Estimation of Tail Risk based on Extreme Expectiles

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

[a] Toulouse School of Economics, University of Toulouse Capitole, France
[b] INRIA Grenoble Rhône-Alpes / LJK Laboratoire Jean Kuntzmann - MISTIS
[c] Aix Marseille Université, CNRS, EHESS, Centrale Marseille, GREQAM UMR 7316, 13002 Marseille, France
[d] School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

Abstract

We use tail expectiles to estimate alternative measures to the Value at Risk (VaR) and Marginal Expected Shortfall (MES), two 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 are M-quantiles as the minimizers of an asymmetric convex loss function, but expectiles are the only M-quantiles that are coherent risk measures. Moreover, expectiles define the only coherent risk measure that is also elicitable. The estimation of expectiles has not, however, received any attention yet from the perspective of extreme values. Two estimation methods are proposed here, either making use of quantiles or relying directly on least asymmetrically weighted squares. A main tool is to first estimate large values of expectile-based VaR and MES located within the range of the data, and then to extrapolate the obtained estimates to the very far tails. We establish the limit distributions of both of the resulting intermediate and extreme estimators. We show via a detailed simulation study the good performance of the procedures, and present concrete applications to medical insurance data and three large US investment banks.

Keywords: Asymmetric squared loss; Coherency; 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 a random variable Y in much the same way that quantiles do. It is a natural generalization of the usual mean $\mathbb{E}(Y)$ just as the class of quantiles generalizes the median. Both expectiles and quantiles are 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. Later, Newey and Powell (1987) substituted the L^1 loss function with the L^2 loss to define the population expectile of order $\tau \in (0, 1)$ as the minimizer

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

where $\eta_\tau(y) = |\tau - \mathbb{1}(y \leqslant 0)| \, y^2$, with $\mathbb{1}(\cdot)$ being the indicator function. Although formulated using a quadratic loss, problem (1) is well-defined as soon as $\mathbb{E}|Y|$ is finite, thanks to the presence of the term $\eta_\tau(Y)$. 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 (see the R package 'expectreg'). The second advantage, following Newey and Powell (1987), Abdous and Remillard (1995) and Sobotka and Kneib (2012), among others, is that inference on expectiles is much easier than inference on quantiles, and their estimation makes more efficient use of the available data since 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. Furthermore, unlike sample quantiles, sample expectiles provide a class of smooth curves as functions of the level τ (see, *e.g.*, Schulze Waltrup *et al.*, 2015).

Value at Risk (VaR) and Marginal Expected Shortfall (MES) are two instruments of risk protection of utmost importance in actuarial science and statistical finance. They are traditionally based on the use of tail quantiles as a main tool for quantifying the riskiness implied by the variability of losses and the tails of their distribution. In this article we focus on the less-discussed problem of estimating the concepts of VaR and MES when quantiles are replaced therein by expectiles. The use of expectiles as an alternative tool for quantifying tail risk has recently attracted a lot of interest, see for instance Martin (2014). Motivating advantages are that expectiles are more alert than quantiles to the magnitude of infrequent catastrophic losses, and they depend on both the tail realizations of Y and their probability, while quantiles only depend on the frequency of tail realizations (see Kuan *et al.*, 2009). Both families of quantiles and expectiles were embedded in the more general class of M-quantiles by Breckling and Chambers (1988). 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. Most importantly, Bellini *et al.* (2014) have proved that the only M-quantiles that are coherent risk measures are the expectiles. Very recently, Ziegel (2016) has proved that expectiles are the only coherent law-invariant measure of risk which is also elicitable. Elicitability corresponds to the existence of a natural backtesting methodology. It has been shown that the Expected Shortfall (ES), viewed as the most popular coherent risk measure, is not elicitable (Gneiting, 2011), but jointly elicitable with VaR (Fissler and Ziegel, 2016).

In terms of interpretability, the τ-quantile is the point below which $100\tau\%$ of the mass of Y lies, while the τ-expectile specifies the position ξ_τ such that the average distance from the data below ξ_τ to ξ_τ itself is $100\tau\%$ of the average distance between ξ_τ and all the data:

$$\tau = \mathbb{E}\left\{|Y - \xi_\tau|\,\mathbb{1}(Y \leqslant \xi_\tau)\right\}/\mathbb{E}\,|Y - \xi_\tau|. \tag{2}$$

Thus, the τ-expectile shares an intuitive interpretation similar to the τ-quantile, replacing

the number of observations by the distance. Bellini and Di Bernardino (2017) provide a transparent financial meaning of expectiles in terms of gain-loss ratio, which is a popular performance measure in portfolio management and is well-known in the literature on no good deal valuation in incomplete markets (see Bellini and Di Bernardino, 2017 and references therein). Also, Ehm *et al.* (2016) have shown that expectiles are optimal decision thresholds in certain binary investment problems. Another motivation for the adoption of expectiles in risk management, according to Taylor (2008), is that they are very closely related to the classical mean and the popular ES. Furthermore, the theoretical and numerical results obtained by Bellini and Di Bernardino (2017) seem to indicate that expectiles are perfectly reasonable alternatives to standard VaR and ES. The statistical problem of expectile estimation has not, however, received any attention yet from the perspective of extreme values, unlike VaR estimation (see de Haan and Ferreira, 2006) and ES estimation (see El Methni *et al.*, 2014).

Although least asymmetrically weighted squares estimation of expectiles dates back to Newey and Powell (1987) in the 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, as well as the two recent contributions by Holzmann and Klar (2016) and Krätschmer and Zähle (2017) for advanced theoretical developments. Attention has been, however, restricted to ordinary expectiles of fixed order τ staying away from the tails of the underlying distribution. In the latter two references, several asymptotic results such as uniform consistency and a uniform central limit theorem are shown for expectile estimators, but the order τ therein is assumed to lie within a fixed interval bounded away from 0 and 1. 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), Mao and Yang (2015) and Bellini and Di Bernardino (2017) 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, however, consider the crucial question of their statistical inference.

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: already Embrechts *et al.* (1997, p.9) pointed out that "claims are mostly modelled by heavy-tailed distributions", and more recently Resnick (2007, p.1) has stated that "Record-breaking insurance losses, financial log-returns [...] are all examples of heavy-tailed phenomena". The rival quantile-based risk measures 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 $(1 - 1/n)$, see for instance Embrechts and Puccetti (2007) who studied extreme operational bank losses, Cai *et al.* (2015) for an application to extreme loss returns of banks in the US market, and El Methni and Stupfler (2017a, 2017b) who estimate excess-of-loss risk measures on automobile insurance data and the average value of a catastrophic loss in commercial fire risk. Therefore, estimating the corresponding quantile-based risk measures is a typical extreme value problem. We refer the reader to the books by Embrechts *et al.* (1997), Beirlant *et al.* (2004), and de Haan and Ferreira (2006) for a general overview of the theoretical background.

Let us point out three main contributions of this paper. First, we 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 nonnegative 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. Second, we provide adapted extreme expectile-based tools for the estimation of the MES, an important factor when measuring the systemic risk of financial firms. Denoting by X and Y, respectively, the loss on the 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 estimators of the MES. A rival procedure by Cai *et al.* (2015) is based on extreme quantiles. Finally, we unravel the important question of how to select theoretically the extreme level τ'_n so that each expectile-based risk measure (VaR and MES) at this level coincides with its quantile-based analogue at a given tail probability $\alpha_n \to 1$ as $n \to \infty$. The obtained level $\tau'_n = \tau'_n(\alpha_n)$ needs itself to be estimated, which results in two final composite estimators of the risk measure. Our main results establish the asymptotic distributions of all presented estimators. To the best of our knowledge, this is the first work to actually join together the expectile perspective with the tail restrictions of extreme-value theory.

We organize this paper as follows. Section 2 discusses the basic properties of the expectile-based VaR including its connection with the standard quantile-VaR for high levels $\tau_n \to 1$. Section 3 presents the two estimation methods of intermediate and extreme expectiles. Section 4 considers the problem of estimating the MES when the related variable is extreme. Section 5 addresses the important question of how to select the extreme expectile level in the three studied risk measures. The good performance of our procedures is shown in Section 6 on a simulation study, and concrete applications to medical insurance data and the loss-returns of three large US investment banks are provided in Section 7. The implemented codes for our methods can be found at `https://hal.archives-ouvertes.fr/hal-01142130`.

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 considered as the negative of a series of financial returns. The right-tail of the distribution of Y then corresponds to the negative of extreme losses. Following Newey and Powell (1987), the expectile ξ_τ of order $\tau \in (0,1)$ of the variable Y can be defined as the minimizer (1) of a piecewise-quadratic loss function or, equivalently, as

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

where $y_+ := \max(y, 0)$ and $y_- := \min(y, 0)$. The presence of terms Y_+^2 and Y_-^2 makes this problem well-defined indeed as soon as $Y \in L^1$ [i.e. $\mathbb{E}|Y| < \infty$]. The related first-order necessary condition for optimality can be written in several ways, one of them being

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

This equation has a unique solution for all $Y \in L^1$. Thenceforth expectiles of a distribution function F_Y with finite absolute first moment are well-defined, and we will assume in the sequel that $\mathbb{E}|Y| < \infty$. Expectiles 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, as well as to quantify the "riskiness" implied by the return distribution under consideration, may be based on the following elementary properties [Newey and Powell (1987), Abdous and Remillard (1995) and Bellini *et al.* (2014)]:

(i) Law invariance: two integrable random variables Y and $\widetilde{Y}$, with continuous densities, have the same distribution 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 $Y = c \in \mathbb{R}$ with probability 1, 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 $Y \leq \widetilde{Y}$ with probability 1, then $\xi_{Y,\tau} \leq \xi_{\widetilde{Y},\tau}$ for any τ.

