Estimation of Tail Risk based on Extreme Expectiles
Abdelaati Daouia, Stephane Girard, Gilles Stupfler

To cite this version:

HAL Id: hal-01142130
https://hal.archives-ouvertes.fr/hal-01142130v4
Submitted on 17 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract

We use tail expectiles to estimate alternative measures to the 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 large expectile-based VaR, ES and MES when they are covered by the range of the data, 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.

Keywords: 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 $\mu(Y)$, 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 \( \tau \), as the minimizer

\[
\xi_\tau = \arg\min_{\theta \in \Theta} \mathbb{E} \{ \eta_r(Y - \theta) - \eta_r(Y) \},
\]

where \( \eta_r(y) = |\tau - \mathbb{I}(y \leq 0)| y^2 \), with \( \mathbb{I}(\cdot) \) 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 \( \eta_r(Y) \). 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) and Sobotka and Kneib (2012), among others, is that expectiles make more efficient use of the available data since 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. Also, according to Kuan et al. (2009), while quantiles only depend on the frequency of tail realizations of \( Y \) and not on their values, expectiles depend by construction on both the tail realizations and their probability. This benefit in terms of more valuable tail information comes at the price of increased sensitivity to the magnitude of extremes and/or outliers, see Sobotka and Kneib (2012). 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 adequate 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 expectile regression is closer to the notion of explained variance in least squares estimation and expectile-specific parameters can be interpreted with respect to variance heteroscedasticity (see Sobotka and Kneib (2012, p.756)). Furthermore, sample expectiles provide a class of smooth curves as functions of the level \( \tau \), which is not the case for sample quantiles (see, e.g., Schulze Waltrup et al. (2015)). Most importantly, inference on expectiles is much easier than inference on quantiles (Abdous and Remillard, 1995). 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 (see also Newey and Powell, 1987). In terms of interpretability, the \( \tau \)-quantile determines the point below which 100\( \tau \)% of the mass of \( Y \) lies, while the \( \tau \)-expectile specifies the position \( \xi_\tau \) such that the average distance from the
data below $\xi_\tau$ to $\xi_\tau$ itself is $100\tau\%$ of the average distance between $\xi_\tau$ and all the data, i.e.,

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

(2)

Thus, as pointed out by Fan and Gijbels (1996, p.231), the $\tau$-expectile shares an intuitive interpretation similar to the $\tau$-quantile, replacing the number of observations by the distance. This reduced probabilistic interpretability of expectiles should not be viewed as a serious disadvantage however, since Efron (1991) already suggested an elegant device to recover quantiles and their strong intuitive appeal from a set of expectiles. Koenker (1993) derived a first example of a distribution where the $\tau$th expectile and quantile coincide for all $\tau \in (0,1)$. Even more strongly, 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 $\tau$. Very recently, Zou (2014) has extended the Koenker’s argument to a more general 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 $\tau > \frac{1}{2}$ 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 (2016) among others. It is generally accepted that elicitability is a desirable property for model selection, computational efficiency, forecasting and testing algorithms. Actually, expectiles define the only coherent risk measure that is also elicitable. It has been shown that ES, the most popular coherent risk measure, is not elicitable (Gneiting, 2011), but jointly elicitable with VaR (Fissler and Ziegel, 2016).

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. (2016) 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 Taylor (2008), 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, as well as the two recent contributions by Holzmann and Klar (2016) and Krätschmer and Zähle (2016) for advanced theoretical developments. Attention has been, however, restricted to ordinary expectiles of fixed order \( \tau \) 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 \( \tau \) 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 \( \tau = \tau_n \rightarrow 0 \) or \( \tau_n \rightarrow 1 \) 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 \( \xi_{\tau_n} \) 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) have indeed 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 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 \( \tau_n \) at an extremely high level that can be even larger than \((1 - 1/n)\), see for instance Steenbergen et al. (2004) in the context of flood risk assessment, 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, El Methni and Stupfler (2016) who estimate several risk measures including excess-of-loss risk measures on automobile insurance data and de Valk (2016) for an application to oceanographic data. 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), and de Haan and Ferreira (2006) for a general overview of the theoretical background.

Let us point out some conceptual results of this paper. We first estimate the intermediate tail expectiles of order \( \tau_n \rightarrow 1 \) such that \( n(1 - \tau_n) \rightarrow \infty \), and then extrapolate these estimates to the very extreme expectile level \( \tau_n \) which approaches one at an arbitrarily fast rate in the sense that \( n(1 - \tau_n) \rightarrow c \), 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.

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 \( \tau \rightarrow 1 \). 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 7 and concrete applications to medical insurance data and three large US investment banks are provided in Section 8.
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. 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 $\xi_\tau$ 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 presence of terms $Y^2_+$ and $Y^2_-$ makes indeed this problem well-defined as soon as $Y \in L^1$ [i.e. $\mathbb{E} |Y| < \infty$]. The first-order necessary condition for optimality related to this problem can be written in several ways, one of them being

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

These equations have 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. 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 collection of elementary properties [Newey and Powell (1987), Abdous and Remillard (1995) and Bellini et al. (2014)]:

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

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

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

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

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

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


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

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

$$d_W(Y, \tilde{Y}) = \int_{-\infty}^{\infty} |F_Y(y) - F_{\tilde{Y}}(y)| dy = \int_{0}^{1} |F_Y^{-1}(t) - F_{\tilde{Y}}^{-1}(t)| dt.$$

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

The sign convention we have chosen for values of $Y$ as the negative of returns implies that extreme losses correspond to levels $\tau$ 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 $\xi_\tau$ and how they are linked to extreme quantiles $q_\tau$ 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

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

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 tail index $\gamma$ 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: in addition to the monographs of Embrechts et al. (1997) and Resnick (2007), see for instance Chavez-Demoulin et al. (2015) and the references therein. See also Alm (2015) for a recent study in the context of the Swedish insurance market.

Writing $F_Y := 1 - F_Y$, Bellini et al. (2014) have shown in the case $\gamma < 1$ that

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

or equivalently $\frac{F_Y(\xi_\tau)}{1-\tau} \sim \gamma^{-1} - 1$ as $\tau \to 1$. It follows that extreme expectiles $\xi_\tau$ are larger than extreme quantiles $q_\tau$ (i.e. $\xi_\tau > q_\tau$) when $\gamma > \frac{1}{2}$, whereas $\xi_\tau < q_\tau$ when $\gamma < \frac{1}{2}$, for all large $\tau$. 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 $1/F_Y$, defined by
\[ \forall t > 1, \ U(t) = \inf \left\{ y \in \mathbb{R} \ \left| \frac{1}{F_Y(y)} \geq t \right. \right\}, \]
satisfies the second-order condition indexed by $(\gamma, \rho, A)$, that is, there exist $\gamma > 0$, $\rho \leq 0$, and a function $A(\cdot)$ converging to 0 at infinity and having constant sign such that
\[ C_2(\gamma, \rho, A) \] for all $x > 0,$
\[ \lim_{t \to \infty} \frac{1}{A(t)} \left[ \frac{U(tx)}{U(t)} - x^\gamma \right] = x^\gamma x^\rho - 1 \rho. \]

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 Beirlant et al. (2004) and de Haan and Ferreira (2006) along with abundant examples of commonly used continuous distributions satisfying $C_2(\gamma, \rho, A)$: for instance, the (Generalized) Pareto, Burr, Fréchet, Student, Fisher and Inverse-Gamma distributions all satisfy this condition, and more generally so does any distribution whose distribution function $F$ satisfies
\[ 1 - F(x) = x^{-1/\gamma} \left( a + bx^{\rho/\gamma} + o(x^{\rho/\gamma}) \right) \ \text{as} \ x \to \infty, \]
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 $C_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1$. Then
\[ \frac{F_Y(\xi)}{1 - \tau} = (\gamma^{-1} - 1)(1 + \varepsilon(\tau)) \]
with $\varepsilon(\tau) = -\frac{(\gamma^{-1} - 1)^{\gamma} E(Y)}{q_\tau} (1 + o(1)) - \frac{(\gamma^{-1} - 1)^{-\rho}}{\gamma(1 - \rho - \gamma)} A((1 - \tau)^{-1})(1 + o(1))$ as $\tau \uparrow 1$.

