Extremiles: A new perspective on asymmetric least squares

May 17, 2018

Abstract

Quantiles and expectiles of a distribution are found to be useful descriptors of its tail in the same way as the median and mean are related to its central behavior. This paper considers a valuable alternative class to expectiles, called extremiles, which parallels the class of quantiles and includes the family of expected minima and expected maxima. The new class is motivated via several angles, which reveals its specific merits and strengths. Extremiles suggest better capability of fitting both location and spread in data points and provide an appropriate theory that better displays the interesting features of long-tailed distributions. We discuss their estimation in the range of the data and beyond the sample maximum. A number of motivating examples are given to illustrate the utility of estimated extremiles in modeling noncentral behavior. There is in particular an interesting connection with coherent measures of risk protection.

1 Introduction

Given a data set from a random variable Y, quantiles and expectiles are found to be useful descriptors of its extreme tails in the same way as the median and mean are related to its central behavior. Both can be defined with the help of appropriate loss functions. For a fixed τ in $(0, 1)$, the τth quantile of Y can be obtained by minimizing asymmetrically weighted mean absolute deviations (Koenker and Bassett, 1978):

$$q_\tau \in \operatorname{argmin}_\theta \mathbb{E}\left\{|\tau - 1\!\!1(Y \leqslant \theta)| \cdot |Y - \theta| - |\tau - 1\!\!1(Y \leqslant 0)| \cdot |Y|\right\}, \tag{1}$$

with $1\!\!1(\cdot)$ being the indicator function. The τth expectile is obtained in a similar way with absolute deviations replaced by squared deviations (Newey and Powell, 1987):

$$e_\tau = \operatorname{argmin}_\theta \mathbb{E}\left\{|\tau - 1\!\!1(Y \leqslant \theta)| \cdot |Y - \theta|^2 - |\tau - 1\!\!1(Y \leqslant 0)| \cdot |Y|^2\right\}. \tag{2}$$

The special case $\tau = \frac{1}{2}$ leads to the median and the mean of Y in (1) and (2), respectively. Although different in their construction, both quantiles and expectiles have similar properties. The reason for this, as proved by Jones (1994), is that expectiles are precisely quantiles but for a transformation of the original distribution. Abdous and Remillard (1995) showed that quantiles and expectiles of the same distribution coincide under the hypothesis of weighted-symmetry.

Despite their strong intuitive appeal, quantiles are not always satisfactory. They can be criticized for being somewhat difficult to compute as the corresponding loss function is not continuously differentiable (though modern efficient linear programming algorithms are available). Most importantly, they are relatively inefficient against long-tailed distributions as they are based on absolute rather than squared loss minimization. Finally, the asymptotic variance of the diverse known quantile estimators involves the so-called sparsity function (value of the density function at those quantiles) whose estimators converge very slowly and depend heavily on the choice of the smoothing parameter (see, *e.g.*, Cheng and Parzen, 1997). The use of expectiles reduces some of these vexing inconveniences.

The key advantage of the expectile e_τ over the quantile q_τ is its efficiency and computing expedience (using iteratively-reweighted least squares), although it does not have an interpretation as direct as q_τ in terms of relative frequency. Its unique interpretation is given by the equation $\mathbb{E}\left\{|Y - e_\tau| \mathbb{I}(Y \leqslant e_\tau)\right\} / \mathbb{E}|Y - e_\tau| = \tau$. Also, inference on e_τ is not governed by the estimation of the density function and it only requires that the second moment of Y be bounded (see Corollary 4 in Holzmann and Klar, 2016). However, the complicated cost function defining expectiles makes their use in statistics of extremes and tail analysis a hard mathematical problem (see Daouia *et al.*, 2018). Perhaps most importantly, the expectile terminology coined for e_τ is frustrated by the absence of any closed form expression in terms of tail expectations in the same way that the quantile q_τ is explicitly determined by the generalized inverse of the distribution function.