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

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

The convention we have chosen for values of Y as the negative of returns implies that extreme losses correspond to levels τ close to 1. Only Bellini *et al.* (2014), Mao *et al.* (2015), Mao and Yang (2015) and Bellini and Di Bernardino (2017) have described what happens for large population expectiles ξ_τ and their link with high 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

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

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 index γ tunes the tail heaviness of the distribution function F_Y, whose first moment does not exist when $\gamma > 1$. For most applicational purposes in risk management, it has been found in previous studies that the class of heavy-tailed distributions describes sufficiently well the tail structure of actuarial and financial data: in addition to the monographs of Embrechts *et al.* (1997) and Resnick (2007), see for instance Chavez-Demoulin *et al.* (2014) and the references therein. See also Alm (2016) for a recent study in the context of the Swedish insurance market.

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 $\overline{F}_Y(\xi_\tau)/(1 - \tau) \sim \gamma^{-1} - 1$ as $\tau \to 1$. It follows that extreme expectiles ξ_τ are larger than extreme quantiles q_τ (*i.e.* $\xi_\tau > q_\tau$) when $\gamma > \frac{1}{2}$, whereas $\xi_\tau < q_\tau$ when $\gamma < \frac{1}{2}$, for all large τ. The connection (4) between high expectiles and quantiles can actually be refined appreciably by considering a strengthened yet classical version of condition (3). Assume that the tail quantile function U of Y, namely the left-continuous inverse of $1/\overline{F}_Y$, defined by $U(t) = \inf\{y \in \mathbb{R} \,|\, 1/\overline{F}_Y(y) \geqslant t\}$, is such that there exist $\gamma > 0$, $\rho \leqslant 0$, and a function $A(\cdot)$ converging to 0 at infinity and having constant sign such that

$\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{A})$ 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 read as $\log x$ when $\rho = 0$. The interpretation of this condition can be found in Beirlant *et al.* (2004) and de Haan and Ferreira (2006) along with abundant examples of commonly used continuous distributions satisfying $\mathcal{C}_2(\gamma, \rho, A)$: for instance, this is the case for any distribution whose distribution function F satisfies

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

where a and b are positive constants and $\rho < 0$. This contains in particular the Hall-Weiss class of models (see Hua and Joe, 2011).

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

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

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

One can actually establish the precise bias term in the asymptotic expansion of ξ_τ / q_τ.

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

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

$$\text{with } r(\tau) = \frac{\gamma(\gamma^{-1} - 1)^\gamma}{q_\tau}(\mathbb{E}(Y) + \mathrm{o}(1))$$

$$+ \left(\frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \rho - \gamma} + \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho} + \mathrm{o}(1)\right)A((1 - \tau)^{-1}) \text{ as } \tau \uparrow 1.$$

Other refinements under similar second order regular variation conditions can also be found in Mao *et al.* (2015) and Mao and Yang (2015). An extension to a subset of the challenging Gumbel domain of attraction is also derived in Proposition 2.4 in Bellini and Di Bernardino (2017). 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 have been widely used in applied statistical analyses and extensively investigated 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. This motivated us to construct estimators of large expectiles ξ_{τ_n} and derive their limit distributions when they are located within or beyond the range of the data, where their empirical counterparts usually fail due to data sparseness. We shall assume the extended regular variation condition $\mathcal{C}_2(\gamma, \rho, A)$ to obtain our convergence results.

3 Estimation of the expectile-based VaR

Our main objective in this section is to estimate ξ_{τ_n} for high levels τ_n that may approach 1 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 nonnegative constant c. We assume that the available data consist of independent copies $(Y_1, \ldots, Y_n)$ of Y, and denote by $Y_{1,n} \leqslant \cdots \leqslant Y_{n,n}$ their ascending order statistics.

3.1 Intermediate expectile estimation

3.1.1 Estimation based on intermediate quantiles

The rationale for this first method relies on the heavy-tailed property (3) and on the asymptotic equivalence (4). Given that F_Y satisfies (3), we expect that (4) entails:

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

This result is actually an immediate consequence of Corollary 1 above and can be found in Proposition 2.3 of Bellini and Di Bernardino (2017) as well. Therefore, for a suitable estimator $\widehat{\gamma}$ of γ, we suggest estimating the intermediate expectile ξ_{τ_n} by $\widehat{\xi}_{\tau_n} := (\widehat{\gamma}^{-1} - 1)^{-\widehat{\gamma}} \widehat{q}_{\tau_n}$, where $\widehat{q}_{\tau_n} := Y_{n-\lfloor n(1-\tau_n)\rfloor, n}$ and $\lfloor \cdot \rfloor$ stands for the floor function. This estimator is based on the intermediate quantile-VaR $\widehat{q}_{\tau_n}$ and crucially hinges on the estimated tail index $\widehat{\gamma}$. A simple and widely used estimator $\widehat{\gamma}$ is the popular Hill estimator (Hill, 1975):

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

where $k = k(n)$ is an intermediate sequence, namely $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$. Beirlant *et al.* (2004) and de Haan and Ferreira (2006) give a nice summary of the properties of $\widehat{\gamma}_H$ and review other efficient estimation methods with an extensive bibliography.

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 $k = \lceil n(1 - \tau_n) \rceil$, 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 centered Gaussian distribution with variance γ^2. 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 for $\widehat{\gamma} = \widehat{\gamma}_H$.

Corollary 2. *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 $n(1 - \tau_n) \to \infty$. 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*

$$\sqrt{n(1 - \tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 \right) \xrightarrow{d} \mathcal{N}\left(\frac{m(\gamma)}{1 - \rho}\lambda_2 - \lambda,\, v(\gamma) \right),$$

with $m(\gamma)$ and λ as in Theorem 1, and

$$v(\gamma) = \gamma^2 \left[1 + \left(\frac{1}{1 - \gamma} - \log\left(\frac{1}{\gamma} - 1 \right) \right)^2 \right].$$

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 $\widehat{q}_{\tau_n}$; it should be pointed out though that one may design a bias-reduced version of the estimator $\widehat{\xi}_{\tau_n}$. Indeed, the bias components λ_1 and λ_2 appearing in Theorem 1 can be estimated, respectively, by using $\widehat{\lambda}_1 = \sqrt{n(1 - \tau_n)}\widehat{q}_{\tau_n}^{-1}$ and by applying the methodology of Caeiro *et al.* (2005) in conjunction with the Hall-Welsh class of models to get an estimator $\widehat{\lambda}_2$ of λ_2. Plugging these, along with the empirical mean $\overline{Y}$, the estimator $\widehat{\gamma}$, and a consistent estimator $\widehat{\rho}$ of the second-order parameter ρ (a review of possible estimators $\widehat{\rho}$ is given in Gomes and Guillou, 2015), into the expression of λ, we get a consistent estimator $\widehat{\lambda}$ of this bias component. This in turn enables one to define a bias-reduced version of $\widehat{\xi}_{\tau_n}$, for instance as

$$\widehat{\xi}_{\tau_n}^{RB} := \widehat{\xi}_{\tau_n} \left(1 - \left[\frac{m(\widehat{\gamma})}{1 - \widehat{\rho}}\widehat{\lambda}_2 - \widehat{\lambda} \right] \frac{1}{\sqrt{n(1 - \tau_n)}} \right).$$

Of course, one should expect the value of the asymptotic variance of this estimator to be even higher than that of $\widehat{\xi}_{\tau_n}$, similarly to what is observed when bias reduction techniques are applied to the Hill estimator (see *e.g.* Theorem 3.2 in Caeiro *et al.*, 2005).

Another efficient way of estimating ξ_{τ_n}, which we develop in the next section, is by joining together 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"* of the R package *'expectreg'*. Clearly

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

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

Note that (ψ_n) is a sequence of almost surely continuous and convex random functions. 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. Let us recall that we denote by Y_- the negative part of Y, *i.e.*, $Y_- = \min(Y, 0)$.