Even more strongly, one can establish the precise bias term in the asymptotic approximation of $(\xi/\tau)$ itself.

**Corollary 1.** Assume that condition $C_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1$. If $F_Y$ is strictly increasing, then
\[ \frac{\xi}{q_\tau} = (\gamma^{-1} - 1)^{-\gamma}(1 + r(\tau)) \]
with $r(\tau) = \frac{\gamma(\gamma^{-1} - 1)^{\gamma} E(Y)}{q_\tau} (1 + o(1))$
\[ + \left( \frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \rho - \gamma} + \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho} + o(1) \right) A((1 - \tau)^{-1}) \ \text{as} \ \tau \uparrow 1. \]
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 (2015). In practice, the tail quantities $\xi_\tau$, $q_\tau$ and $\gamma$ 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 $\gamma$ are used widely in applied work and investigated extensively in theoretical statistics, the problem of estimating $\xi_\tau$, 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 $\xi_{\tau_n}$ 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 $C_2(\gamma, \rho, A)$ to obtain some convergence results.

3 Estimation of the expectile-based VaR

Our main objective in this section is to estimate $\xi_{\tau_n}$ for high levels $\tau_n$ 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 $Y_{1,n} \leq \cdots \leq Y_{n,n}$ 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 $F_Y$ is regularly varying at infinity with index $-1/\gamma$ [i.e. it satisfies, for any $x > 0$, the property $F_Y(tx)/F_Y(t) \to x^{-1/\gamma}$ as $t \to \infty$], it follows that $U$ is regularly varying as well with index $\gamma$. Hence, (4) entails that

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

This result is also an immediate consequence of Corollary 1 above and can be found in Proposition 2.3 of Bellini and Di Bernardino (2015) as well. Therefore, for a suitable estimator $\hat{\gamma}$ of $\gamma$, we may suggest estimating the intermediate expectile $\xi_{\tau_n}$ by

$$\hat{\xi}_{\tau_n} := (\hat{\gamma}^{-1} - 1)^{-\hat{\gamma}} \hat{q}_{\tau_n}, \quad \text{where} \quad \hat{q}_{\tau_n} := Y_{n - \lfloor n(1 - \tau_n) \rfloor, n}.$$
Theorem 1. Assume that \( F_Y \) is strictly increasing, that condition \( C_2(\gamma, \rho, A) \) holds with \( 0 < \gamma < 1 \), that \( \tau_n \uparrow 1 \) and \( n(1 - \tau_n) \to \infty \). Assume further that

\[
\sqrt{n(1 - \tau_n)} \left( \frac{\gamma - \hat{\gamma}_n}{\hat{q}_n} - 1 \right) \xrightarrow{d} (\Gamma, \Theta). \tag{7}
\]

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

\[
\sqrt{n(1 - \tau_n)} \left( \frac{\hat{\xi}_n}{\bar{q}_n} - 1 \right) \xrightarrow{d} m(\gamma)\Gamma + \Theta - \lambda
\]

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

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

When using the Hill estimator (6) of \( \gamma \) 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 \( \Gamma \) is then Gaussian with mean \( \lambda_2/(1 - \rho) \) and variance \( \gamma^2 \), while \( \Theta \) 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 \( C_2(\gamma, \rho, A) \) with \( 0 < \gamma < 1 \) and \( \tau_n \to 1 \) is such that \( n(1 - \tau_n) \to \infty \),

\[
\sqrt{n(1 - \tau_n)} \left( \frac{\hat{\xi}_n}{\bar{q}_n} - 1 \right) \xrightarrow{d} \mathcal{N} \left( \frac{m(\gamma)}{1 - \rho} \lambda_2 - \lambda, v(\gamma) \right),
\]

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

\[
v(\gamma) = 1 + \left( \frac{\gamma}{1 - \gamma} - \gamma \log \left( \frac{1}{\gamma} - 1 \right) \right)^2.
\]
Yet, a drawback to the resulting estimator $\hat{\xi}_{\tau_n}$ lies in its heavy dependency on the estimated quantile $\hat{q}_{\tau_n}$ and tail index $\hat{\gamma}$ in the sense that $\hat{\xi}_{\tau_n}$ may inherit the vexing defects of both $\hat{q}_{\tau_n}$ and $\hat{\gamma}$. Note also that $\hat{\xi}_{\tau_n}$ is asymptotically biased, which is not the case for $\hat{q}_{\tau_n}$; it should be pointed out though that one may design a bias-reduced version of the estimator $\hat{\xi}_{\tau_n}$. Indeed, the bias components $\lambda_1$ and $\lambda_2$ appearing in Theorem 1 can be estimated, respectively, by using $\hat{\lambda}_1 = \sqrt{n(1 - \tau_n)} \hat{\gamma}^{-1}$ and by applying the methodology of Caeiro et al. (2005) in conjunction with the Hall-Welsh class of models to get an estimator $\hat{\lambda}_2$ of $\lambda_2$. Along with the empirical mean $\bar{Y}$, the estimator $\hat{\gamma}$, and a consistent estimator $\hat{\rho}$ of the second-order parameter $\rho$ (a review of possible estimators $\hat{\rho}$ is given in Gomes and Guillou, 2015), it is possible to come up with a consistent estimator $\hat{\lambda}$ as suggested by Caeiro et al. (2005). This in turn enables one to define a bias-reduced version of $\hat{\xi}_{\tau_n}$, for instance, as

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

Of course, one should expect the value of the asymptotic variance of this estimator to be even higher than that of $\hat{\xi}_{\tau_n}$, as 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 $\xi_{\tau_n}$, which we develop in the next section, 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 $\xi_{\tau_n}$ by its empirical counterpart defined through

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

where $\eta_{\tau}(y) = |\tau - I(y \leq 0)|y^2$ 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

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

with

$$
\psi_n(u): = \frac{1}{2\sigma_{\tau_n}^2} \sum_{i=1}^{n} \left[ \eta_{\tau_n}(Y_i - \xi_{\tau_n} - u\xi_{\tau_n}/\sqrt{n(1 - \tau_n)}) - \eta_{\tau_n}(Y_i - \xi_{\tau_n}) \right].
$$
It follows from the continuity and the convexity of $\eta$, that $(\psi_n)$ 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 $(\psi_n)$. Built on this idea, we get the asymptotic normality of the LAWS estimator $\tilde{\xi}_{\tau_n}$ by applying standard techniques involving sums of independent and identically distributed random variables. We will denote in the sequel by $Y_-$ the negative part of $Y$, i.e., $Y_- = Y - \max(Y, 0)$.

**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

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

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 $C_2(\gamma, \rho, A)$ or any bias condition. A mild moment assumption and the condition $0 < \gamma < 1/2$ suffice. It has been found in previous studies by many authors, too many to review here, that the model assumption of Pareto-type tails along with these conditions (required for losses to have at least a finite variance) deliver competitive results for most applicational purposes in risk management. In these studies the realized values of $\gamma$ were found well below 1/2, being in line with the recent findings of Cai et al. (2015) and our findings in Section 8. We refer the reader to the R package ‘CASdatasets’ which contains a large variety of dataset examples where realized values of $\gamma$ often vary between 1/4 and 1/2. Most importantly, unlike the indirect expectile estimator $\hat{\xi}_{\tau_n}$, the new estimator $\tilde{\xi}_{\tau_n}$ does not hinge by construction on any particular type of quantile or tail index estimators. A comparison of the asymptotic variance $V(\gamma)$ of $\tilde{\xi}_{\tau_n}$ with the asymptotic variance $v(\gamma)$ of $\hat{\xi}_{\tau_n}$ 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, $V(\gamma)$ is increasing and 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 estimates of order $\tau_n \to 1$, such that $n(1 - \tau_n) \to \infty$, to the very extreme level $\tau'_n$. 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 $\tau'_n$ and $\tau_n$ are typically set to be $\tau'_n = 1 - p_n$ for a $p_n$ much smaller than $1/n$, and $\tau_n = 1 - \frac{k(n)}{n}$ for an intermediate sequence of integers $k(n).$
We then show that the indirect estimator

\[ \xi_{\tau_n} = \frac{U((1-\tau_n')^{-1})}{U((1-\tau_n)^{-1})} \approx \left( \frac{1-\tau_n'}{1-\tau_n} \right)^{-\gamma} \]

and thus

\[ \frac{\xi_{\tau_n'}}{\xi_{\tau_n}} \approx \left( \frac{1-\tau_n'}{1-\tau_n} \right)^{-\gamma} \]