The present paper proposes a new least squares analogue of quantiles, called extremiles, which defines a valuable alternative option to both quantiles and expectiles for general statistical diagnoses. As shown in the following section, the new class is a generalization of the usual central moment $\mathbb{E}(Y)$, which summarizes the distribution of Y in the same way as the expected values of extreme order statistics do. Extremiles are by construction more alert/sensitive to the magnitude of extreme values than quantiles and make more efficient use of the available data since they rely on both the distance to observations and their probability, while quantiles only depend on the frequency of observations below or above the predictor and not on their values. Unlike expectiles, extremiles benefit from various equivalent explicit formulations and more intuitive interpretations. They are proportional to specific probability-weighted moments (PWMs) and can be estimated by L-statistics, M-statistics, linearized M-statistics and PWM-estimators. In addition, inference on extremiles is much easier than inference on both expectiles and quantiles, since their various estimators have closed form expressions and are computationally efficient. Also, these estimators steer an advantageous middle course between the excessive robustness of ordinary quantile estimators and severe sensitivity of extreme quantile estimators in the sense that the extremile estimators of order τ tend to be more tail sensitive than the τth quantile estimators for ordinary levels τ, but become more resistant for extremely high/low levels τ. Finally, extremiles are very closely connected to quantiles from an extreme-value perspective, and provide an appropriate theory that better displays the interesting features of long-tailed population distributions. Although extremiles have been used in Daouia and Gijbels (2011) in the context of estimating a frontier cost function (involving a conditional expectation), the current paper is the first to introduce the notion of extremiles as a new class of least squares analogues of quantiles, and to give a full study of this class, including relationships with expectiles and quantiles, as well as extensive statistical inference within this class.

The presented M- and L-estimates of extremiles, as well as their linearized M- and PWM-estimates, may suffer from instability at very far tails due to data sparseness, especially for heavy-tailed distributions. This motivated us to extend their estimation and the underlying asymptotic theory far enough into the tails, which translates into considering the level $\tau = \tau_n \to 1$ as the sample size n goes to infinity. We show that these high τ_nth extremiles enjoy a very interesting connection with the extreme τ_nth quantiles and the tail index of the underlying distribution. Our first estimation method is based on this asymptotic connection, so that modern extreme-value estimates of large quantiles and of the tail index can be used for extrapolation beyond the range of the data. Our second method relies directly on asymmetric least squares estimation. Although both approaches work quite well and either might be used in practice, we have a particular preference for the second due to some theoretical findings and simulation evidence.

The extremile and the quantile of Y with the same level τ are actually identical to the mean and the median, respectively, of a common asymmetric distribution $K_\tau(F)$, where F stands for the distribution function of Y and K_τ is a well-specified power transformation. The use of extremiles appears then naturally in the context of any decision theory where 'optimistic' and 'pessimistic' judgements are contrasted such as, for instance, extreme risk analysis, survival analysis and medical decision making. In this paper, we focus on risk management, where heavy-tailed distributions describe quite well the tail structure of actuarial and financial data [see, *e.g.*, Embrechts *et al.* (1997) and Resnick (2007)]. Extremiles bear much better than quantiles the burden of representing an alert risk measure to the magnitude of infrequent catastrophic losses. Although expectiles have recently attracted a lot of interest as measures of risk [see Daouia *et al.* (2018) and the references therein], their absence of comonotonic additivity is a serious problem for a regulatory risk standard [Acerbi and Szekely (2014)]. Theoretical and numerical results indicate that extremiles are perfectly reasonable alternatives to both quantiles and expectiles. In particular, they are comonotonically additive, coherent spectral risk measures of law–invariant type. Also, they belong to the class of Wang (1996)'s distortion risk measures with concave distortion function that acts to depress the likelihood of the most favorable outcomes and to accentuate the likelihood of the least favorable ones, yielding thus the desired form of pessimistic decision theory [Bassett *et al.* (2004)]. To illustrate the discussed ideas, we consider data examples on Trended Hurricane Losses and Medical Insurance Large Claims.

The paper is further organized as follows. Section 2 presents a detailed description of the proposed new class, including its motivation, interpretation and basic properties. Section 3 deals with estimation of extremiles in the range of the data and beyond the sample maximum. Section 4 provides an interesting connection between extremiles and coherent law-invariant measures of risk in actuarial and financial management. Section 5 concludes. All the necessary mathematical proofs are given in the supplementary file.

