perspective by means of their realized losses. A more comprehensive description of this comparison including Monte Carlo verification and validation is given in Supplement B.2, where the resulting average values of the realized losses seem to favor $\widetilde{\xi}^{\star}_{\tau'_n}$ over $\hat{\xi}^{\star}_{\tau'_n}$ for making and evaluating point forecasts. We also investigate the normality of the estimators $\hat{\xi}^{\star}_{\tau'_n}$ and $\widetilde{\xi}^{\star}_{\tau'_n}$ in Supplement B.3, where the Q–Q-plots indicate that the limit Theorem 3 and its Corollaries 3 and 4 provide adequate approximations for finite sample sizes.

Other simulation experiments have been undertaken to assess the finite-sample performance of the 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 $\widehat{\mathrm{XES}}^{\star}(\tau_n')$ and $\widehat{\mathrm{XES}}^{\dagger}(\tau_n')$ [respectively, $\widetilde{\mathrm{XES}}^{\star}(\tau_n')$ and $\widetilde{\mathrm{XES}}^{\dagger}(\tau_n')$] are very similar.

6.2 Expectile-based MES

When it comes to estimate $\mathrm{XMES}(\tau_n')$, in contrast to the expectile-based VaR and ES, this section provides Monte-Carlo evidence that the indirect estimation method is superior to the direct one in terms of both MSE and bias. Also, the indirect method seems to provide better adequacy in terms of asymptotic normality approximations for finite sample sizes. Recall that the direct estimator $\widetilde{\mathrm{XMES}}^{*}(\tau_n')$ is obtained via LAWS estimation, while the indirect estimator $\widehat{\mathrm{XMES}}^{*}(\tau_n')$ is built on the quantile-based MES estimator of Cai *et al.* (2015).

To investigate the finite sample performance of the two estimators, the simulation experiments employ the Student t_ν-distribution on $(0,\infty)^2$ with density $f_\nu(x,y)$ described in (17). It can be shown that this distribution satisfies the conditions $\mathcal{JC}_2(R,\beta,\kappa)$ and $\mathcal{C}_2(\gamma_X,\rho_X,A_X)$ of Theorems 4 and 5 (see Cai *et al.* (2015) for the case $\nu = 3$). Other motivating examples of distributions that satisfy these conditions can also be found in section 3 of Cai *et al.* (2015). All the experiments have $\nu \in \{3,5,7,9\}$ and $n = 1000$. For the choice of the intermediate and extreme expectile levels τ_n and τ_n', we used the same considerations as in Section 6.1.

A comparison of the two estimators is shown in Figure 5, where we present the root-MSE (top panels) and bias estimates (bottom panels) computed over $10,000$ simulated samples. Each picture displays the evolution of the obtained Monte-Carlo results, for the two normalized estimators $\widetilde{\mathrm{XMES}}^{*}/\mathrm{XMES}(\tau_n')$ and $\widehat{\mathrm{XMES}}^{*}/\mathrm{XMES}(\tau_n')$, as functions of the sample fraction k. Surprisingly, we observe that the latter estimator is clearly the winner in all cases in terms of both root-MSE and bias. The LAWS method is thus more suitable for estimating the expectile-based VaR and ES than the MES. As can be seen also in Supplement B.3, the limit Theorems 4 and 5 provide adequate approximations for finite sample sizes, with a slight advantage for the estimator $\widehat{\mathrm{XMES}}^{*}(\tau_n')$.

7 Applications

In this section, we apply our estimation methods to first estimate the tail VaR and ES for the Society of Actuaries (SOA) Group Medical Insurance Large Claims, and then to estimate the tail MES for three large investment banks in the USA.