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

\[ \tilde{\xi}_{\tau_n}^* \equiv \tilde{\xi}_{\tau_n}^*(\tau_n) := \left( \frac{1-\tau_n'}{1-\tau_n} \right)^{-\hat{\gamma}} \xi_{\tau_n} \tag{9} \]

where \( \hat{\gamma} \) is an estimator of \( \gamma \), and \( \tilde{\xi}_{\tau_n} \) stands for either the estimator \( \hat{\xi}_{\tau_n} \) or \( \tilde{\xi}_{\tau_n} \) of the intermediate expectile \( \xi_{\tau_n} \). As a matter of fact, we have \( \frac{\tilde{\xi}_{\tau_n}^*}{\xi_{\tau_n}} = \frac{\hat{\xi}_{\tau_n}^*}{\hat{\xi}_{\tau_n}} \) where \( \hat{\xi}_{\tau_n} = Y_n - \lfloor n(1-\tau_n) \rfloor n \) is the intermediate quantile estimator introduced above, and \( \hat{\xi}_{\tau_n}^* \) is the extreme Weissman quantile estimator defined as

\[ \hat{\xi}_{\tau_n}^* \equiv \hat{\xi}_{\tau_n}^*(\tau_n) := \left( \frac{1-\tau_n'}{1-\tau_n} \right)^{-\hat{\gamma}} \hat{\xi}_{\tau_n}. \tag{10} \]

We then show that \( \frac{\tilde{\xi}_{\tau_n} - 1}{\tilde{\xi}_{\tau_n}} \) has the same limit distribution as \( (\hat{\gamma} - \gamma) \) with a different scaling.

**Theorem 3.** Assume that \( F_Y \) is strictly increasing, that condition \( C_2(\gamma, \rho, A) \) holds with \( \rho < 0 \), that \( \tau_n, \tau_n' \uparrow 1 \), with \( n(1-\tau_n) \to \infty \) and \( n(1-\tau_n') \to c \), and moreover

\[ \sqrt{n(1-\tau_n)} \left( \frac{\tilde{\xi}_{\tau_n} - 1}{\xi_{\tau_n}} \right) \xrightarrow{d} \Delta \quad \text{and} \quad \sqrt{n(1-\tau_n)}(\hat{\gamma} - \gamma) \xrightarrow{d} \Gamma, \]

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

\[ \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \left( \frac{\tilde{\xi}_{\tau_n}^* - 1}{\xi_{\tau_n}} \right) \xrightarrow{d} \Gamma. \]
More specifically, we can choose $\tilde{\xi}_{\tau_n}$ in (9) to be either the indirect intermediate expectile estimator $\hat{\xi}_{\tau_n}$, the resulting extreme expectile estimator $\hat{\xi}_{\tau_n}^* := \tilde{\xi}_{\tau_n}$ being
\[
\hat{\xi}_{\tau_n}^* = \left( \frac{1 - \tau_n}{1 - \tau_n'} \right)^{-\gamma} \hat{\xi}_{\tau_n} = \left( \gamma^{-1} - 1 \right)^{-\gamma} \tilde{q}_{\tau_n'}^*.
\]
(11)
or we may choose $\xi_{\tau_n}$ to be the LAWS estimator $\tilde{\xi}_{\tau_n}$, yielding the extreme expectile estimator
\[
\tilde{\xi}_{\tau_n}^* = \left( \frac{1 - \tau_n}{1 - \tau_n'} \right)^{-\gamma} \tilde{\xi}_{\tau_n}.
\]
(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 $C_2(\gamma, \rho, \widehat{A})$ holds with $0 < \gamma < 1$ and $\rho < 0$, and that $\tau_n$, $\tau_n' \uparrow 1$ with $n(1 - \tau_n) \to \infty$ and $n(1 - \tau_n') \to c < \infty$. Assume further that
\[
\sqrt{n(1 - \tau_n)} \left( \hat{\gamma} - \gamma, \frac{\tilde{q}_{\tau_n}}{q_{\tau_n}} - 1 \right) \overset{d}{\to} (\Gamma, \Theta).
\]
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
\[
\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau_n')] \left( \frac{\hat{\xi}_{\tau_n}^*}{\xi_{\tau_n}^*} - 1 \right) \overset{d}{\to} \Gamma.
\]

**Corollary 4.** Assume that $F_Y$ is strictly increasing, there is $\delta > 0$ such that $\mathbb{E}|Y|^2 + \delta < \infty$, condition $C_2(\gamma, \rho, \widehat{A})$ holds with $0 < \gamma < 1/2$ and $\rho < 0$, and that $\tau_n$, $\tau_n' \uparrow 1$ with $n(1 - \tau_n) \to \infty$ and $n(1 - \tau_n') \to c < \infty$. If in addition
\[
\sqrt{n(1 - \tau_n)} (\hat{\gamma} - \gamma) \overset{d}{\to} \Gamma
\]
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
\[
\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau_n')] \left( \frac{\hat{\xi}_{\tau_n}^*}{\xi_{\tau_n}^*} - 1 \right) \overset{d}{\to} \Gamma.
\]

4 **Expectile-based expected shortfall**

The conventional quantile-based VaR was often criticized for being insensitive to the magnitude of extreme losses 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 method
known as Expected Shortfall (ES). This proposal was criticized though for its dependency only on the tail event. The formulation of the ES remains still intrinsically linked to quantiles as can be seen from (13) and (14) below. 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. Yet, 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. Already in October 2013, The Basel Committee on Banking Supervision proposed to change the traditional VaR with the ES for calculating losses. This motivates us to introduce a new variant of ES which is purely built on expectiles, inherits their coherency, but is more alert to extreme risks.

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 $\tau$. It is traditionally expressed at the $100(1 - \tau)\%$ security level as the Quantile-based Expected Shortfall (QES):

$$QES(\tau) := \frac{1}{1 - \tau} \int_{\tau}^{1} q_{\alpha} d\alpha. \quad (13)$$

When the financial position $Y$ is continuous, $QES(\tau)$ is just the conditional expectation

$$QTCE(\tau) := E[Y|Y > q_{\tau}] \quad (14)$$

of $Y$ given that it exceeds the VaR $q_{\tau}$. In this sense, it is referred to as Tail Conditional Expectation (TCE), with $-QES(\tau)$ 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

$$XES(\tau) := \frac{1}{1 - \tau} \int_{\tau}^{1} \xi_{\alpha} d\alpha. \quad (15)$$

We show below in Proposition 4 that $XES(\tau)$, in contrast to the expectile-based TCE introduced earlier by Taylor (2008) as

$$XTCE(\tau) := E[Y|Y > \xi_{\tau}], \quad (16)$$

defines a new coherent risk measure. We also show in Proposition 2, under the model assumption of Pareto-type distributions $F_{Y}(\cdot)$ with tail index $\gamma < 1$, that $XES(\tau)$ is asymptotically equivalent to $XTCE(\tau)$ as $\tau \to 1$, and hence inherits its direct meaning as a conditional expectation for all $\tau$ large enough. On the other hand, the choice between the expectile-based
ES/TCE and the quantile-based versions will depend on the value at hand of $\gamma \leq \frac{1}{2}$ as is the case in the duality between the expectile-based VaR and quantile-VaR. More precisely, the $\text{XES}(\tau)$ in (15) and $\text{XTCE}(\tau)$ in (16) are more extreme (respectively, less extreme) than their quantile-based analogues $\text{QES}(\tau)$ in (13) and $\text{QTCE}(\tau)$ in (14), for all $\tau$ 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

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

These connections are very useful when it comes to proposing estimators for $\text{XES}(\tau)$ and $\text{XTCE}(\tau)$. 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-based ES estimators.