2 The class of extremiles

2.1 Definition and motivation

Given $\tau \in (0,1)$, the quantile q_τ can uniquely be defined as the generalized inverse $q_\tau = F^{-1}(\tau) := \inf\{y : F(y) \geqslant \tau\}$ of the underlying cumulative distribution function F. For ease of presentation, we assume throughout the paper that F is continuous. It is not hard to verify that q_τ is identical to the median of a random variable Z_τ having cumulative distribution function $F_{Z_\tau} = K_\tau(F)$, where

$$K_\tau(t) = \begin{cases} 1-(1-t)^{s(\tau)} & \text{if} \quad 0 < \tau \leqslant 1/2 \\ t^{r(\tau)} & \text{if} \quad 1/2 \leqslant \tau < 1 \end{cases}$$

is a distribution function with support $[0,1]$, and $r(\tau) = s(1-\tau) = \log(1/2)/\log(\tau)$. Hence, the τth quantile can be derived from the alternative minimization problem

$$q_\tau \in \text{argmin}_\theta \mathbb{E}\left\{J_\tau\left(F(Y)\right) \cdot \left[|Y-\theta| - |Y|\right]\right\},$$

with the special weight-generating function $J_\tau(\cdot) = K_\tau'(\cdot)$ on $(0,1)$. This does not seem to have been appreciated in the literature before. The parallel concept to the quantile q_τ which we call extremile of order τ of Y is then defined in a similar way by substituting the squared deviations in place of the absolute deviations:

$$\xi_\tau = \text{argmin}_\theta \mathbb{E}\left\{J_\tau\left(F(Y)\right) \cdot \left[|Y-\theta|^2 - |Y|^2\right]\right\}. \tag{3}$$

As a matter of fact, while q_τ coincides with the median of the transformation Z_τ, it is easily seen that, whenever $\mathbb{E}|Z_\tau| < \infty$,

$$\xi_\tau = \mathbb{E}(Z_\tau). \tag{4}$$

We shall see in Proposition 2 that the condition $\mathbb{E}|Z_\tau| < \infty$ is implied by $\mathbb{E}|Y| < \infty$, and therefore extremiles of any order exist as soon as Y has a finite first moment. Denote by $y_\ell = \inf\{y : F(y) > 0\}$ and $y_u = \sup\{y : F(y) < 1\}$ the lower and, respectively, upper endpoint of the support of F. One way of defining the extremile ξ_τ, for $0 \leqslant \tau \leqslant 1$, is as the explicit quantity

$$\xi_\tau = \begin{cases} -\displaystyle\int_{y_\ell}^0 \left\{1 - [1-F(y)]^{s(\tau)}\right\} dy + \int_0^{y_u} [1-F(y)]^{s(\tau)}\, dy & \text{for} \quad 0 \leqslant \tau \leqslant 1/2 \\ -\displaystyle\int_{y_\ell}^0 [F(y)]^{r(\tau)}\, dy + \int_0^{y_u} \left(1 - [F(y)]^{r(\tau)}\right) dy & \text{for} \quad 1/2 \leqslant \tau \leqslant 1. \end{cases} \tag{5}$$

This follows from a general property of expectations (see Shorack, 2000, p.117). Clearly, F_{Z_τ} reduces to F when $\tau = 1/2$, and $\xi_{1/2}$ is just the expectation of Y, while the endpoints y_ℓ and y_u of the support of F coincide respectively with the lower and upper extremiles ξ_0 and ξ_1, since $s(0) = r(1) = \infty$ in (5). A thorough description of basic properties of ξ_τ is given in Section 2.3.