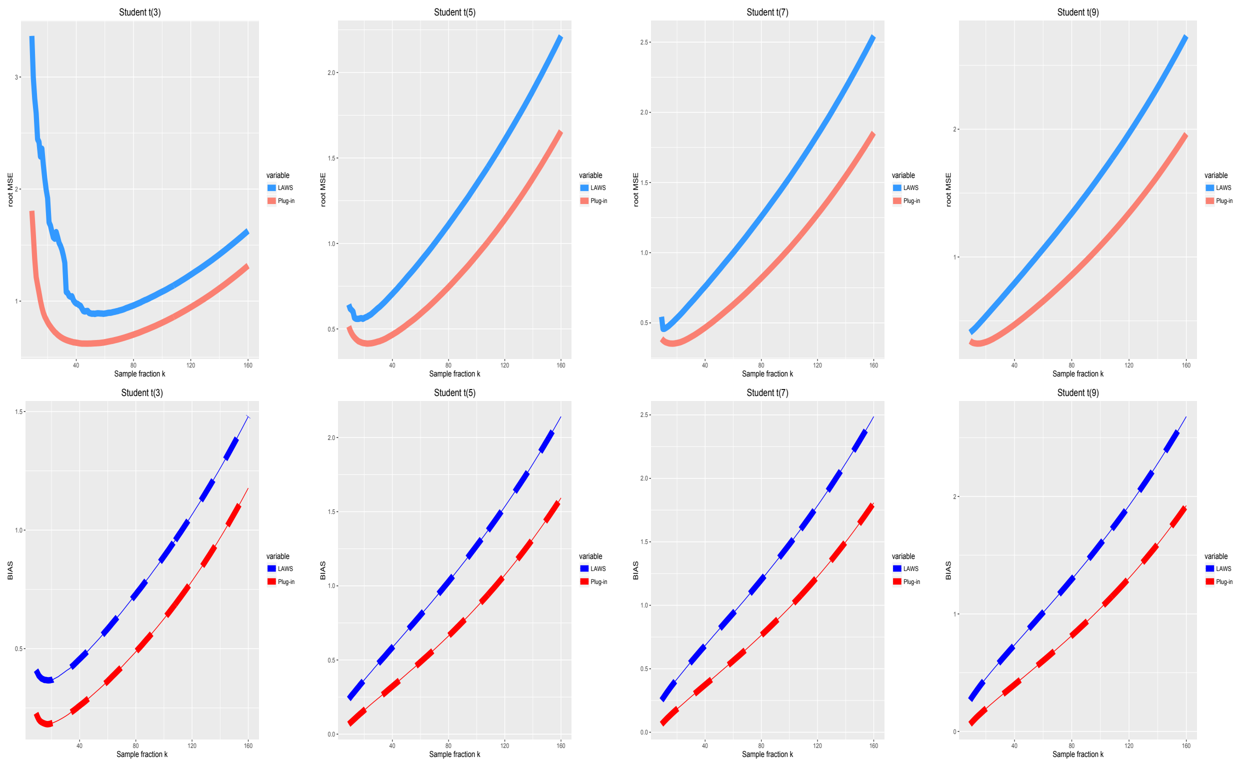

Figure 5: *Root MSE estimates (top) and Bias estimates (bottom) of* $\widetilde{XMES}^*/XMES$ *(blue) and* $\widehat{XMES}^*/XMES$ *(red), for the* t_3*,* t_5*,* t_7*,* t_9*-distributions, respectively, from left to right.*

7.1 VaR and ES for medical insurance data

The 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 6 (top) gives evidence of an important right-skewness. Accordingly, nothing guarantees that the future does not hold some unexpected higher claim amounts. Insurance companies 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 (10), where $\widehat{\gamma}_H$ is the Hill estimator defined in (6), 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 6 (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 estimate does not succeed in exceeding the sample maximum $Y_{n,n} = 4{,}518{,}420$ (indicated by the horizontal pink line), which is consistent with the earlier analysis of Beirlant *et al.* (2004, p.125 and p.159). The effect of $\widehat{\gamma}_H$ 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 $\widehat{\gamma}_H$). This Hill estimate of the extreme-value index γ seems to mainly vary within the interval $[0.27, 0.43]$.

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 (11), is by construction more liberal than the quantile-VaR $\hat{q}^\star_{1-p_n}$. Its plot graphed in Figure 6 (bottom) in yellow indicates a more optimistic VaR between 3.02 and 3.40 million, for $k \in [150, 500]$. The "direct" asymmetric-least-squares based estimator $\widetilde{\xi}^\star_{1-p_n}$ of the expectile-VaR, defined in (12), 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 appears to be more conservative than the indirect version $\hat{\xi}^\star_{1-p_n}$. It varies between 3.18 and 3.57 million over $k \in [150, 500]$.

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 6 (bottom) in black line indicates a stable region for $k \in [150, 500]$ 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 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 (13) and (14), 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 6 in gray and red lines, indicate a more realistic averaged risk estimate of around 5 million for $k \in [150, 500]$, which might be a good compromise 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 6 in cyan and magenta lines, indicate a slightly higher averaged risk estimate of around 5.30 million, for $k \in [150, 500]$.