**Proposition 3.** Assume that condition $C_2(\gamma, \rho, A)$ holds, with $0 < \gamma < 1$. Then, as $\tau \to 1$,

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

$$\left. + \frac{1-\gamma}{(1-\rho-\gamma)^2} (\gamma^{-1} - 1)^{-\rho} A((1-\tau)^{-1}) (1 + o(1)) \right),$$

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

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-based expected shortfall $\text{QES}(\tau)$ and the expectile-based VaR, we show that the expectile-based expected shortfall $\text{XES}(\tau)$ satisfies all of their requirements, namely Translation invariance, Monotonicity, Positive homogeneity, and Subadditivity. We also establish the first three requirements for the Tail Conditional Expectation variant $\text{XTCE}(\tau)$, as it is the case for the quantile-based version $\text{QTCE}(\tau)$. It should however be clear that, in contrast to $\text{QES}(\tau)$, the coherence of $\text{XES}(\tau)$ is actually a straightforward consequence of the coherence of the expectile-based VaR above the median in conjunction with the fact that the expectile-based ES is an increasing linear functional of the expectile-based VaR above some high level, in the sense that

$$\xi_{\alpha}^{(1)} \leq \xi_{\alpha}^{(2)} \quad \forall \alpha \in (\tau, 1) \quad \Rightarrow \quad \text{XES}^{(1)}(\tau) = \frac{1}{1-\tau} \int_{\tau}^{1} \xi_{\alpha}^{(1)} d\alpha \leq \frac{1}{1-\tau} \int_{\tau}^{1} \xi_{\alpha}^{(2)} d\alpha = \text{XES}^{(2)}(\tau).$$

**Proposition 4.** For all $\tau \geq 1/2$, the expectile-based expected shortfall $\text{XES}(\tau)$ induces a coherent risk measure, and $\text{XTCE}(\tau)$ is translation invariant, positive homogeneous and monotonic.
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

When analyzing the very far tails of the involved distribution, as required in modern regulatory frameworks (such as the European Union Solvency II directive), financial institutions and insurance companies are typically interested in the region $\tau = \tau_n \uparrow 1$, as the sample size $n \to \infty$. This is particularly required to manage extreme events. For example, Acharya et al. (2012) handle once-in-a-decade events with one year of data. Gilli and Kellezi (2006) apply extreme value theory to estimate risk measures for several stock market indices. Hartmann et al. (2004, 2005, 2010) employ extreme value techniques to evaluate systemic risk in the European and American banking systems. Also, Ibragimov et al. (2009) develop models for catastrophe insurance markets based on real data. The asymptotic equivalence $X_{ES}^{\cdot}(\tau_n) \sim (1 - \gamma)^{-1} \xi_{\tau_n}$, established in Proposition 2, suggests the following estimators of the expectile-based ES:

$$\hat{X}_{ES}^\cdot(\tau_n) = (1 - \hat{\gamma})^{-1} \cdot \hat{\xi}_{\tau_n}$$ \text{ and } $$\hat{X}_{ES}^{\cdot\cdot}(\tau_n) = (1 - \hat{\gamma})^{-1} \cdot \hat{\xi}_{\tau_n}^\cdot$$

where $\hat{\xi}_{\tau_n}$ and $\hat{\xi}_{\tau_n}^\cdot$ are the extreme expectile estimators defined above in (11)-(12), and $\hat{\gamma}$ is an estimator of $\gamma$. Another option motivated by the second asymptotic equivalence $X_{ES}^{\cdot}(\tau_n) \sim \frac{\xi_{\tau_n}^\cdot}{\hat{q}_{\tau_n}}$. QES$(\tau_n)$ would be to estimate $X_{ES}(\tau_n)$ by

$$\hat{X}_{ES}^{\cdot\cdot}(\tau_n) = \hat{\xi}_{\tau_n}^\cdot \cdot \frac{QES^{\cdot}(\tau_n)}{\hat{q}_{\tau_n}} \text{ or } \hat{X}_{ES}^{\cdot\cdot}(\tau_n) = \hat{\xi}_{\tau_n}^\cdot \cdot \frac{QES^{\cdot}(\tau_n)}{\hat{q}_{\tau_n}^\cdot}$$

for a suitable estimator $\hat{QES}^{\cdot}(\tau_n)$ of QES$(\tau_n)$ [see, e.g., El Methni et al. (2014)], with $\hat{q}_{\tau_n}^\cdot$ being the extreme Weissman quantile estimator defined in (10). Our experience with real and simulated data indicates, however, that the estimates $\hat{X}_{ES}^\cdot(\tau_n)$ and $\hat{X}_{ES}^{\cdot\cdot}(\tau_n)$ [respectively, $\hat{X}_{ES}^\cdot(\tau_n)$ and $\hat{X}_{ES}^{\cdot\cdot}(\tau_n)$] point toward very similar results. We therefore restrict our theoretical treatment to the initial versions given in (17). Our first asymptotic result is for the extreme XES estimator $\hat{X}_{ES}^\cdot(\tau_n)$:

**Corollary 5.** Under the conditions of Corollary 3,

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau_n)]} \left( \frac{\hat{X}_{ES}^\cdot(\tau_n)}{XES(\tau_n)} - 1 \right) \overset{d}{\to} \Gamma.$$

In what concerns the asymmetric least squares-type estimator $\hat{X}_{ES}^{\cdot\cdot}(\tau_n)$, we have the following result.
Corollary 6. Under the conditions of Corollary 4,

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

Both results are derived by noticing that, on the one hand, the extreme expectile estimators \( \widehat{\xi}_{\tau_n} \) and \( \widehat{\zeta}_{\tau_n} \) 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.

When it comes to estimate the Tail Conditional Expectation \( \text{XTCE}(\tau_n') \), the asymptotic equivalence \( \text{XTCE}(\tau_n') \sim \text{XES}(\tau_n') \) as \( n \to \infty \), obtained in Proposition 2, suggests to use the same estimators of \( \text{XES}(\tau_n') \) in (17) and (18) for \( \text{XTCE}(\tau_n') \) itself. To derive the limit distributions here, the basic arguments go as above.

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 (2016) and Engle et al. (2015) 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

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

where \( q_{Y,\tau} \) is the \( \tau \)th quantile of the distribution of \( Y \). Typically, a systemic crisis defined as an extreme tail event corresponds to a probability \( \tau \) 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 QMES(τ). 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 (2016) and Engle et al. (2015), but cannot handle extreme events required for systemic risk measures (i.e. 1 − τ = O(1/n)). Very recently, Cai et al. (2015) have proposed adapted extreme-value tools for the estimation of QMES(τ) without recourse to any parametric structure on (X, Y). Here, instead of the extreme τth quantile qY,τ, we will explore the use of the τth expectile analogue ξY,τ in the marginal expected shortfall

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

at least for the following respect: As claimed by Newey and Powell (1987), Kuan et al. (2009) and Sobotka and Kneib (2012) among others, expectiles make 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 XMES(τ) and QMES(τ) is provided below in Proposition 5. It is also 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 XMES(τ) 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 XMES(τ) 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 \( p_{X,Y} \). Here, the right-hand upper tail dependence between X and Y is described by the following joint convergence condition:

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

\[
\lim_{t \to \infty} t \mathbb{P}(F_X(X) \leq x/t, F_Y(Y) \leq y/t) := R(x, y)
\]

exists, with \( F_X = 1 - F_X \) and \( F_Y = 1 - F_Y \). The limit function \( R \) completely determines the so-called tail dependence function \( \ell \) [Drees and Huang (1998)] via the identity \( \ell(x, y) = x + y - R(x, y) \) for all \( x, y \geq 0 \) [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 \( 1/F_X \) and \( 1/F_Y \). 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 \( XMES(\tau) \).