Extremiles can be of considerable importance for modeling extremes of natural phenomena. When $\tau \geqslant 1/2$ with $r(\tau) = 1, 2, \ldots$ the τth extremile has a nice interpretation:

it equals the expectation of the maximum of $r(\tau)$ independent copies $Y^1,\ldots,Y^{r(\tau)}$ of Y, *i.e.* $\xi_\tau=\mathbb{E}[\max(Y^1,\ldots,Y^{r(\tau)})]$. Of interest is the case $\tau\uparrow 1$, or equivalently $r(\tau)\to\infty$, which leads to access the upper endpoint y_u of the support of F. Likewise, if $\tau\leqslant 1/2$ with $s(\tau)=1,2,\ldots$ we have $\xi_\tau=\mathbb{E}[\min(Y^1,\ldots,Y^{s(\tau)})]$, the expectation of the minimum of $s(\tau)$ i.i.d. observations from Y. Of interest is also the case $\tau\downarrow 0$, or equivalently $s(\tau)\to\infty$, which leads to access the lower bound y_ℓ of the support of Y. For a general order τ, we have

$$\begin{aligned}
\mathbb{E}\left[\max\left(Y^1,\ldots,Y^{\lfloor r(\tau)\rfloor}\right)\right]\leqslant \quad \xi_\tau \quad &\leqslant \mathbb{E}\left[\max\left(Y^1,\ldots,Y^{\lfloor r(\tau)\rfloor+1}\right)\right] \quad \text{if} \quad \frac{1}{2}\leqslant\tau<1,\\
\mathbb{E}\left[\min\left(Y^1,\ldots,Y^{\lfloor s(\tau)\rfloor+1}\right)\right]\leqslant \quad \xi_\tau \quad &\leqslant \mathbb{E}\left[\min\left(Y^1,\ldots,Y^{\lfloor s(\tau)\rfloor}\right)\right] \quad \text{if} \quad 0<\tau\leqslant\frac{1}{2},
\end{aligned}$$

where $\lfloor\cdot\rfloor$ denotes the integer part (or floor function) and $Y^1,Y^2,\ldots$ are i.i.d. observations from Y. The bracketing of ξ_τ becomes narrower when $\tau\uparrow 1$ or $\tau\downarrow 0$.

Yet, there is still another way of looking at ξ_τ for orders τ in $(0,1)$. These extremiles are likely to be most useful when the quantile function q can be written in closed form, for then we have

$$\xi_\tau=\int_0^1 F_{Z_\tau}^{-1}(u)du=\int_0^1 q_t\,dK_\tau(t)=\int_0^1 J_\tau(t)q_t\,dt. \tag{6}$$

These expressions are key when it comes to proposing an estimator for an extremile. They clearly link the τth extremile of the random variable Y to its quantile function as a weighted average of q, which is the most convenient way of evaluating ξ_τ.

2.2 Relating extremiles and PWMs

Extremiles are closely related to the concept of probability-weighted moments (PWMs) introduced by Greenwood *et al.* (1979) and defined by the quantities

$$M_{p,r,s}=\mathbb{E}\left[Y^p\{F(Y)\}^r\{1-F(Y)\}^s\right],$$

where p,r,s are nonnegative real numbers. These moments have been extensively utilized in extreme-value procedures (see, *e.g.*, Beirlant *et al.*, 2004 and de Haan and Ferreira, 2006). Generally, a distribution is characterized either by the moments $M_{1,0,s}$

($s = 0, 1, 2, \ldots$) or by the moments $M_{1,r,0}$ ($r = 0, 1, 2, \ldots$). In our approach, the moments $M_{1,0,s}$ (with $s \in \mathbb{R}_+$) are favored for describing the distribution of the population in the left tail (*i.e.* for $\tau \leqslant 1/2$), but a more convenient definition for ξ_τ in the right tail ($\tau \geqslant 1/2$) is formulated by considering the moments $M_{1,r,0}$ (with $r \in \mathbb{R}_+$). Indeed, for $0 < \tau < 1$, we can rewrite (6) as

$$\xi_\tau = \int_0^1 J_\tau(t) q_t \, dt = \mathbb{E}\left[Y J_\tau\left(F(Y)\right)\right], \tag{7}$$

or equivalently

$$\xi_\tau = \begin{cases} s(\tau) M_{1,0,s(\tau)-1} & \text{for} \quad 0 < \tau \leqslant 1/2 \\ r(\tau) M_{1,r(\tau)-1,0} & \text{for} \quad 1/2 \leqslant \tau < 1. \end{cases} \tag{8}$$