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 6.13 million 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 5 million USD or at most

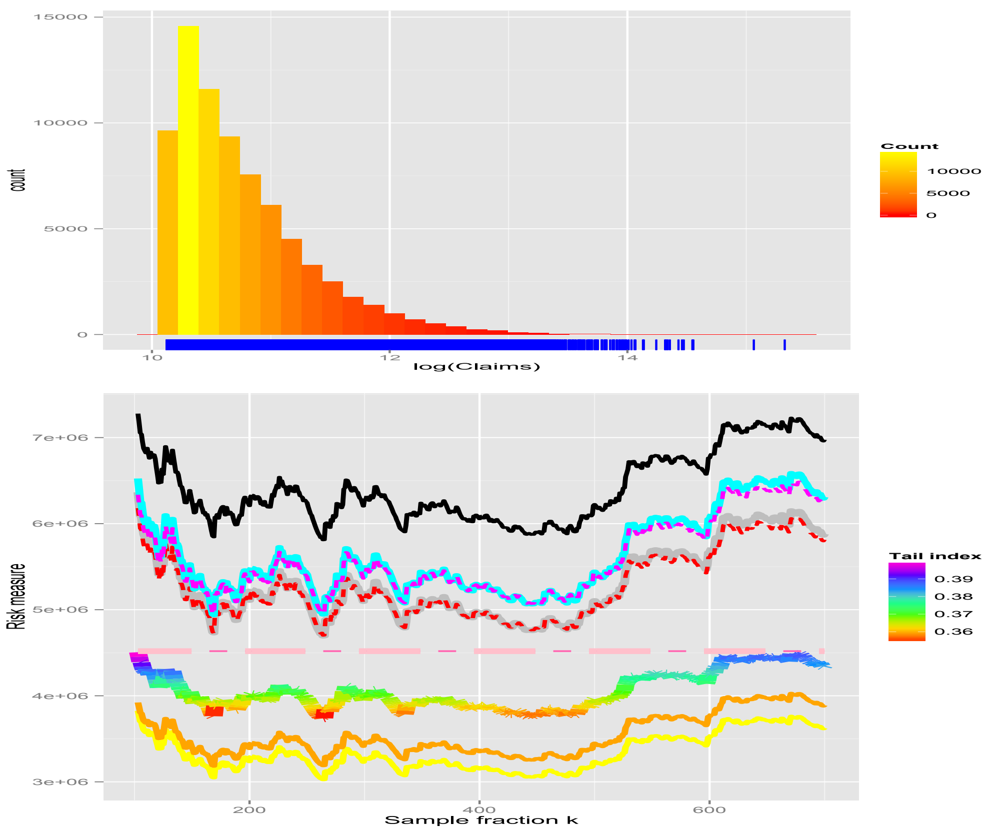

Figure 6: *SOA Group Medical Insurance data. (top) Histogram and scatterplot of the log-claim amounts; (bottom) The expectile-based VaR and ES plots* $\{(k, \hat{\xi}^{\star}_{1-p_n}(k))\}_k$ *in yellow,* $\{(k, \widetilde{\xi}^{\star}_{1-p_n}(k))\}_k$ *in orange,* $\{(k, \widehat{XES}^{\star}_k(1-p_n))\}_k$ *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.*

5.30 million 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 might be favored as it 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 indeed more liberal than their quantile analogues, since they are by construction less spread in the usual encountered practical settings where $\widehat{\gamma} < \frac{1}{2}$.

7.2 MES of three large US financial institutions

We consider the same investment banks as in the studies of Brownlees and Engle (2012) and Cai *et al.* (2015), namely Goldman Sachs, Morgan Stanley and T. Rowe Price. For the three banks, the dataset consists of the loss returns (X_i) on their equity prices at a daily frequency from July 3rd, 2000, to June 30th, 2010. We follow the same set-up as in Cai *et al.* (2015) to extract, for the same time period, daily loss returns (Y_i) of a value-weighted market index aggregating three markets: the New York Stock Exchange, American Express stock exchange and the National Association of Securities Dealers Automated Quotation system.