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

\[
\lim_{\tau \uparrow 1} \frac{XMES(\tau)}{U_X(1/F_Y(\xi_{Y,\tau}))} = \int_0^\infty R(x^{-1/\gamma_X}, 1)dx, \quad (19)
\]

\[
\lim_{\tau \uparrow 1} \frac{XMES(\tau)}{QMES(\tau)} = (\gamma_Y^{-1} - 1)^{-\gamma_X}. \quad (20)
\]

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

\[
f_\nu(x, y) = \frac{2}{\pi} \left( 1 + \frac{x^2 + y^2}{\nu} \right)^{-(\nu+2)/2}, \quad x, y > 0, \quad (21)
\]

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

Figure 2: \( QMES(\tau) \) in solid line and \( XMES(\tau) \) in dashed line, as functions of \( \tau \in [0.95, 1] \). Case of Student \( t_\nu \)-distribution on \((0, \infty)^2\). From left to right, \( \nu = 3, 5, 7, 9 \).
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 (19). Subsequently we shall deal with a nonparametric estimator derived from the asymptotic connection (20) with the tail quantile-based MES.

5.3.1 Asymmetric least squares estimation

On the basis of the limit (19) and then of the heavy-tailedness assumption on $X$, we have for $\tau < \tau' < 1$ that, as $\tau \to 1$,

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

It follows then from Proposition 1 that

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

(22)

Hence, to estimate XMES($\tau'$) at an arbitrary extreme level $\tau' = \tau'_n$, we first consider the estimation of XMES($\tau$) 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

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

where $\tilde{\xi}_{Y,\tau_n}$ is the LAWS estimator of $\xi_{Y,\tau_n}$. As a matter of fact, in actuarial settings, we typically have a positive loss variable $X$, and hence $\mathbb{I}\{X_i > 0\} = 1$. 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 $\hat{\text{XMES}}(\tau_n)$ is $\sqrt{n(1 - \tau_n)}$-relatively consistent. By plugging this estimator into approximation (22) together with a $\sqrt{n(1 - \tau_n)}$-consistent estimator $\hat{\gamma}_X$ of $\gamma_X$, we obtain the following estimator of XMES($\tau'_n$):

$$\text{XMES}^*(\tau'_n) = \hat{\text{XMES}}^*(\tau'_n, \tau_n) := \left( \frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\hat{\gamma}_X} \hat{\text{XMES}}(\tau_n).$$

(23)

To determine the limit distribution of this estimator, we need to quantify the rate of convergence in condition $JC(R)$ as follows:
\( \mathcal{JC}_2(R, \beta, \kappa) \) Condition \( \mathcal{JC}(R) \) holds and there exist \( \beta > \gamma_X \) and \( \kappa < 0 \) such that

\[
\sup_{x \in [0, \infty)} \left| \frac{t \mathbb{P}(F_X(X) \leq x/t, F_Y(Y) \leq y/t) - R(x,y)}{\min(x^{\beta}, 1)} \right| = O(t^\kappa) \text{ \ as \ } t \to \infty.
\]

This is exactly condition (a) in Cai et al. (2015) under which an extrapolated estimator of \( \text{QMES}(\tau_n') \) 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 \( U_X \) satisfies the second-order condition \( C_2(\gamma_X, \rho_X, A_X) \) (resp. \( C_2(\gamma_Y, \rho_Y, A_Y) \)). The following generic theorem gives the asymptotic distribution of \( \text{XMES}^*(\tau_n') \). The asymptotic normality follows by using for example the Hill estimator \( \hat{\gamma}_X \) of the tail index \( \gamma_X \).

**Theorem 4.** Suppose that condition \( \mathcal{JC}_2(R, \beta, \kappa) \) holds, that there is \( \delta > 0 \) such that \( \mathbb{E}|Y_\cdot|^{2+\delta} < \infty \), and that \( U_X \) and \( U_Y \) satisfy conditions \( C_2(\gamma_X, \rho_X, A_X) \) and \( C_2(\gamma_Y, \rho_Y, A_Y) \) 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 \( \alpha < \min \left( \frac{-2\kappa}{-2\kappa + 1}, \frac{2\gamma_X \rho_X}{2\gamma_X \rho_X + \rho_X - 1} \right) \);

(iii) The bias conditions \( \sqrt{n(1 - \tau_n)} q_{Y_n, \tau_n}^{-1} \to \lambda_1 \in \mathbb{R}, \sqrt{n(1 - \tau_n)} A_X ((1 - \tau_n)^{-1}) \to \lambda_2 \in \mathbb{R} \) 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

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

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

(v) \( \mathbb{E}|X_-|^{1/\gamma_X} < \infty \); \hspace{1cm} (24)

(vi) \( n(1 - \tau_n) = o \left( (1 - \tau_n')^{2\kappa(1 - \gamma_X)} \right) \text{ \ as \ } n \to \infty. \hspace{1cm} (25) \)

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 \( C_2(\gamma_X, \rho_X, A_X) \). 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 \( X \).
5.3.2 Estimation based on tail QMES

On the basis of the limit (20), we consider the alternative estimator

$$\overline{\text{QMES}}(\tau'_n) := \left( \frac{1}{1-\tau'_n} \right)^{-\gamma_X} \overline{\text{QMES}}(\tau'_n),$$

(26)

where $\overline{\gamma}_X$, $\overline{\gamma}_Y$ and $\overline{\text{QMES}}(\tau'_n)$ are suitable estimators of $\gamma_X$, $\gamma_Y$ and $\text{QMES}(\tau'_n)$, respectively. Here, we use the Weissman-type device

$$\overline{\text{QMES}}(\tau'_n) = \left( \frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\gamma_X} \overline{\text{QMES}}(\tau_n)$$

(27)

of Cai et al. (2015) to estimate $\text{QMES}(\tau'_n)$, where

$$\overline{\text{QMES}}(\tau_n) = \frac{1}{\left[ n(1 - \tau_n) \right]} \sum_{i=1}^{n} X_i \mathbb{I}\{X_i > 0, Y_i > \tilde{q}_{Y,\tau_n}\},$$

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

**Theorem 5.** Suppose that condition $JC_2(R, \beta, \kappa)$ holds, and $U_X$ and $U_Y$ satisfy conditions $C_2(\gamma_X, \rho_X, A_X)$ and $C_2(\gamma_Y, \rho_Y, A_Y)$ with $\gamma_X \in (0, 1/2)$ 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 $\sqrt{n(1 - \tau_n)/(1 - \tau'_n)}) \to \infty$ as $n \to \infty$;

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

(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) $\sqrt{n(1 - \tau_n)}(\overline{\gamma}_X - \gamma_X) \overset{d}{\to} \Gamma$ and $\sqrt{n(1 - \tau_n)}(\overline{\gamma}_Y - \gamma_Y) = O_F(1)$.

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

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[\log(1 - \tau_n)/(1 - \tau'_n)]} \left( \frac{\overline{\text{QMES}}^*(\tau'_n)}{\overline{\text{QMES}}(\tau'_n)} - 1 \right) \overset{d}{\to} \Gamma.$$

This convergence remains still valid if $X \in \mathbb{R}$ provided that (24) and (25) hold.
6 Extreme expectile level selection

An important question that remains to be addressed is the choice of the extreme expectile level \( \tau_n \) in the three instruments of risk protection \( \xi_{\tau_n} \), \( \text{XES}(\tau_n) \) and \( \text{XMES}(\tau_n) \).