Theorem 2. *Assume that there is $\delta > 0$ such that $\mathbb{E}|Y_-|^{2+\delta} < \infty$, that $0 < \gamma < 1/2$ and $\tau_n \uparrow 1$ is such that $n(1-\tau_n) \to \infty$. Then*

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

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. A mild moment assumption and the condition $0 < \gamma < 1/2$ suffice. It has been found in many instances (*e.g.* recently by Chavez-Demoulin *et al.*, 2014, Cai *et al.*, 2015, Alm, 2016, El Methni and Stupfler, 2017a; see also the R package 'CASdatasets') that the model assumption of Pareto-type tails along with these finite-variance conditions deliver competitive results for most applicational purposes in risk management; our findings in Section 7 also go in this sense. 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 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 from the left panel that both asymptotic variances are stable and close for values of $\gamma < 0.3$, with an advantage for $V(\gamma)$ in dashed line as visualized more clearly in the right panel. Then $V(\gamma)$ becomes appreciably larger than $v(\gamma)$ for $\gamma > 0.3$ and explodes in a neighborhood of $1/2$, while $v(\gamma)$ in solid line remains lower than the level 1.25.
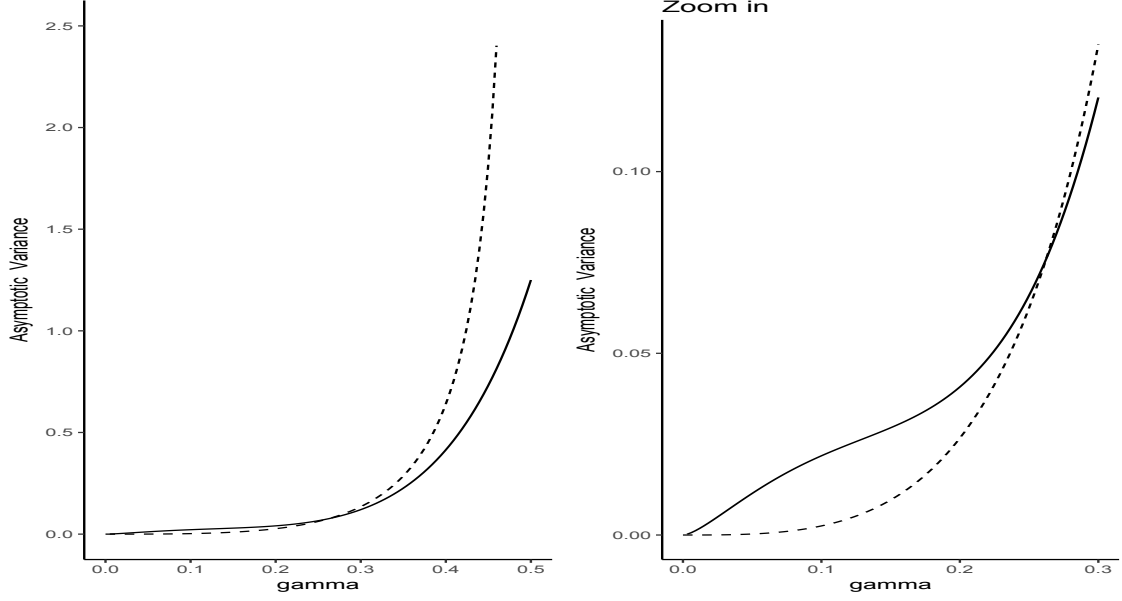
Figure 1: *Asymptotic variances $V(\gamma)$ of the LAWS estimator $\widetilde{\xi}_{\tau_n}$ in dashed line, and $v(\gamma)$ of the indirect estimator $\widehat{\xi}_{\tau_n}$ in solid line. From left to right, $\gamma \in (0, 1/2)$ and $\gamma < 0.3$.*

3.2 Extreme expectile estimation

We now discuss the important issue of estimating extreme tail expectiles $\xi_{\tau'_n}$, where $\tau'_n \to 1$ 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 not greater than $1/n$, and $\tau_n = 1 - k(n)/n$ 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, using (5), suggests that when τ_n and τ'_n satisfy suitable conditions, we may write:

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

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

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

where $\widehat{\gamma}$ is an estimator of γ, and $\overline{\xi}_{\tau_n}$ is either the estimator $\widehat{\xi}_{\tau_n}$ or $\widetilde{\xi}_{\tau_n}$ of the intermediate expectile ξ_{τ_n}. We actually have $\overline{\xi}^{\star}_{\tau'_n}/\overline{\xi}_{\tau_n} = \widehat{q}^{\star}_{\tau'_n}/\widehat{q}_{\tau_n}$, where $\widehat{q}_{\tau_n} = Y_{n-\lfloor n(1-\tau_n)\rfloor, n}$ is the intermediate quantile estimator introduced above, and $\widehat{q}^{\star}_{\tau'_n}$ is the Weissman extreme quantile estimator

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

11

We then 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 $0 < \gamma < 1$ and $\rho < 0$, and that τ_n, $\tau'_n \uparrow 1$, with $n(1 - \tau_n) \to \infty$ and $n(1 - \tau'_n) \to c < \infty$. If moreover*

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

with $\sqrt{n(1 - \tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$, $\sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$ and $\sqrt{n(1 - \tau_n)}/\log[(1 - \tau_n)/(1 - \tau'_n)] \to \infty$, 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 $\widehat{\xi}^\star_{\tau'_n} := \overline{\xi}^\star_{\tau'_n}$ being

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

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 $\sqrt{n(1 - \tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$, $\sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$ and $\sqrt{n(1 - \tau_n)}/\log[(1 - \tau_n)/(1 - \tau'_n)] \to \infty$, then

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

Corollary 4. *Assume that F_Y is strictly increasing, there is $\delta > 0$ such that $\mathbb{E}|Y_-|^{2+\delta} < \infty$, 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 $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to \lambda_1 \in \mathbb{R}$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_2 \in \mathbb{R}$ *and* $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$, *then*

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

4   Marginal expected shortfall

4.1   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), Engle *et al.* (2015) and Brownlees and Engle (2017) define systemic risk as the propensity of a financial institution to be undercapitalized when the financial system as a whole is undercapitalized. 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, namely 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 the firm's return X conditional on the loss Y in the aggregated return of the financial market being extreme. More specifically, the MES at probability level $(1-\tau)$ is defined as

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

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 $\text{QMES}(\tau)$. The estimation procedure in Acharya *et al.* (2012) relies on daily data from only one year and assumes a specific linear relationship between X and Y. A nonparametric kernel estimation method has been performed in Engle *et al.* (2015) and Brownlees and Engle (2017), 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 $\text{QMES}(\tau)$ without recourse to any parametric structure on (X,Y). Here, instead of the extreme τth quantile $q_{Y,\tau}$, we shall explore the use of the τth expectile analogue $\xi_{Y,\tau}$ in the marginal expected shortfall

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

at least for the following reason: as claimed by Newey and Powell (1987), Kuan *et al.* (2009) and Sobotka and Kneib (2012) among others, expectiles make a more efficient use of

the available data since they rely on the distance of observations from the predictor, 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. An interesting asymptotic connection between $\text{XMES}(\tau)$ and $\text{QMES}(\tau)$ is provided below in Proposition 2. The overall objective is to establish estimators of the tail expectile-based MES and unravel their asymptotic behavior in a general setting.

4.2 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, assumed to be increasing in what follows. Given that our goal is to estimate $\text{XMES}(\tau)$ 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:

$\mathcal{JC}(R)$ 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 $R(1, 1) > 0$. Here $\overline{F}_X = 1 - F_X$ 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 $z > 0$,

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

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 $\text{XMES}(\tau)$.

Proposition 2. *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*

$$\lim_{\tau \uparrow 1} \frac{XMES(\tau)}{U_X(1/\overline{F}_Y(\xi_{Y,\tau}))} \ = \ \int_0^\infty R(x^{-1/\gamma_X}, 1)dx \tag{13}$$

$$and \ \ \lim_{\tau \uparrow 1} \frac{XMES(\tau)}{QMES(\tau)} \ = \ \left(\gamma_Y^{-1} - 1\right)^{-\gamma_X}. \tag{14}$$

14

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 positive multiplicative constant. Since as usual in the financial setting $0 < \gamma_X, \gamma_Y < 1/2$, the second result shows that XMES(τ) is less extreme than QMES(τ) as $\tau \to 1$. This is visualised in Figure 2 in the case of the restriction of the standard bivariate Student t_ν-distribution to $(0, \infty)^2$, having the density

$$f_\nu(x, y) = \frac{2}{\pi} \left(1 + \frac{x^2 + y^2}{\nu} \right)^{-(\nu+2)/2} \quad , \quad x, y > 0, \tag{15}$$

where $\nu = 3$ on the left panel and $\nu = 5$ on the right panel. It can be seen that QMES(τ) becomes overall much more extreme than XMES(τ) as τ approaches 1.

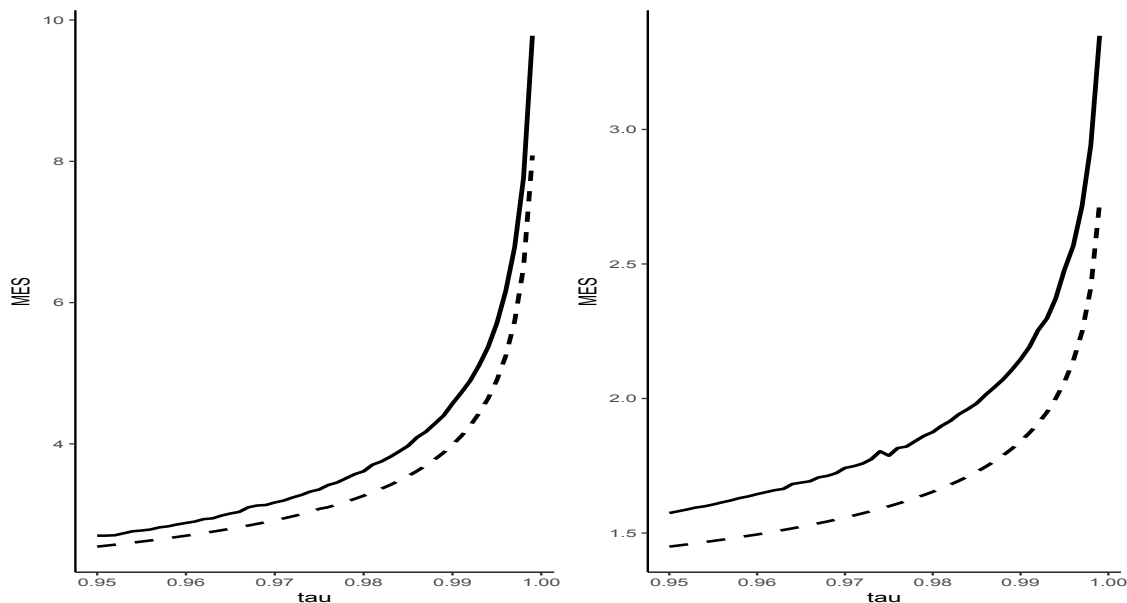Figure 2: *QMES(τ) in solid line and XMES(τ) in dashed line, as functions of $\tau \in [0.95, 1)$. Case of the Student t_ν-distribution on $(0, \infty)^2$. Left: $\nu = 3$, right: $\nu = 5$.*

4.3   Estimation and results

The asymptotic equivalences in Proposition 2 are of particular interest when it comes to proposing estimators for the tail expectile-based MES. Two approaches will be distinguished. We consider first asymmetric least squares estimation by making use of the asymptotic equivalence (13). Subsequently we shall deal with a nonparametric estimator derived from the asymptotic connection (14) with the tail quantile-based MES.