Thus, extremiles are proportional to specific PWMs. The weight-generating function $J_\tau(\cdot)$ being monotonically increasing for $\tau \geqslant 1/2$ and decreasing for $\tau \leqslant 1/2$, the extremile ξ_τ (see equation (7)) depends by construction on all feasible values of Y, putting more weight to the high values for $\tau \geqslant 1/2$ and more weight to the low values for $\tau \leqslant 1/2$. Therefore ξ_τ is sensible to the magnitude of extreme values for any order $\tau \in (0, 1)$. In contrast, the τth quantile q_τ is determined solely by the probability level (relative frequency) τ, and so it may be unaffected by extreme values whatever the shape of the tail distribution, unless τ is very extreme. In addition, when sample quantiles break down at $\tau \downarrow 0$ or $\tau \uparrow 1$, the various extremile estimators discussed below in Section 3 remain more resistant to extreme values thanks to their formulation as L-functionals whose weighting function J_τ converges to 0 pointwise on $(1/2, 1)$ at a geometrically fast rate as $\tau \uparrow 1$. Hence, the new class steers an advantageous middle course between the robustness of ordinary quantiles and the sensitivity of the extreme ones.

The interpretation in terms of expected minimum and expected maximum of a sample from Y, the impact on the lower and upper tails of the underlying distribution, the sensitivity and resistance properties to extremes and the proportionality to PWMs are of particular interest in extreme-value theory. This inspired the name *extremiles* for this class. As a matter of taste, we prefer to regard equation (7), or equivalently (6), as the most convenient definition of extremiles from a mathematical perspective. It should be clear that the first-order necessary condition for optimality related to the initial definition

(3) leads to

$$\xi_\tau = \frac{\mathbb{E}\left[Y J_\tau\left(F(Y)\right)\right]}{\mathbb{E}\left[J_\tau\left(F(Y)\right)\right]},$$

which in turn gives the identity (7), since $\mathbb{E}\left[J_\tau\left(F(Y)\right)\right] = 1$ for all $\tau \in (0,1)$ by continuity of F. Note also that the transformed random variable Z_τ in (4) has the same expectation ξ_τ as the random variable $Y J_\tau\left(F(Y)\right)$, but not necessarily the same continuous distribution. In the particular case where $\tau \geqslant 1/2$ and $r(\tau)$ is an integer, it is easily seen that Z_τ is the maximum of $r(\tau)$ independent copies of Y; similarly, when $\tau \leqslant 1/2$ and $s(\tau)$ is an integer, Z_τ is the minimum of $s(\tau)$ independent copies of Y. In the general setting, we have the following characterization of Z_τ.

Proposition 1 *For a random variable Y with continuous distribution function F and quantile function q, and for $\tau \in (0,1)$, define*

$$\phi_\tau = q \circ K_\tau^{-1} \circ F \quad \text{i.e.} \quad \phi_\tau(y) = q_{K_\tau^{-1}(F(y))}.$$

Then $Z_\tau \overset{d}{=} \phi_\tau(Y)$.

2.3 Basic properties

An alternative justification for the use of extremiles to describe probability distributions may be based on the following propositions. As established in Proposition 2, extremiles have similar basic properties as quantiles and expectiles. The proofs of Proposition 2 and of all further theoretical results are provided in the Supplementary Material document.

Proposition 2 *(i) If $\mathbb{E}|Y| < \infty$ then ξ_τ exists for any $\tau \in (0,1)$ and (if Y is not a constant) defines a continuous increasing function which maps $(0,1)$ onto the set $\{y \in \mathbb{R} \,|\, 0 < F(y) < 1\}$.*

(ii) Two integrable random variables Y and $\widetilde{Y}$ have the same distribution if and only if $\xi_{Y,\tau} = \xi_{\widetilde{Y},\tau}$ for every $\tau \in (0,1)$.

(iii) The τth extremile of the linear transformation $\widetilde{Y} = a + bY$, $a, b \in \mathbb{R}$, is given by

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

(iv) If Y has a symmetric distribution with mean μ, then $\xi_{1-\tau} = 2\mu - \xi_\tau$ for any $\tau \in (0,1)$.