In the case of quantile-based risk measures \( q_{\alpha_n}, \text{QES}(\alpha_n) \) and \( \text{QMES}(\alpha_n) \), 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., Brownlees and Engle (2016) and Cai et al. (2015)). In the case of expectiles, we propose to select \( \tau_n \) so that each expectile-based risk measure has the same intuitive interpretation as its quantile-based analogue. This translates into choosing \( \tau_n \) such that \( \xi_{\tau_n} = q_{\alpha_n} \) for a given relative frequency \( \alpha_n \). Bellini and Di Bernardino (2015) have already suggested to pick out \( \tau_n \) which satisfies \( \xi_{\tau_n} = q_{\alpha_n} \), 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 \( \xi_{\tau_n} = q_{\alpha_n} \) that

\[
1 - \tau_n(\alpha_n) = \frac{\mathbb{E} \{|Y - q_{\alpha_n}| \mathbb{I}(Y > q_{\alpha_n})\}}{\mathbb{E}|Y - q_{\alpha_n}|}.
\]

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 \( \alpha_n \) and not on the quantile \( q_{\alpha_n} \) itself.

**Proposition 6.** Suppose \( F_Y \) satisfies (3) with \( 0 < \gamma < 1 \). Then

\[
1 - \tau_n(\alpha_n) \sim (1 - \alpha_n) \frac{\gamma}{1 - \gamma}, \quad n \to \infty.
\]

Hence, by substituting the estimated value

\[
\hat{\tau}_n(\alpha_n) = 1 - (1 - \alpha_n) \frac{\hat{\gamma}_n}{1 - \hat{\gamma}_n}
\]

in place of \( \tau_n(\alpha_n) \), both extreme expectile estimators \( \hat{\xi}_{\tau_n} \) in (11) and \( \hat{\xi}_{\tau_n} \) in (12) estimate the same Value at Risk \( \xi_{\tau_n(\alpha_n)} = q_{\alpha_n} \) as the Weissman quantile estimator \( \hat{q}_{\alpha_n} \) in (10). It is easily seen that the latter estimator is actually identical to the indirect expectile estimator \( \hat{\xi}_{\tau_n(\alpha_n)} \).

This quantile-based estimator \( \hat{q}_{\alpha_n} = \hat{\xi}_{\tau_n(\alpha_n)} \) may be criticized for being optimistic (or liberal) because it relies on a single order statistic \( \hat{q}_{\tau_n} = Y_{n-[n(1-\tau_n)]n} \), and hence may not respond properly to the very extreme losses. By contrast, the direct expectile-based estimator \( \hat{\xi}_{\tau_n(\alpha_n)} \) relies on the asymmetric least squares estimator \( \hat{\xi}_{\tau_n} \), 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
Assume further that 

(i) Under the conditions of Theorem 6 (i),

\[ \frac{\sqrt{n(1-\tau_n)}(\hat{\gamma} - \gamma)}{\log[(1-\tau_n)/(1-\alpha_n)]} \xrightarrow{d} (\Gamma, \Theta). \]

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 \( \sqrt{n(1-\tau_n)/\log[(1-\tau_n)/(1-\alpha_n)]]} \to \infty \), then

\[ \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]} \left( \frac{\hat{\xi}^*_{\tau_n}(\alpha_n)}{q_{\alpha_n}} - 1 \right) \xrightarrow{d} \Gamma. \]

(ii) Assume that \( F_Y \) is strictly increasing, that condition \( C_2(\gamma, \rho, A) \) holds with \( 0 < \gamma < 1/2 \) and \( \rho < 0 \), and that \( \tau_n, \alpha_n \uparrow 1 \) with \( n(1-\tau_n) \to \infty \) and \( n(1-\alpha_n) \to c < \infty \). If in addition

\[ \sqrt{n(1-\tau_n)}(\hat{\gamma} - \gamma) \xrightarrow{d} \Gamma \]

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 \( \sqrt{n(1-\tau_n)/\log[(1-\tau_n)/(1-\alpha_n)]]} \to \infty \), then

\[ \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]} \left( \frac{\tilde{\xi}^*_{\tau_n}(\alpha_n)}{q_{\alpha_n}} - 1 \right) \xrightarrow{d} \Gamma. \]

Likewise, \( \widetilde{\operatorname{XES}}^*(\tau'_n) \) and \( \overline{\operatorname{XES}}^*(\tau'_n) \) in (17) as well as \( \overline{\operatorname{XES}}^*(\tau'_n) \) and \( \overline{\operatorname{XES}}^*(\tau'_n) \) in (18), with \( \tau'_n = \tau'_n(\alpha_n) \), estimate the same expected shortfall \( \operatorname{XES}(\tau'_n(\alpha_n)) \equiv \overline{\operatorname{QES}}(\alpha_n) \) as the quantile-based estimator \( \widetilde{\operatorname{QES}}^*(\alpha_n) \) described in (28). Note also that \( \overline{\operatorname{XES}}^*(\tau'_n(\alpha_n)) \) coincides with \( \widetilde{\operatorname{QES}}^*(\alpha_n) \). Moreover, our numerical illustrations indicate that \( \widetilde{\operatorname{QES}}^*(\alpha_n) \) points towards similar estimates as \( \widetilde{\operatorname{XES}}^*(\tau'_n(\alpha_n)) \), but the direct expectile-based estimators \( \overline{\operatorname{XES}}^*(\tau'_n(\alpha_n)) \) and \( \overline{\operatorname{XES}}^*(\tau'_n(\alpha_n)) \) tend to be more conservative. The next result can be proved by making use of the proof of the previous theorem just as Corollaries 5 and 6 follow from the proofs of Corollaries 3 and 4.

**Theorem 7.** (i) Under the conditions of Theorem 6 (i),

\[ \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]} \left( \frac{\operatorname{XES}^*(\tau'_n(\alpha_n))}{\overline{\operatorname{QES}}(\alpha_n)} - 1 \right) \xrightarrow{d} \Gamma. \]

(ii) Under the conditions of Theorem 6 (ii),

\[ \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]} \left( \frac{\operatorname{XES}^*(\tau'_n(\alpha_n))}{\overline{\operatorname{QES}}(\alpha_n)} - 1 \right) \xrightarrow{d} \Gamma. \]
Let us now turn to $\hat{\text{XMES}}_{\tau_n}(\alpha_n)$ in (23) and $\hat{\text{XMES}}_{\tau_n}(\alpha_n)$ in (26) that estimate the same marginal expected shortfall $\text{XMES}(\tau_n(\alpha_n)) \equiv \text{QMES}(\alpha_n)$ as Cai et al. (2015)’s estimator $\hat{\text{QMES}}_{\tau_n}(\alpha_n)$ defined in (27). As a matter of fact, $\text{XMES}_{\tau_n}(\alpha_n)$ is nothing but $\hat{\text{QMES}}_{\tau_n}(\alpha_n).

**Theorem 8.** (i) Suppose the conditions of Theorem 4 hold with $\alpha_n$ in place of $\tau_n$. Then

$$\sqrt{n(1-\tau_n)} \log[(1-\tau_n)/(1-\alpha_n)] \left( \frac{\text{XMES}_{\tau_n}(\alpha_n)}{\text{QMES}(\alpha_n)} - 1 \right) \rightarrow \Gamma.$$ 

(ii) Suppose the conditions of Theorem 5 hold with $\alpha_n$ in place of $\tau_n$. Then

$$\sqrt{n(1-\tau_n)} \log[(1-\tau_n)/(1-\alpha_n)] \left( \frac{\text{XMES}_{\tau_n}(\alpha_n)}{\text{QMES}(\alpha_n)} - 1 \right) \rightarrow \Gamma.$$ 

### 7 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 7.1 and tail XMES estimators in section 7.2. 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.

#### 7.1 Expectile-based VaR and ES

To evaluate finite-sample performance of the extreme expectile estimators $\hat{\xi}_{\tau_n}(\alpha_n) \equiv \hat{\xi}_{\tau_n}(\alpha_n)$ and $\hat{\xi}_{\tau_n}(\alpha_n) \equiv \hat{\xi}_{\tau_n}(\alpha_n)$, we have considered simulated samples from the Student $t$-distribution ($\nu = 3, 5, 7, 9$), which corresponds to real-valued profit-loss variables, and from the marginal of the bivariate Student $t$-distribution described in (21), which corresponds to non-negative loss variables. We shall refer to this marginal distribution on $(0, \infty)$ as ‘truncated Student $t$-distribution’. We used in all our simulations the Hill estimator of $\gamma$, 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 $\hat{\xi}_{\tau_n}(\alpha_n)$ and $\hat{\xi}_{\tau_n}(\alpha_n)$, as functions of the sample
fraction $k$. Our tentative conclusion is that the accuracy of the direct estimator $\hat{\xi}_{r_n}$ is quite respectable relative to the indirect estimator $\hat{\xi}_{p_n}$. 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 $t_\nu$-distributions with $\nu \leq 2$.