Cai *et al.* (2015) used $\widehat{\mathrm{QMES}}^*(\tau_n')$, as defined in (21), to estimate the quantile-based MES, $\mathrm{QMES}(\tau_n') = \mathbb{E}\{X|Y > q_{Y,\tau_n'}\}$, where $\tau_n' = 1 - \frac{1}{n} = 1 - 1/2513$, with two intermediate sequences involved in $\widehat{\gamma}_X$ and $\widehat{\mathrm{QMES}}(\tau_n)$ to be chosen in two steps. Instead, we use our expectile-based method to estimate $\mathrm{XMES}(\tau_n') = \mathbb{E}\{X|Y > \xi_{Y,\tau_n'}\}$, with the same extreme level τ_n' that corresponds to a once-per-decade systemic event. As a benchmark, we employ $\widehat{\mathrm{QMES}}^*(\tau_n')$ with the same intermediate sequence $\tau_n = 1 - \frac{k}{n}$ in both $\widehat{\gamma}_X$ and $\widehat{\mathrm{QMES}}(\tau_n)$. The conditions required by the procedure were already checked empirically in Cai *et al.* (2015). It only remains to verify that $\gamma_Y < \frac{1}{2}$ as it is the case for γ_X. This assumption is confirmed by the plot of the Hill estimates of γ_Y against the sample fraction k (green curve) in Figure 7 (a). Indeed, the first stable region appears for $k \in [70, 100]$ with an averaged estimate $\widehat{\gamma}_Y = 0.349$. Hence, by Proposition 5, our estimates $\widehat{\mathrm{XMES}}^*(\tau_n')$ and $\widetilde{\mathrm{XMES}}^*(\tau_n')$ of $\mathrm{XMES}(\tau_n')$ are expected to be less conservative than the benchmark values $\widehat{\mathrm{QMES}}^*(\tau_n')$. This is visualised in Figure 7 (b)-(d), where the three estimates are graphed as functions of k for each bank: (b) Goldman Sachs; (c) Morgan Stanley; (d) T. Rowe Price. The first stable region of the plots (b)-(d) appears, respectively, for $k \in [85, 105]$, $k \in [85, 115]$ and $k \in [70, 100]$. The final estimates based on averaging the estimates from these stable regions are reported in Table 1. It may be seen that both expectile- and quantile-based MES levels for Morgan Stanley are largely higher than those for Goldman Sachs and T. Rowe Price. It may also be noted that the estimates $\widehat{\mathrm{QMES}}^*(\tau_n')$, obtained here with a single intermediate sequence, are slightly smaller than those obtained in Table 1 of Cai *et al.* (2015) by using two intermediate sequences. These estimates represent the average daily loss return for a once-per-decade market crisis. They are appreciably more pessimistic than the $\widehat{\mathrm{XMES}}^*(\tau_n')$ and $\widetilde{\mathrm{XMES}}^*(\tau_n')$ estimates. This substantial difference is visualised in Figure 7 (b)-(d) for all values of k, where it may also be seen that the curves of $\widehat{\mathrm{XMES}}^*$ (pink) and $\widetilde{\mathrm{XMES}}^*$ (black) show a very similar evolution for the three banks. The effect of the Hill estimates $\widehat{\gamma}_X$ on the MES estimates is highlighted by a "colouring-scheme" of $\widehat{\mathrm{QMES}}^*(\tau_n')$, ranging from dark red (low) to dark violet (high).