4.3.1 Asymmetric least squares estimation

On the basis of the limit (13) and then of the heavy-tailedness assumption on X, we have for $\tau < \tau' < 1$ that, as $\tau \to 1$,

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

It follows then from Proposition 1 that

$$\mathrm{XMES}(\tau') \approx \left(\frac{1-\tau'}{1-\tau}\right)^{-\gamma_X} \mathrm{XMES}(\tau). \tag{16}$$

Hence, to estimate $\mathrm{XMES}(\tau')$ at an arbitrary extreme level $\tau' = \tau'_n$, we first consider the estimation of $\mathrm{XMES}(\tau)$ at an intermediate level $\tau = \tau_n$, and then we use the extrapolation technique of Weissman (1978). For estimating $\mathrm{XMES}(\tau_n) = \mathbb{E}\{X|Y > \xi_{Y,\tau_n}\}$ 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

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

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 $\widetilde{\mathrm{XMES}}(\tau_n)$ is $\sqrt{n(1-\tau_n)}$-relatively consistent. By plugging this estimator into approximation (16) together with a $\sqrt{n(1-\tau_n)}$-consistent estimator $\widehat{\gamma}_X$ of γ_X, we obtain the following estimator of $\mathrm{XMES}(\tau'_n)$:

$$\widetilde{\mathrm{XMES}}^\star(\tau'_n) \equiv \widetilde{\mathrm{XMES}}^\star(\tau'_n; \tau_n) := \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}_X} \widetilde{\mathrm{XMES}}(\tau_n). \tag{17}$$

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 $\mathrm{QMES}(\tau'_n)$ 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 $\widetilde{\mathrm{XMES}}^\star(\tau'_n)$. 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, that there is $\delta > 0$ such that $\mathbb{E}|Y_-|^{2+\delta} < \infty$, and that 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*

(i) *$\tau_n, \tau_n' \uparrow 1$, with $n(1 - \tau_n) \to \infty$, $n(1 - \tau_n') \to c < \infty$ and $\sqrt{n(1 - \tau_n)}/\log[(1 - \tau_n)/(1 - \tau_n')] \to \infty$ as $n \to \infty$;*

(ii) *$1 - \tau_n = O(n^{\alpha - 1})$ for some $\alpha < \min\left(\dfrac{-2\kappa}{-2\kappa + 1}, \dfrac{2\gamma_X \rho_X}{2\gamma_X \rho_X + \rho_X - 1}\right)$;*

(iii) *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;*

(iv) *$\sqrt{n(1 - \tau_n)}(\hat{\gamma}_X - \gamma_X) \xrightarrow{d} \Gamma$.*

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

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

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

(v) *$\mathbb{E}|X_-|^{1/\gamma_X} < \infty$;* (18)

(vi) *$n(1 - \tau_n) = o\left((1 - \tau_n')^{2\kappa(1 - \gamma_X)}\right)$ as $n \to \infty$.* (19)

Let us point out here that condition (ii), which also appears in Theorem 1 of Cai *et al.* (2015), is a strengthening of the condition $1 - \tau_n = o(1)$. It essentially allows to control additional bias terms that appear in conditions $\mathcal{JC}_2(R, \beta, \kappa)$ and $\mathcal{C}_2(\gamma_X, \rho_X, A_X)$. Condition (vi), which is also utilized in Cai *et al.* (2015), is another bias condition that makes it possible to control the bias coming from the left tail of X.

4.3.2   Estimation based on tail QMES

On the basis of the limit (14), we consider the alternative estimator

$$\widehat{XMES}^\star(\tau_n') := \left(\hat{\gamma}_Y^{-1} - 1\right)^{-\hat{\gamma}_X} \widehat{QMES}^\star(\tau_n'),$$ (20)

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

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

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

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

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 use 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 $\widehat{\mathrm{XMES}}^{\star}(\tau_n')$.

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*

(i) τ_n, $\tau_n' \uparrow 1$, with $n(1 - \tau_n) \to \infty$, $n(1 - \tau_n') \to c < \infty$ and $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$ as $n \to \infty$;

(ii) $1 - \tau_n = \mathrm{O}(n^{\alpha-1})$ for some $\alpha < \min\left(\dfrac{-2\kappa}{-2\kappa+1}, \dfrac{2\gamma_X\rho_X}{2\gamma_X\rho_X + \rho_X - 1}\right)$;

(iii) 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;

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

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

This convergence remains still valid if $X \in \mathbb{R}$ provided that (18) and (19) hold.

5 Extreme expectile level selection

An important question that remains to be addressed is the choice of the extreme expectile level τ_n' in the instruments of risk protection $\xi_{\tau_n'}$ and $\mathrm{XMES}(\tau_n')$.

In the case of quantile-based risk measures q_{α_n} and $\mathrm{QMES}(\alpha_n)$, it is customary to choose tail probabilities $\alpha_n \to 1$ with $n(1 - \alpha_n) \to c$, a finite constant, as the sample size $n \to \infty$, to allow for more 'prudent' risk management. In response to the many turbulent episodes that have been experienced by financial markets during the last few decades, academics are nowadays more interested in once-in-a-decade or twice-per-decade events (see, *e.g.*, Cai *et al.*, 2015 and Brownlees and Engle, 2017). In the case of expectiles, we propose to select τ_n' so that each expectile-based risk measure has the same intuitive interpretation as its quantile-based analogue. This translates into choosing τ_n' such that $\xi_{\tau_n'} \equiv q_{\alpha_n}$ for a given relative frequency α_n. Bellini and Di Bernardino (2017) have already suggested to pick out τ_n' which satisfies $\xi_{\tau_n'} \equiv q_{\alpha_n}$, but for a normally distributed Y. Here, we wish to extend this elegant device to a general random variable Y without any a priori specification.

Thanks to the connection (2), it is immediate from $\xi_{\tau_n'} \equiv q_{\alpha_n}$ that $\tau_n'(\alpha_n) := \tau_n'$ satisfies

$$1 - \tau_n'(\alpha_n) = \frac{\mathbb{E}\left\{|Y - q_{\alpha_n}| \; \mathbb{1}(Y > q_{\alpha_n})\right\}}{\mathbb{E}\,|Y - q_{\alpha_n}|}.$$

As a matter of fact, under the model assumption of Pareto-type tails, it turns out that the expectile level $\tau_n'(\alpha_n)$ depends asymptotically only on the quantile level α_n and on the tail index γ, but not on the quantile q_{α_n} itself.

Proposition 3. *Suppose F_Y satisfies (3) with $0 < \gamma < 1$. Then*

$$1 - \tau_n'(\alpha_n) \sim (1 - \alpha_n)\frac{\gamma}{1 - \gamma}, \quad as \;\; n \to \infty.$$

Hence, by substituting the estimated value

$$\widehat{\tau}_n'(\alpha_n) = 1 - (1 - \alpha_n)\frac{\widehat{\gamma}}{1 - \widehat{\gamma}} \tag{22}$$

in place of τ_n', both extreme expectile estimators $\widehat{\xi}_{\tau_n'}^\star$ in (11) and $\widetilde{\xi}_{\tau_n'}^\star$ in (12) estimate the same Value at Risk $\xi_{\tau_n'(\alpha_n)} \equiv q_{\alpha_n}$ as the Weissman quantile estimator $\widehat{q}_{\alpha_n}^\star$ in (10). It is easily seen that the latter estimator is actually identical to the indirect expectile estimator $\widehat{\xi}_{\widehat{\tau}_n'(\alpha_n)}^\star$. Indeed, we have in view of (10), (11) and (22) that

$$\widehat{\xi}_{\widehat{\tau}_n'(\alpha_n)}^\star = \left(\widehat{\gamma}^{-1} - 1\right)^{-\widehat{\gamma}}\left(\frac{1 - \widehat{\tau}_n'(\alpha_n)}{1 - \tau_n}\right)^{-\widehat{\gamma}}\widehat{q}_{\tau_n} = \widehat{q}_{\alpha_n}^\star.$$

This quantile-based estimator $\widehat{q}_{\alpha_n}^\star \equiv \widehat{\xi}_{\widehat{\tau}_n'(\alpha_n)}^\star$ may be criticized for its reliance on a single order statistic $\widehat{q}_{\tau_n} = Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$, and hence because it may not respond properly to the very extreme losses. By contrast, the direct expectile-based estimator $\widetilde{\xi}_{\widehat{\tau}_n'(\alpha_n)}^\star$ relies on the asymmetric least squares estimator $\widetilde{\xi}_{\tau_n}$, and hence bears much better the burden of representing a sensitive risk measure to the magnitude of infrequent catastrophic losses. The next result shows that the asymptotic behavior of the original extrapolated estimators $\widehat{\xi}_{\tau_n'}^\star$ and $\widetilde{\xi}_{\tau_n'}^\star$, established in Corollaries 3 and 4, remains still valid for the resulting composite estimators $\widehat{\xi}_{\widehat{\tau}_n'(\alpha_n)}^\star$ and $\widetilde{\xi}_{\widehat{\tau}_n'(\alpha_n)}^\star$, under the same technical conditions.