![Figure 3](image)

**Figure 3:** Root MSE estimates (top panels) and Bias estimates (bottom panels) of $\hat{\xi}_{r_n}(k)/\xi_{r_n}$ (solid line) and $\hat{\xi}_{p_n}(k)/\xi_{p_n}$ (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 truncated Student distributions, displayed in Figure 4, indicate that the indirect estimator $\hat{\xi}_{p_n}$ is superior to the direct estimator $\hat{\xi}_{r_n}$. The use of the Pareto distribution $F_Y(y) = 1 - y^{-1/\gamma}$, $y > 1$, and the Fréchet distribution $F_Y(y) = e^{-y^{-1/\gamma}}$, $y > 0$, leads to the same conclusion. It may also be seen in both Student and truncated Student scenarios 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.

Another way of validating the presented estimation procedures for $\xi_{r_n}$ 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 $\hat{\xi}_{r_n}$ and $\hat{\xi}_{p_n}$ 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 A.2,
where the resulting average values of the realized losses seem to favor $\xi_{\tau_n}^*$ over $\xi_{\tau_n}^\dagger$ in the case of Student t-distributions, while they tend to prefer $\xi_{\tau_n}^\dagger$ over $\xi_{\tau_n}^*$ in the case of truncated Student t-distributions. Note that one can also compare the two forecasters $\xi_{\tau_n}^*$ and $\xi_{\tau_n}^\dagger$ by applying the modern and promising tool of ‘Murphy diagram’, recently developed by Ehm et al. (2015).

We also investigate the normality of the estimators $\xi_{\tau_n}^*$ and $\xi_{\tau_n}^\dagger$ in Supplement A.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-based ES estimators $\overline{\text{XES}}^*(\tau'_n)$, $\overline{\text{XES}}^\dagger(\tau'_n)$, $\overline{\text{XES}}(\tau'_n)$ and $\overline{\text{XES}}^\dagger(\tau'_n)$. The experiments all employed the same families of Student’s and truncated Student’s t-distributions as before. The lessons were similar to those from the expectile-based VaR setting, hence the results are not reported here. It may also be noticed that the Monte-Carlo estimates corresponding to $\overline{\text{XES}}^*(\tau'_n)$ and $\overline{\text{XES}}^\dagger(\tau'_n)$ [respectively, $\overline{\text{XES}}^*(\tau'_n)$ and $\overline{\text{XES}}^\dagger(\tau'_n)$] are very similar.
7.2 Expectile-based MES

Here, we compare the composite estimators \( \hat{X}_{\text{MES}}(\hat{\tau}_n^\nu(\alpha)) \) and \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n)) \) that estimate the same MES, \( X_{\text{MES}}(\tau_n^\nu(\alpha_n)) \equiv \text{QMES}(\alpha_n) \), as Cai et al. (2015)’s estimator \( \hat{Q}_{\text{MES}}(\alpha_n) \). The latter is actually identical to the indirect estimator \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n)) \). 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 \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n)) \) and \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n)) \), the simulation experiments first employ the Student \( t_\nu \)-distribution on \( (0, \infty)^2 \) with density \( f_\nu(x,y) \) described in (21). It can be shown that this distribution satisfies the conditions \( JC_2(R, \beta, \kappa) \) and \( C_2(\gamma_X, \rho_X, A_X) \) 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 \( \tau_n \), we used the same considerations as in Section 7.1.

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 \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n))/X_{\text{MES}}(\tau_n^{\nu}(\alpha_n)) \) and \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n))/X_{\text{MES}}(\tau_n^{\nu}(\alpha_n)) \), as functions of the sample fraction \( k \). 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.3, the limit Theorems 4 and 5 provide adequate approximations for finite sample sizes, with a slight advantage for \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n)) \).

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

\[
(X, Y) = \left( Z_1^{\nu^2/4} \mathbb{I}(Z_1 \geq 0) - (-Z_1)^{\nu^2/8} \mathbb{I}(Z_1 < 0), Z_2 \right),
\]

where \((Z_1, Z_2)\) denotes a standard Student \( t_\nu \)-distribution on \( \mathbb{R}^2 \) with density

\[
\frac{1}{2\pi} \left( 1 + (x^2 + y^2) / \nu \right)^{-(\nu+2)/2}, \quad x, y \in \mathbb{R}.
\]

The resulting Monte-Carlo estimates for \( \nu \in \{3, 5\} \), displayed in Figure 6, indicate that \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n)) \) is more efficient relative to \( \hat{X}_{\text{MES}}(\hat{\tau}_n^{\nu}(\alpha_n)) \). This superiority of the direct estimator is, however, no more 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).

8 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.
8.1 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 and scatterplot shown in Figure 7 (top) give evidence of an important right-skewness. 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 $\hat{q}_{\alpha_n} = Y_{n-k,n} \left( \frac{k}{n \tau_n} \right)^{\hat{\gamma}_H}$ as described in (10), where $\hat{\gamma}_H$ 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 (see Basel Committee, 2011), 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 (bottom) shows the quantile-VaR estimates $\hat{q}_{\alpha_n}$ against the sample fraction $k$ (rainbow curve). A commonly used heuristic
Figure 6: Root MSE estimates (top panels) and Bias estimates (bottom panels) of $\overline{\text{XMES}}$/XMES (solid line) and $\hat{\text{XMES}}$/XMES (dashed line), as functions of $k$, for the transformed $t_3$ and $t_5$-distributions on $\mathbb{R}^2$, respectively, from left to right.

An 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 $\hat{\gamma}_H$ on $\hat{\alpha}_{\alpha_n}$ is highlighted by a colour-scheme, ranging from dark red (low $\hat{\gamma}_H$) to dark violet (high $\hat{\gamma}_H$). This Hill estimate of the tail index $\gamma$ seems to mainly vary within the interval $[0.27, 0.43]$.

The alternative expectile-based estimator $\hat{\xi}_{\tau_1\alpha_n}$ introduced in Section 6, which estimates the same VaR $q_{\alpha_n} = \xi_{\tau_1\alpha_n}$ as the quantile-based estimator $\hat{q}_{\alpha_n} = \hat{\xi}_{\tau_1\alpha_n}$, is also graphed in Figure 7 in yellow. As an asymmetric-least-squares estimator, it is more affected by the infrequent great claim amounts visualized in the top figure. Its plot in yellow indicates a more conservative risk measure between 3.92 and 4.33 million, over the stable region $k \in [150, 500]$. Yet, it is less severe than the maximal recorded claim amount $Y_{n,n}$.

An 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

\[ \hat{QES}^* (\alpha_n) = \frac{1}{k} \sum_{i=1}^{n} Y_i \mathbb{I}(Y_i > Y_{n-k,n}) \cdot \left( \frac{k}{np_n} \right)^{\hat{a}Y} \]  

[see El Methni et al. (2014)]. Its graph shown in Figure 7 (bottom) in red dashed 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. The graph of \( \hat{QES}^* (\alpha_n) \) is very close to the plot in gray of the indirect expectile-based estimator \( \hat{XES}^* (\hat{\tau}_n (\alpha_n)) \) which estimates the same risk measure \( XES(\tau'_n (\alpha_n)) \equiv QES(\alpha_n) \). In contrast to \( \hat{XES}^* (\hat{\tau}_n (\alpha_n)) \), which indicates an averaged risk estimate of around 6.14 million over the stable region \( k \in [150, 500] \), both direct expectile-based estimators \( \hat{XES}^* (\hat{\tau}_n (\alpha_n)) \) in cyan and \( \hat{XES}^* (\hat{\tau}_n (\alpha_n)) \) in dashed black are clearly more pessimistic as they rely on the asymmetric-least-squares estimator \( \xi_{\tau'_n (\alpha_n)} \). The averaged values of these ES estimates over \( k \in [150, 500] \) are, respectively, around 6.5 and 6.48 million. They exceed the traditional quantile-based ES estimate by 0.37 million.