(v) If Y and $\widetilde{Y}$ are comonotone [i.e. there exists a third random variable $\bar{Y}$ and increasing functions u and v such that $Y = u(\bar{Y})$ and $\widetilde{Y} = v(\bar{Y})$], then $\xi_{Y+\widetilde{Y},\tau} = \xi_{Y,\tau} + \xi_{\widetilde{Y},\tau}$.

The wide range of distributions covered by property *(i)* can be extended further by relaxing the condition of finite "absolute" first moments. By *(ii)*, a distribution with finite absolute first moment is uniquely defined by its class of extremiles. Expectiles satisfy this law invariance property as well, but only for distributions with continuous densities [see Theorem 1 in Newey and Powell (1987)]. Extremiles are also location and scale equivariant by property *(iii)* in the same way as quantiles and expectiles are [see equation (A.1) in the Supplementary Material document for quantiles and Theorem 1 in Newey and Powell (1987) for expectiles]. The desirable properties *(iv)* for symmetric distributions and *(v)* on comonotonic additivity are also shared by quantiles. The implications of these properties in regression analysis are clear. For example, the conditional extremile curves will be parallel to each other if the conditional distributions of the response are homogeneous. Also, for any τ, the lower τth conditional extremile curve and the upper τth conditional extremile curve will be symmetric about the mean curve if the conditional distributions of the response are symmetric. Expectiles also satisfy property *(iv)* [Newey and Powell (1987)], but they are not comonotonically additive [Acerbi and Szekely (2014)]. They have recently attracted a lot of interest as risk measures, but absence of comonotonic additivity is a serious problem for a regulatory risk standard, as justified by Acerbi and Szekely (2014). We shall discuss further properties of extremiles as risk measures as well as their relation to expectiles in Sections 4.1 and 4.2.

We now describe what happens for large extremiles ξ_τ and how they are linked to extreme quantiles q_τ when F is attracted to the Fisher-Tippett distributions of extreme-value types:

(Fréchet) $\Phi_\gamma(y) = \exp\{-y^{-1/\gamma}\}$ with support $[0,\infty)$ and $\gamma > 0$;

(Weibull) $\Psi_\gamma(y) = \exp\{-(-y)^{-1/\gamma}\}$ with support $(-\infty, 0]$ and $\gamma < 0$;

(Gumbel) $\Lambda(y) = \exp\{-e^{-y}\}$ with support $\mathbb{R}$.

Let $\mathrm{DA}(\cdot)$ denote the maximum domain of attraction of an extreme-value distribution, *i.e.*, the set of distribution functions whose asymptotic distributions of suitably normalized maxima are of an extreme-value type. Let $\Gamma(\cdot)$ be the Gamma function.

Proposition 3 *Suppose that* $\mathbb{E}|Y| < \infty$.

(i) If $F \in DA(\Phi_\gamma)$ *with* $\gamma < 1$*, then*

$$\frac{\xi_\tau}{q_\tau} \sim \Gamma(1-\gamma)\{\log 2\}^\gamma \quad \textit{as} \quad \tau \uparrow 1.$$

(ii) If $F \in DA(\Psi_\gamma)$*, then* $y_u < \infty$ *and*

$$\frac{y_u - \xi_\tau}{y_u - q_\tau} \sim \Gamma(1-\gamma)\{\log 2\}^\gamma \quad \textit{as} \quad \tau \uparrow 1.$$

(iii) If $F \in DA(\Lambda)$ *and* $y_u = \infty$*, then* $\xi_\tau \sim q_\tau$ *as* $\tau \uparrow 1$*. If on the contrary* $y_u < \infty$*, then*

$$y_u - \xi_\tau \sim y_u - q_\tau \quad \textit{as} \quad \tau \uparrow 1.$$

Note that the moments of $F \in \mathrm{DA}(\Phi_\gamma)$ do not exist when $\gamma > 1$. The index γ tunes the tail heaviness of the distribution function F, with higher positive values indicating heavier tails. A consequence of Proposition 3 is that the extremiles of distributions with heavy tails of index $\gamma < 1$ are asymptotically more spread than the quantiles since $\Gamma(1-\gamma)\{\log 2\}^\gamma > 1$. Indeed, the function $\varphi : y \mapsto \log\left(\Gamma(1-y)\{\log 2\}^y\right)$ has derivative