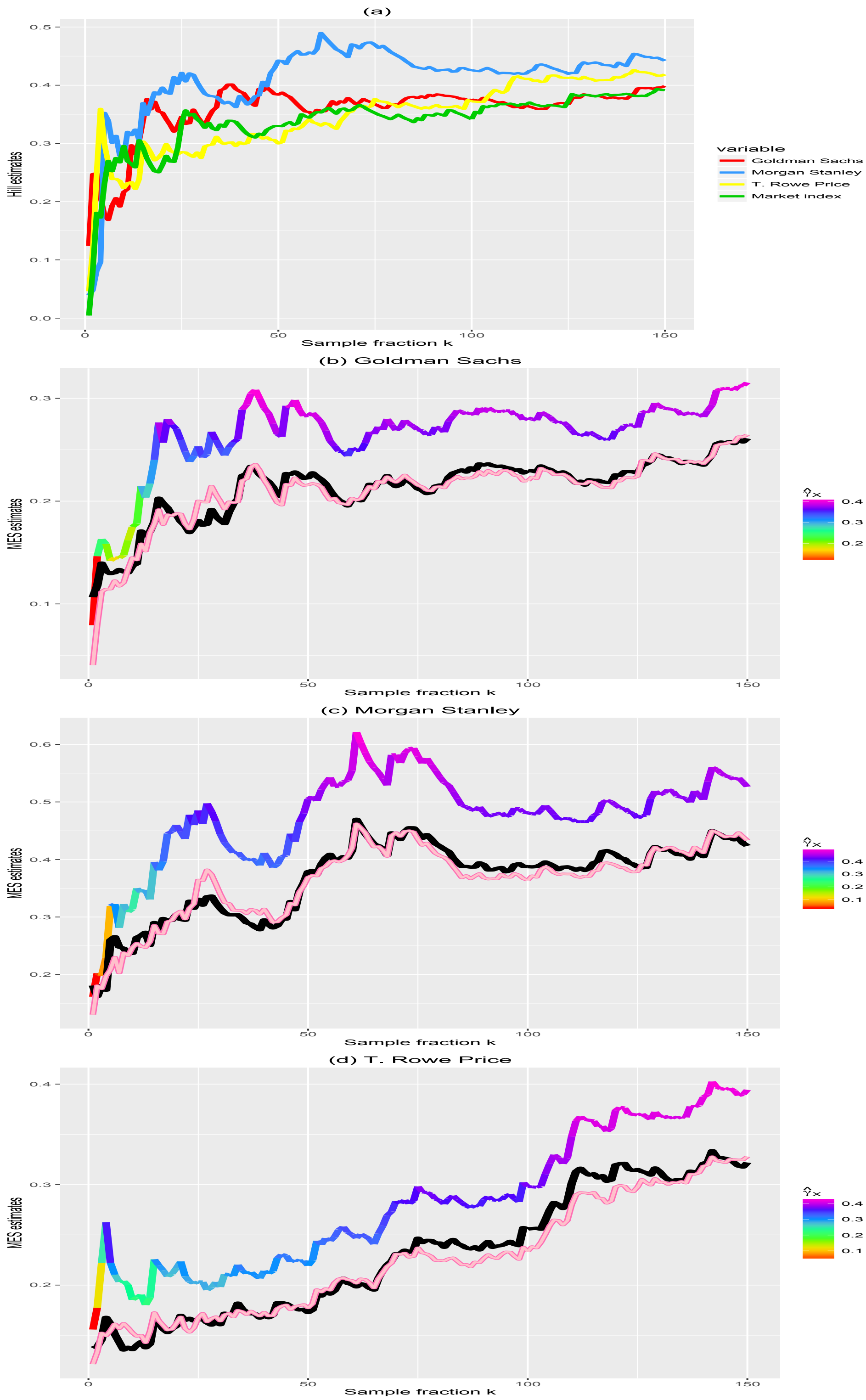

Figure 7: *(a) Hill estimates* $\widehat{\gamma}_Y$ *based on daily loss returns of market index (green), along with* $\widehat{\gamma}_X$ *based on daily loss returns of three investment banks. (b)-(d) The estimates* $\widehat{XMES}^*$ *(pink),* $\widetilde{XMES}^*$ *(black) and* $\widehat{QMES}^*$ *(rainbow) for the three banks.*

Bank	$\widehat{\text{XMES}}^*$	$\widetilde{\text{XMES}}^*$	$\widehat{\text{QMES}}^*$
Goldman Sachs	0.225	0.231	0.285
Morgan Stanley	0.372	0.387	0.480
T. Rowe Price	0.227	0.240	0.286

Table 1: *Expectile- and quantile-based MES of the three investment banks.*

Supplementary materials

The supplement to this article contains additional simulations, technical lemmas and the proofs of all theoretical results in the main article.

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

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

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

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

[6] Bellini, F. and Di Bernardino, E. (2015). Risk Management with Expectiles, *The European Journal of Finance*. DOI:10.1080/1351847X.2015.1052150

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

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

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

[10] Cai, J., Einmahl, 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.

[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] Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function, *Journal of Multivariate Analysis*, **64**, 25–47.

[14] Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. (2015). Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations, and Forecast Rankings. URL arXiv:1503.08195v2.

[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. (2014). Statistics and Quantitative Risk Management for Banking and Insurance, *Annual Review of Statistics and Its Application*, **1**, 493–514.

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

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

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

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

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

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

[23] 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.

[24] Mao, T., Ng, K. and Hu, T. (2015). Asymptotic Expansions of Generalized Quantiles and Expectiles for Extreme Risks, *Probability in the Engineering and Informational Sciences*, **29**, 309–327.

[25] Mao, T. and Yang, F. (2015). Risk concentration based on Expectiles for extreme risks under FGM copula, *Insurance: Mathematics and Economics*, **64**, 429–439.

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

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

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

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

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

[31] Ziegel, J.F. (2014). Coherence and elicitability, *Mathematical Finance*, DOI: 10.1111/mafi.12080 (to appear).

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