Theorem 6. *(i) Suppose the conditions of Corollary 3 hold with α_n in place of τ_n'. Then*

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

(ii) Suppose the conditions of Corollary 4 hold with α_n in place of τ_n'. Then

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

19

Let us now turn to $\widetilde{XMES}^{\star}(\hat{\tau}'_n(\alpha_n))$ in (17) and $\widehat{XMES}^{\star}(\hat{\tau}'_n(\alpha_n))$ in (20) that estimate the same marginal expected shortfall $XMES(\tau'_n(\alpha_n)) \equiv QMES(\alpha_n)$ as Cai *et al.* (2015)'s estimator $\widehat{QMES}^{\star}(\alpha_n)$ defined in (21). Actually, $\widehat{XMES}^{\star}(\hat{\tau}'_n(\alpha_n))$ is nothing but $\widehat{QMES}^{\star}(\alpha_n)$.

Theorem 7. *(i) Suppose the conditions of Theorem 4 hold with α_n in place of τ'_n. Then*

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

(ii) Suppose the conditions of Theorem 5 hold with α_n in place of τ'_n. Then

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

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 expectile estimators in Section 6.1 and tail XMES estimators in Section 6.2. Both sections provide Monte-Carlo evidence that the direct estimation method is more efficient relative to the indirect method in the case of real-valued profit-loss variables, whereas the rival indirect method tends to be the winner in the case of non-negative loss distributions. The latter method seems to be also superior in the case of extremely heavy tails.

6.1 Expectile-based VaR

To evaluate the finite-sample performance of the extreme expectile estimators $\widetilde{\xi}^{\star}_{\tau'_n} \equiv \widetilde{\xi}^{\star}_{\tau'_n}(\tau_n)$ and $\widehat{\xi}^{\star}_{\tau'_n} \equiv \widehat{\xi}^{\star}_{\tau'_n}(\tau_n)$, we have considered simulated samples from the Student t_ν-distribution ($\nu = 3, 5, 7, 9$), which corresponds to real-valued profit-loss variables, and from the marginal of the bivariate Student t_ν-distribution described in (15), which corresponds to non-negative loss variables. We shall refer to this marginal distribution on $(0, \infty)$ as 'positive Student t_ν-distribution'. 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 here the results for $n = 1000$ and $\nu \in \{3, 5\}$, a full comparison including additional results for optimal k is given in Supplement A.1.

In the case of Student t-distributions, Figure 3 gives the root Mean-Squared Error (MSE) in top panels and bias estimates in bottom panels, 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 $\widehat{\xi}^{\star}_{\tau'_n}(k)/\xi_{\tau'_n}$, as functions of the effective

sample size k. Our tentative conclusion is that the accuracy of the direct estimator $\widetilde{\xi}^{\star}_{\tau'_n}$ is quite respectable relative to the indirect estimator $\widehat{\xi}^{\star}_{\tau'_n}$. Our experience with other simulated data indicates, however, that the direct estimator is no longer the winner in the case of extremely heavy-tailed distributions such as, for instance, Student t_ν-distributions with $1 < \nu \leqslant 2$.
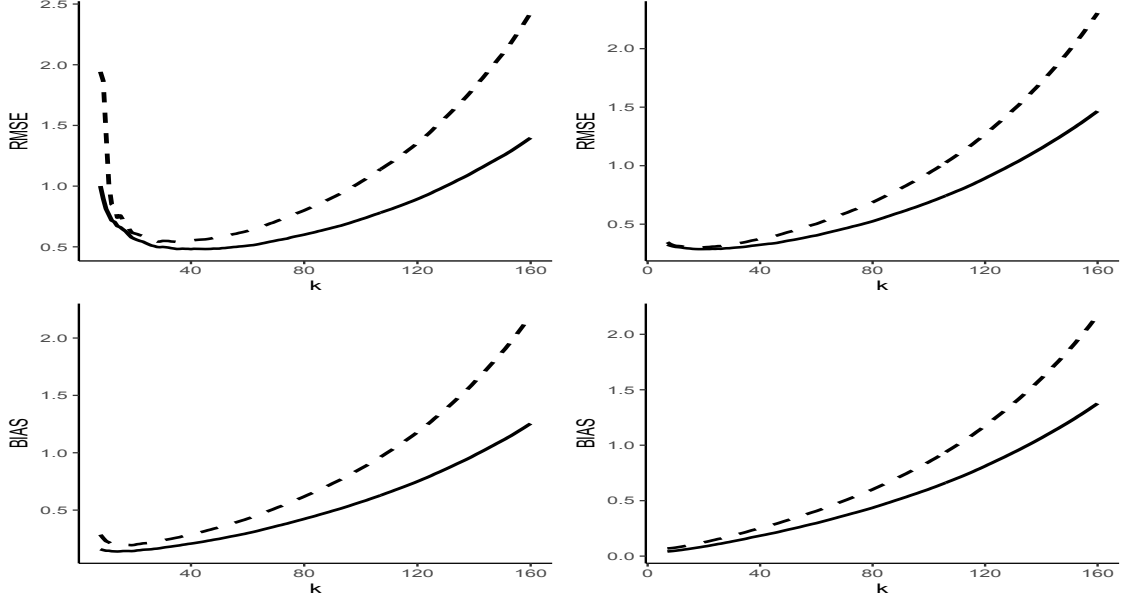

Figure 3: *Root MSE estimates (top panels) and Bias estimates (bottom panels) of $\widetilde{\xi}^{\star}_{\tau'_n}(k)/\xi_{\tau'_n}$ (solid line) and $\widehat{\xi}^{\star}_{\tau'_n}(k)/\xi_{\tau'_n}$ (dashed line), as functions of k, for the t_3 and t_5-distributions, respectively, from left to right.*

The resulting Monte-Carlo estimates in the case of positive Student distributions, displayed in Figure 4, indicate that the indirect estimator $\widehat{\xi}^{\star}_{\tau'_n}$ is superior to the direct estimator $\widetilde{\xi}^{\star}_{\tau'_n}$: the use of the Pareto distribution $F_Y(y) = 1 - y^{-1/\gamma}$, $y > 1$, and of the Fréchet distribution $F_Y(y) = e^{-y^{-1/\gamma}}$, $y > 0$, lead to the same conclusion. It may also be seen in both Student and positive Student scenarios that most of the error is due to variance, the squared bias being much smaller in all cases. 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. It is interesting that in almost all cases the bias was positive.

We also investigate the normality of the estimators $\widehat{\xi}^{\star}_{\tau'_n}$ and $\widetilde{\xi}^{\star}_{\tau'_n}$ in Supplement A.2, where the Q–Q-plots indicate that the limit Theorem 3 and its Corollaries 3 and 4 provide adequate approximations for finite sample sizes.

6.2 Expectile-based MES

Here, we compare the composite estimators $\widetilde{\mathrm{XMES}}^{\star}(\widehat{\tau}'_n(\alpha_n))$ and $\widehat{\mathrm{XMES}}^{\star}(\widehat{\tau}'_n(\alpha_n))$ that estimate the same MES, $\mathrm{XMES}(\tau'_n(\alpha_n)) \equiv \mathrm{QMES}(\alpha_n)$, as the Cai *et al.* (2015) estimator
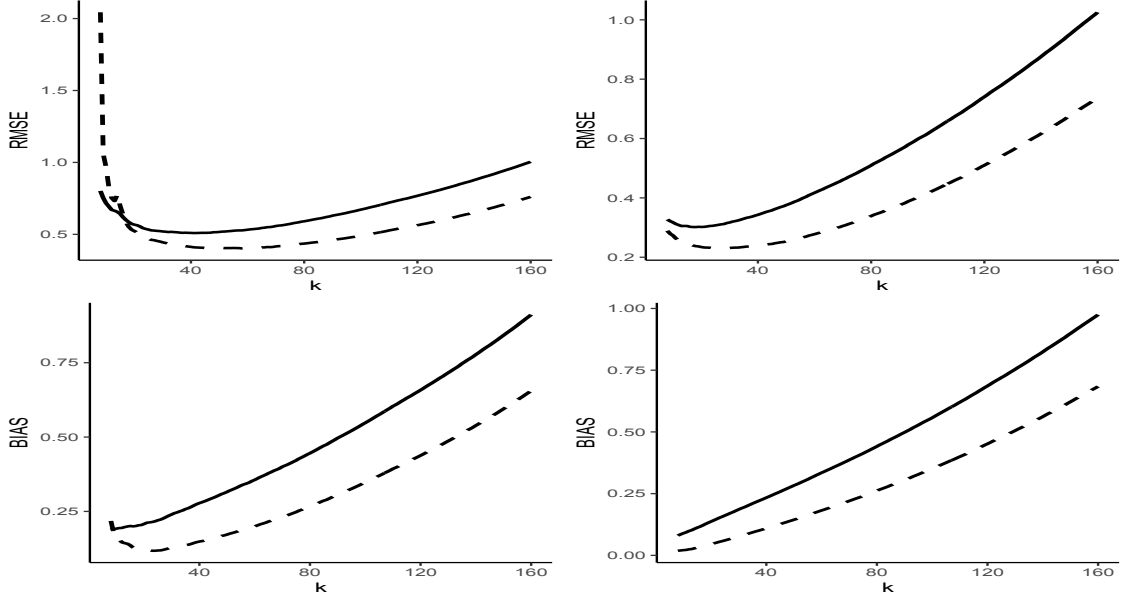
Figure 4: *As before—Results for the positive Student t_3 and t_5-distributions.*

$\widehat{\text{QMES}}^{\star}(\alpha_n)$. The latter is actually identical to the indirect estimator $\widehat{\text{XMES}}^{\star}(\hat{\tau}'_n(\alpha_n))$. All the experiments have sample size $n = 1000$ and extreme level $\alpha_n = 0.9994$.