That eternal maxim of the cautious aunt and misanthropic uncle, “expect the worst, and you won’t be disappointed” [Bassett et al. (2004)] might thus be transformed into a concrete calculus via the asymmetric least squares-based VaR and ES estimates.

### 8.2 MES of three large US financial institutions

We consider the same investment banks as in the studies of Brownlees and Engle (2016) 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, 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 \( \hat{QMES}^* (\alpha_n) \), as defined in (27), to estimate the quantile-based MES, \( QMES(\alpha_n) = \mathbb{E}\{X|Y > q_{Y,\alpha_n}\} \), where \( \alpha_n = 1 - \frac{1}{n} = 1 - 1/2513 \), with two intermediate sequences involved in \( \hat{\gamma}_X \) and \( \hat{QMES}(\tau_n) \) to be chosen in two steps. Instead, we use our expectile-based method to estimate \( QMES(\alpha_n) = XES(\tau_n (\alpha_n)) = \mathbb{E}\{X|Y > \xi_{Y,\tau'_n (\alpha_n)}\} \), with the same extreme relative frequency \( \alpha_n \) that corresponds to a once-per-decade systemic event. We employ the rival estimator \( \hat{QMES}^* (\alpha_n) \) with the same intermediate sequence \( \tau_n = 1 - \frac{k}{n} \) in both \( \hat{\gamma}_X \) and \( \hat{QMES}(\tau_n) \). 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 \( \gamma_X \). This assumption is confirmed by the plot of the Hill estimates of \( \gamma_Y \) against the sample fraction \( k \) (green curve) in Figure 8 (a). Indeed, the first stable region appears for
Figure 7: SOA Group Medical Insurance data. (top) Histogram and scatterplot of the log-claim amounts; (bottom) The expectile-based VaR and ES plots \( (k, \tilde{\gamma}_n^\tau (\alpha_n)(k)) \), in yellow, \((k, \tilde{\gamma}_n^\tau (\alpha_n)) \), in gray, \((k, \tilde{\gamma}_n^\tau (\alpha_n)) \), in cyan and \((k, \tilde{\gamma}_n^\tau (\alpha_n)) \), in black, along with the quantile-based VaR and ES plots \((k, \tilde{\gamma}_n^\tau (\alpha_n)) \), as rainbow curve and \((k, \tilde{\gamma}_n^\tau (\alpha_n)) \), in black. The sample maximum \( Y_{n,n} \) is indicated by the horizontal pink line.

\( k \in [70, 100] \) with an averaged estimate \( \tilde{\gamma}_Y = 0.35 \). Hence, by Proposition 5, the estimates \( \tilde{\gamma}_n^\tau (\alpha_n) \) and \( \tilde{\gamma}_n^\tau (\alpha_n) \) are expected to be less extreme than the benchmark values \( \tilde{\gamma}_n^\tau (\alpha_n) \). This is visualised in Figure 18 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 \( \tilde{\gamma}_n^\tau (\alpha_n) \) and \( \tilde{\gamma}_n^\tau (\alpha_n) \) estimate the less extreme risk measure \( \tilde{\gamma}_n^\tau (\alpha_n) \) and not the desired intuitive tail measure \( \tilde{\gamma}_n^\tau (\alpha_n) \).

The interest here is rather on the composite estimators \( \tilde{\gamma}_n^\tau (\alpha_n) \) and \( \tilde{\gamma}_n^\tau (\alpha_n) \), where \( \tilde{\gamma}_n^\tau (\alpha_n) \) is actually nothing but \( \tilde{\gamma}_n^\tau (\alpha_n) \). The two rival estimates \( \tilde{\gamma}_n^\tau (\alpha_n) \) and \( \tilde{\gamma}_n^\tau (\alpha_n) \) 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 region of the plots (b)-(d)
Figure 8: (a) Hill estimates $\hat{\gamma}_Y$ based on daily loss returns of market index (green), along with $\hat{\gamma}_X$ based on daily loss returns of three investment banks: Goldman Sachs (red), Morgan Stanley (blue), T. Rowe Price (yellow). (b)-(d) The estimates $\overline{QMES}^* (\alpha_n)$ in black line and $\overline{XMES}^* (\hat{\gamma}_n (\alpha_n))$ as rainbow curve for the three banks, with $n = 2513$ and $\alpha_n = 1 - 1/n$. 
appears, 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. 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 \( \hat{\text{QMES}}^*(\alpha_n) \), 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 visualised in Figure 8 (b)-(d), where the plots of \( \hat{\text{QMES}}^*(\alpha_n) \), in black line, and \( \hat{\text{XMES}}^*(\tau_1(\alpha_n)) \), as rainbow curve, exhibit a very similar evolution for the three banks. The effect of the Hill estimates \( \hat{\gamma}_X \) on the MES estimates is highlighted by the “colouring-scheme” of \( \hat{\text{XMES}}^*(\tau_1(\alpha_n)) \), ranging from dark red (low) to dark violet (high).

<table>
<thead>
<tr>
<th>Bank</th>
<th>( \text{XMES}^*(\tau_1(\alpha_n)) )</th>
<th>( \text{QMES}^*(\alpha_n) )</th>
<th>( \text{XMES}^*(\tau_1(\alpha_n)) )</th>
<th>( \text{QMES}^*(\alpha_n) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldman Sachs</td>
<td>0.3123</td>
<td>0.3077</td>
<td>0.3423</td>
<td>0.3375</td>
</tr>
<tr>
<td>Morgan Stanley</td>
<td>0.5622</td>
<td>0.5552</td>
<td>0.6495</td>
<td>0.6641</td>
</tr>
<tr>
<td>T. Rowe Price</td>
<td>0.3308</td>
<td>0.3098</td>
<td>0.3407</td>
<td>0.3405</td>
</tr>
</tbody>
</table>

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 \((n = 2513 \text{ and } \alpha_n = 1 - 1/n)\). The last two columns report the results based on weekly loss returns from the same sample period \((n = 522 \text{ and } \alpha_n = 1 - 1/n)\).

In our theorems 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 \( \gamma_Y \) and \( \text{QMES}(\alpha_n) \equiv \text{XMES}(\tau_1(\alpha_n)) \), with \( \alpha_n = 1 - \frac{1}{n} \), are displayed in Figure 9 as functions of \( k \). The averaged estimate \( \hat{\gamma}_Y = 0.37 \) is obtained from the first stable region \( k \in [25, 35] \) of the plot (a). The first stable region of the plots (b)-(d) appears, 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 (2016). 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 extreme value theorems may work under serial dependence with enlarged
asymptotic variances.

Acknowledgements

The authors would like to thank three referees, the Associate Editor and the Joint Editor for
their valuable suggestions, which have significantly improved the paper. We are also grateful
to Juan-Juan Cai and Chen Zhou who kindly send us their codes about MES numerical
illustrations. A. Daouia’s research was supported by the Toulouse School of Economics
Individual Research Fund (IRF/Daouia-20125).

Supplementary materials

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

References

navian Actuarial Journal, DOI: 10.1080/03461238.2015.1017527.
Mathematical Finance, 9, 203–228.
banks and banking systems.
Figure 9: (a) Hill estimates $\hat{\gamma}_Y$ based on weekly loss returns of market index (green), along with $\hat{\gamma}_X$ based on weekly loss returns of three investment banks: Goldman Sachs (red), Morgan Stanley (blue), T. Rowe Price (yellow). (b)-(d) The estimates $QMES^*(\alpha_n)$ in black line and $XMES^*(\hat{\gamma}_n(\alpha_n))$ as rainbow curve for the three banks, with $n = 522$ and $\alpha_n = 1 - 1/n$. 