To investigate the finite sample performance of the two rival estimators $\widetilde{\text{XMES}}^{\star}(\hat{\tau}'_n(\alpha_n))$ and $\widehat{\text{XMES}}^{\star}(\hat{\tau}'_n(\alpha_n))$, the simulation experiments first employ the Student t_ν-distribution on $(0, \infty)^2$ with density $f_\nu(x, y)$ described in (15). 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 $\nu \in \{3, 5, 7, 9\}$. As they point towards the same conclusions, we only present the results for $\nu = 3, 5$. For the choice of the intermediate level τ_n, we used the same considerations as in Section 6.1.

In Figure 5 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 $\widetilde{\text{XMES}}^{\star}(\hat{\tau}'_n(\alpha_n))/\text{XMES}(\tau'_n(\alpha_n))$ and $\widehat{\text{XMES}}^{\star}(\hat{\tau}'_n(\alpha_n))/\text{XMES}(\tau'_n(\alpha_n))$, as functions of the effective sample size k. We observe that the latter indirect estimator is clearly the winner in all cases in terms of both root-MSE and bias. As can also be seen in Supplement A.2, the limit Theorems 4 and 5 provide adequate approximations for finite sample sizes, with a slight advantage for $\widehat{\text{XMES}}^{\star}(\hat{\tau}'_n(\alpha_n))$.

To illustrate the case of real-valued profit-loss random variables, we consider a transformed Student t_ν-distribution on the whole of the plane $\mathbb{R}^2$ defined as

$$(X, Y) = \left(Z_1^{\nu/4} \mathbb{1}(Z_1 \geqslant 0) - (-Z_1)^{\nu/8} \mathbb{1}(Z_1 < 0), Z_2 \right),$$

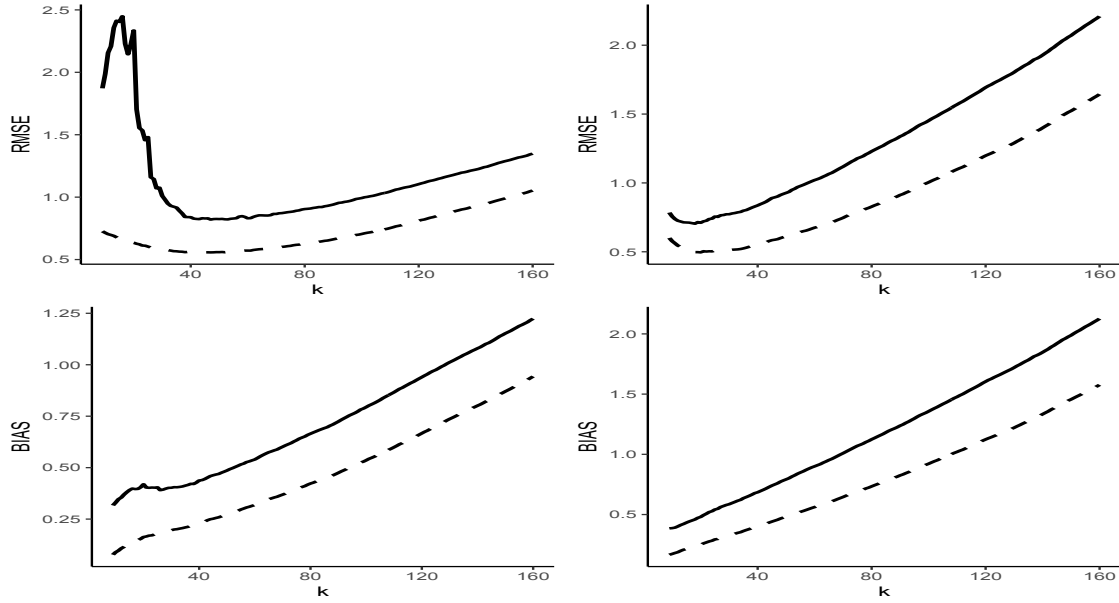where (Z_1, Z_2) is a random pair having a standard Student t_ν-distribution on $\mathbb{R}^2$, namely,

Figure 5: *Root MSE estimates (top panels) and Bias estimates (bottom panels) of $\widetilde{XMES}^\star/XMES$ (solid line) and $\widehat{XMES}^\star/XMES$ (dashed line), as functions of k, for the bivariate t_3 and t_5-distributions on $(0,\infty)^2$, respectively, from left to right.*

with density $(2\pi)^{-1}\left(1+(x^2+y^2)/\nu\right)^{-(\nu+2)/2}$ on $\mathbb{R}^2$. Built on this standard bivariate distribution, the basic idea is to generate in a simple way a sample of data which is neither concentrated on the positive quadrant nor isotropic, and whose marginal tail behavior is straightforward to evaluate. The resulting Monte-Carlo estimates for $\nu \in \{3,5\}$, displayed in Figure 6, indicate that $\widetilde{XMES}^\star(\widehat{\tau}'_n(\alpha_n))$ is more efficient relative to $\widehat{XMES}^\star(\widehat{\tau}'_n(\alpha_n))$. This superiority of the direct estimator is, however, no longer valid in the case of extremely heavy tails such as, for instance, $\nu = 2$ and the transformed Cauchy distribution considered in Cai *et al.* (2015).

7 Applications

In this section, we apply our estimation methods to first estimate the tail VaR 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.

7.1 VaR 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 and scatterplot shown in Figure 7 (a) give evidence of an important right-skewness. Insurance companies are then interested in
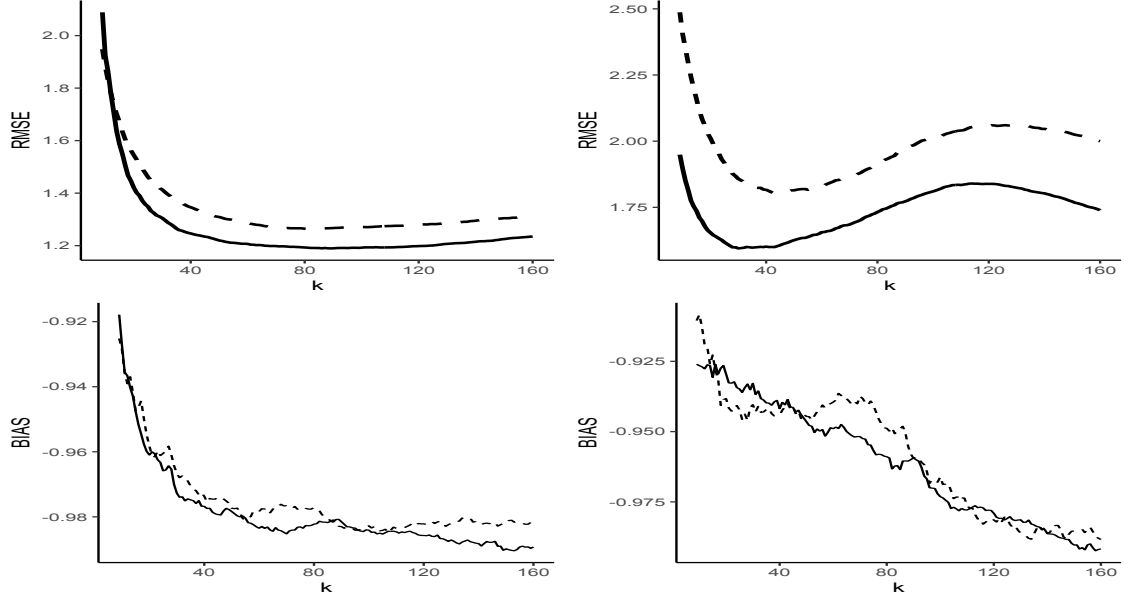
Figure 6: *As before—Results for the transformed t_3 and t_5-distributions on $\mathbb{R}^2$.*

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 $\widehat{q}^{\star}_{\alpha_n} = Y_{n-k,n} \left(\frac{k}{np_n} \right)^{\widehat{\gamma}_H}$ as described in (10), where $\widehat{\gamma}_H$ is the Hill estimator defined in (6), with $\alpha_n = 1 - p_n$ and $\tau_n = 1 - \frac{k}{n}$. According to the earlier study of Beirlant *et al.* (2004, p.123), insurers typically are interested in $p_n = \frac{1}{100,000} \approx \frac{1}{n}$ for these medical insurance data, that is, in an estimate of the claim amount that will be exceeded (on average) only once in 100,000 cases. Similar recent studies in the context of the backtesting problem, which is crucial in the current Basel III regulatory framework, are Chavez-Demoulin *et al.* (2014) and Gong *et al.* (2015), who estimate quantiles exceeded on average once every 100 cases with sample sizes of the order of hundreds. Figure 7 (b) shows the quantile-VaR estimates $\widehat{q}^{\star}_{\alpha_n}$ against the sample fraction k (solid line). 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 exceed the sample maximum $Y_{n,n} = 4,518,420$ (indicated by the horizontal line), which is consistent with the earlier analysis of Beirlant *et al.* (2004, p.125 and p.159).

The alternative expectile-based estimator $\widetilde{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}$ introduced in Section 5, which estimates the same VaR $q_{\alpha_n} \equiv \xi_{\tau'_n(\alpha_n)}$ as the quantile-based estimator $\widehat{q}^{\star}_{\alpha_n} \equiv \widehat{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}$, is also graphed in Figure 7 (b) in dashed line. As an asymmetric-least-squares estimator, it is more affected by the infrequent great claim amounts visualized in the top figure. Its plot indicates a more conservative risk measure between 3.92 and 4.33 million, over the stable region $k \in [150, 500]$.
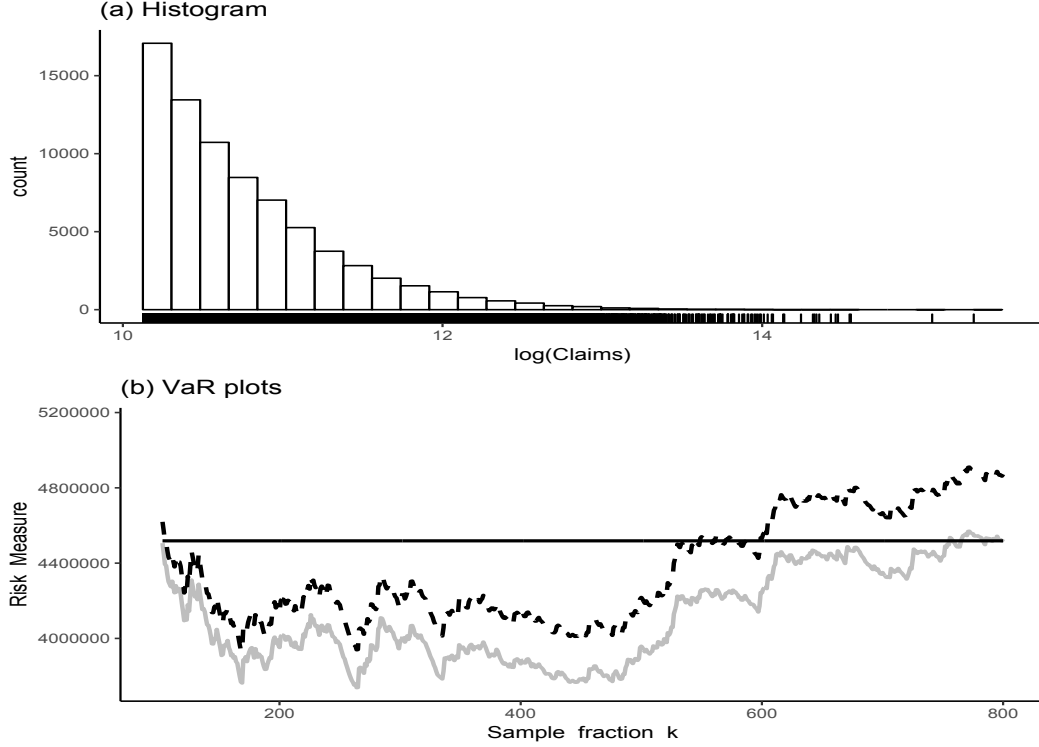
24

(a) Histogram

(b) VaR plots

Figure 7: *SOA Group Medical Insurance data. (a) Histogram and scatterplot of the log-claim amounts. (b) The VaR plots $\{(k, \widetilde{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}(k))\}_k$ in dashed line and $\{(k, \widehat{q}^{\star}_{\alpha_n}(k))\}_k$ in solid line, along with the sample maximum $Y_{n,n}$ in horizontal line.*

7.2 MES of three large US financial institutions

We consider the same investment banks as in the studies of Cai *et al.* (2015) and Brownlees and Engle (2017), namely Goldman Sachs, Morgan Stanley and T. Rowe Price. For the three banks, the dataset consists of the loss returns, *i.e.*, the negative log-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 $\widehat{\mathrm{QMES}}^{\star}(\alpha_n)$, as defined in (21), to estimate the quantile-based MES, $\mathrm{QMES}(\alpha_n) = \mathbb{E}\{X | Y > q_{Y,\alpha_n}\}$, where $\alpha_n = 1 - \frac{1}{n} = 1 - \frac{1}{2513}$, 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 $\mathrm{QMES}(\alpha_n) \equiv \mathrm{XMES}(\tau'_n(\alpha_n)) = \mathbb{E}\{X | Y > \xi_{Y,\tau'_n(\alpha_n)}\}$, with the same extreme relative frequency α_n that corresponds to a once-per-decade systemic event. We employ the rival estimator $\widehat{\mathrm{QMES}}^{\star}(\alpha_n)$ 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 (dashed line) in Figure 8 (a). Indeed, the first stable region appears for $k \in [70, 100]$ with an averaged estimate $\widehat{\gamma}_Y = 0.35$. Hence, by Proposition 2, the estimates $\widehat{\text{XMES}}^\star(\alpha_n)$ and $\widetilde{\text{XMES}}^\star(\alpha_n)$ are expected to be less extreme than the benchmark values $\widehat{\text{QMES}}^\star(\alpha_n)$. This is visualised in Figure 15 in the supplement to this article, where the three estimates are graphed as functions of k for each bank. As a matter of fact, both $\widehat{\text{XMES}}^\star(\alpha_n)$ and $\widetilde{\text{XMES}}^\star(\alpha_n)$ estimate the less extreme risk measure $\text{XMES}(\alpha_n)$ and not the desired intuitive tail measure $\text{XMES}(\tau'_n(\alpha_n)) \equiv \text{QMES}(\alpha_n)$.

The interest here is rather on the composite estimators $\widehat{\text{XMES}}^\star(\widehat{\tau}'_n(\alpha_n))$ and $\widetilde{\text{XMES}}^\star(\widehat{\tau}'_n(\alpha_n))$, where $\widehat{\text{XMES}}^\star(\widehat{\tau}'_n(\alpha_n))$ is actually nothing but $\widehat{\text{QMES}}^\star(\alpha_n)$. The two rival estimates $\widehat{\text{QMES}}^\star(\alpha_n)$ and $\widetilde{\text{XMES}}^\star(\widehat{\tau}'_n(\alpha_n))$ represent the average daily loss return for a once-per-decade market crisis. They are graphed in Figure 8 (b)-(d) as functions of k for each bank: (b) Goldman Sachs; (c) Morgan Stanley; (d) T. Rowe Price. The first stable regions of the plots (b)-(d) appear, respectively, for $k \in [80, 105]$, $k \in [90, 140]$ and $k \in [75, 100]$. The final estimates based on averaging the estimates from these stable regions are reported in the left-hand side of Table 1, along with the asymptotic 95% confidence intervals derived from Theorem 7 with the bias condition $\lambda_2 = 0$ (the asymptotic distribution then being $\mathcal{N}(0, \gamma_X^2)$ due to the use of the Hill estimator of γ_X, see the discussion below Theorem 1). It may be seen that both expectile- and quantile-based MES levels for Goldman Sachs and T. Rowe Price are almost equal. However, the 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 $\widehat{\text{QMES}}^\star(\alpha_n)$, 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. Also, these quantile-based estimates appear to be less conservative than our asymmetric least squares-based estimates, but not by much: this minor difference can already be visualized in Figure 8 (b)-(d), where the plots of $\widehat{\text{QMES}}^\star(\alpha_n)$, in dashed line, and $\widetilde{\text{XMES}}^\star(\widehat{\tau}'_n(\alpha_n))$, in solid line, exhibit a very similar evolution for the three banks.

	Daily loss		Weekly loss	
Bank	$\widetilde{\text{XMES}}^\star(\widehat{\tau}'_n(\alpha_n))$	$\widehat{\text{QMES}}^\star(\alpha_n)$	$\widetilde{\text{XMES}}^\star(\widehat{\tau}'_n(\alpha_n))$	$\widehat{\text{QMES}}^\star(\alpha_n)$
Goldman Sachs	0.3123 (0.20,0.42)	0.3077 (0.19,0.41)	0.3423 (0.17,0.51)	0.3375 (0.16,0.50)
Morgan Stanley	0.5622 (0.34,0.78)	0.5552 (0.33,0.77)	0.6495 (0.26,1.03)	0.6641 (0.26,1.05)
T. Rowe Price	0.3308 (0.25,0.52)	0.3098 (0.23,0.50)	0.3407 (0.19,0.48)	0.3405 (0.19,0.48)

Table 1: *Expectile- and quantile-based MES of the three investment banks. The second and third columns report the results based on daily loss returns ($n = 2513$ and $\alpha_n = 1 - \frac{1}{n}$). The last two columns report the results based on weekly loss returns from the same sample period ($n = 522$ and $\alpha_n = 1 - \frac{1}{n}$). Each MES estimate is followed by the 95% confidence interval.*

In our theoretical results we do not enter into the important question of serial dependence.

We only consider independent and identically distributed random vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$. One way to reduce substantially the potential serial dependence in this application is by using lower frequency data. As suggested by Cai *et al.* (2015), we choose weekly loss returns in the same sample period. This results in a sample of size $n = 522$. The estimates of γ_Y and $\mathrm{QMES}(\alpha_n) \equiv \mathrm{XMES}(\tau'_n(\alpha_n))$, with $\alpha_n = 1 - \frac{1}{n}$, are displayed in Figure 16 of the supplement to this article as functions of k. The averaged estimate $\widehat{\gamma}_Y = 0.37$ is obtained from the first stable region $k \in [25, 35]$ of the plot (a). The first stable regions of the plots (b)-(d) appear, respectively, for $k \in [27, 36]$, $k \in [23, 33]$ and $k \in [25, 33]$. The final results based on averaging the estimates from these stable regions are reported in the right-hand side of Table 1. They are very similar to those obtained in Cai *et al.* (2015) by resorting to two intermediate sequences. Both expectile- and quantile-based MES estimates are qualitatively robust to the change from daily to weekly data: they are still almost equal for Goldman Sachs and T. Rowe Price, while almost twice higher for Morgan Stanley.

There remains a lot to be done, especially on the extension of our expectile-based methods to a time dynamic setting. Already, Taylor (2008) and Kuan *et al.* (2009) have initiated the use of expectiles to estimate VaR and ES in conditional autoregressive expectile models. The use of expectiles to estimate MES may also work by allowing for dynamics in the covariance matrix via a multivariate GARCH model, similarly to the quantile-based method of Brownlees and Engle (2017). From the perspective of extreme values, one way to deal with the heteroskedasticity present in series of financial returns, similarly to Diebold *et al.* (2000), McNeil and Frey (2000) and McNeil *et al.* (2005, p. 283), is by applying our method to residuals standardized by GARCH conditional volatility estimates. Also, similarly to extreme value analysis under mixing conditions in a univariate setting (see *e.g.* Drees, 2000), our theorems may work under serial dependence with enlarged asymptotic variances.

Acknowledgements

Supplementary material

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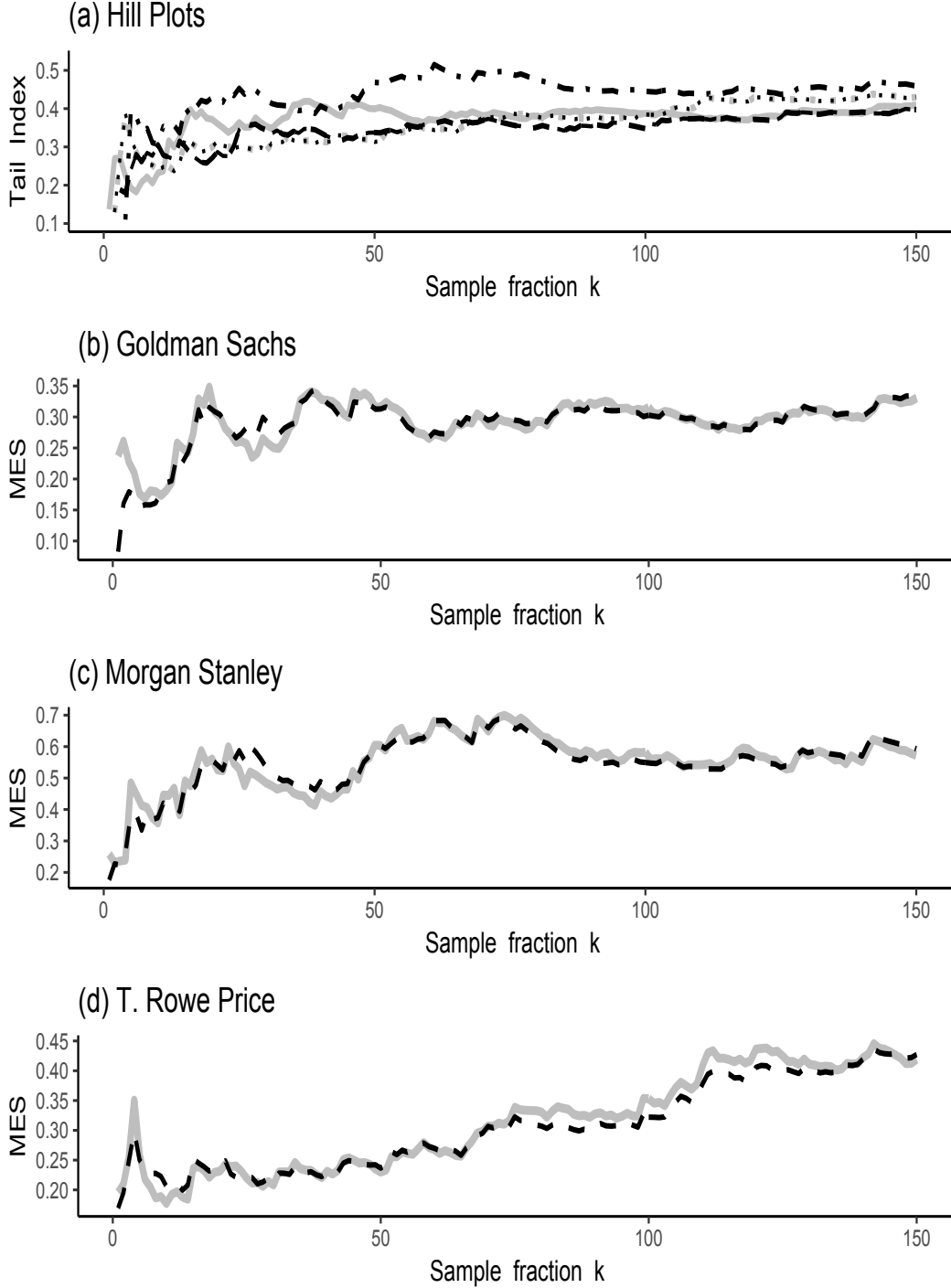
(a) Hill Plots

(b) Goldman Sachs

(c) Morgan Stanley

(d) T. Rowe Price

Figure 8: *(a) Hill estimates $\widehat{\gamma}_Y$ based on daily loss returns of market index (dashed), along with $\widehat{\gamma}_X$ based on daily loss returns of Goldman Sachs (solid), Morgan Stanley (dashed-dotted), and T. Rowe Price (dotted). (b)-(d) The estimates $\widehat{QMES}^{\star}(\alpha_n)$ in dashed line and $\widetilde{XMES}^{\star}(\widehat{\tau}'_n(\alpha_n))$ in solid line for the three banks, with $n = 2513$ and $\alpha_n = 1 - 1/n$.*

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] Acharya, V.V., Pedersen, L.H., Philippon, T. and Richardson, M. (2012). Measuring systemic risk, *Discussion Paper DP8824*, Centre for Economic Policy Research, London.

[3] Alm, J. (2016). Signs of dependence and heavy tails in non-life insurance data, *Scandinavian Actuarial Journal*, **10**, 859–875.

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

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

[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] Brownlees, C.T. and Engle, R. (2017). SRISK: A Conditional Capital Shortfall Measure of Systemic Risk, *The Review of Financial Studies*, **30**, 48–79.

[11] Caeiro, F., Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator, *REVSTAT: Statistical Journal*, **3**, 113–136.

[12] Cai, J.J., Einmahl, J.H.J., de Haan, L. and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme, *Journal of the Royal Statistical Society: Series B*, **77**, 417–442.

[13] Chavez-Demoulin, V., Embrechts, P. and Sardy, S. (2014). Extreme-quantile tracking for financial time series, *Journal of Econometrics*, **181**, 44–52.

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

[15] Diebold, F.X., Schuermann, T. and Stroughair, J.D. (2000). Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management, *Journal of Risk Finance*, **1** (Winter), 30–36.

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

[17] 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.

[18] Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. (2016). Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations, and Forecast Rankings, *Journal of the Royal Statistical Society: Series B*, **78**, 505–562.

[19] 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.

[20] El Methni, J. and Stupfler, G. (2017a). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions, *Statistica Sinica*, **27**(2), 907–930.

[21] El Methni, J. and Stupfler, G. (2017b). Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions, *Econometrics and Statistics* (to appear).

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

[23] Embrechts, P. and Puccetti, G. (2007). Aggregating risk across matrix structured loss data: the case of operational risk, *Journal of Operational Risk*, **3**, 29–44.

[24] Engle, R.F., Jondeau, E. and Rockinger, M. (2015). Systemic Risk in Europe, *Review of Finance*, **19**, 145–190.

[25] Fissler, T. and Ziegel, J.F. (2016). Higher order elicitability and Osband's principle, *Annals of Statistics*, **4**, 1680–1707.

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

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

[28] Gong, J., Li, Y., Peng, L. and Yao, Q. (2015). Estimation of extreme quantiles for functions of dependent random variables, *Journal of the Royal Statistical Society: Series B*, **77**, 1001–1024.

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

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

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

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

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

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

[35] 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.

[36] 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.

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

[38] Martin, R. (2014). Expectiles behave as expected, *Risk Magazine*, **27** (6), 79–83.

[39] McNeil, A.J. and Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach, *Journal of Empirical Finance*, **7**, 271–300.

[40] McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management,* New Jersey: Princeton University Press.

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

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

[43] Schulze Waltrup, L., Sobotka, F., Kneib, T. and Kauermann, G. (2015). Expectile and Quantile Regression - David and Goliath?, *Statistical Modelling*, **15**, 433–456.

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

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

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

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