$$\varphi'(y) = -\digamma(1-y) + \log(\log 2), \quad \text{for all } y \in (0,1),$$

where $\digamma(x) = \Gamma'(x)/\Gamma(x)$ denotes the digamma function. Because $\digamma$ is increasing and $\digamma(1) \approx -0.577$ (see Formulae 6.3.2 and 6.3.21 in Abramovitz and Stegun, 1972), the function φ is increasing. Hence, for any $\gamma \in (0,1)$, $\Gamma(1-\gamma)\{\log 2\}^\gamma > \exp(\varphi(0)) = 1$ as announced. A similar property holds for short-tailed distributions $F \in \mathrm{DA}(\Psi_\gamma)$ depending on the value of the extreme-value index: numerically, when $-0.2907 < \gamma < 0$, extremiles are asymptotically closer to the right endpoint than quantiles are. By contrast, when $\gamma < -0.2907$, we rather have that quantiles are asymptotically closer to the right endpoint than extremiles are. Finally, ξ_τ and q_τ are asymptotically equivalent for light-tailed distributions $F \in \mathrm{DA}(\Lambda)$. A similar proposition can of course be given when F is rather in the minimum domain of attraction of a Fisher-Tippett extreme-value distribution.

3 Estimation of extremiles

This section shows that results for ordinary and trimmed extremiles are easily obtained by means of L-statistics theory. By contrast, asymmetric least squares estimation of high extremiles leads to non-trivial developments from the perspective of extreme-value theory. Our theorems are derived for independent and identically distributed random variables. Although the time series case goes beyond the scope of the present paper, the recent analysis of Daouia *et al.* (2017) suggests that it may be possible to extend the results obtained here under appropriate mixing conditions on the underlying series.

3.1 Ordinary extremiles

Given a random sample $Y_1, Y_2, \ldots, Y_n$ from Y, a natural estimator for the extremile of fixed order $\tau \in (0,1)$ is easily obtained by replacing F with its empirical version $\widehat{F}_n$ in (5), or equivalently, by replacing q_t with its empirical analogue $\widehat{q}_t$ in (6), leading to an L-statistic generated by the measure dK_τ:

$$\widehat{\xi}_\tau^L = \int_0^1 \widehat{q}_t \, dK_\tau(t) = \sum_{i=1}^n \left\{ K_\tau\left(\frac{i}{n}\right) - K_\tau\left(\frac{i-1}{n}\right)\right\} Y_{i,n},$$

where $Y_{1,n} \leqslant Y_{2,n} \leqslant \cdots \leqslant Y_{n,n}$ denotes the ordered sample. Thanks to this special closed-form expression for $\widehat{\xi}_\tau^L$, we can rely for example on Serfling (1980) or Shorack (2000) for proving its consistency, asymptotic normality, and deriving the Berry-Esséen rate of uniform convergence $\mathrm{O}\left(n^{-1/2}\right)$.

Theorem 1 *For any index $\tau \in (0,1)$,*

(i) if $\mathbb{E}|Y|^\kappa < \infty$ for some $\kappa > 1$, then $\widehat{\xi}_\tau^L \xrightarrow{a.s.} \xi_\tau$ as $n \to \infty$.

(ii) if $\mathbb{E}|Y|^\kappa < \infty$ for some $\kappa > 2$, then $\sqrt{n}\left(\widehat{\xi}_\tau^L - \xi_\tau\right)$ has an asymptotic normal distribution with mean zero and variance $\sigma_\tau^2 = \int_0^1\int_0^1 (s \wedge t - s\,t)\, J_\tau(s)\, J_\tau(t)\, dF^{-1}(s)\, dF^{-1}(t)$.

(iii) If $\mathbb{E}|Y|^3 < \infty$, then $\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left(\frac{\sqrt{n}}{\sigma_\tau}\left(\widehat{\xi}_\tau^L - \xi_\tau\right) \leqslant t\right) - \Phi(t)\right| = \mathrm{O}\left(n^{-1/2}\right)$, for any $\tau \in [1-(1/2)^{1/3}, (1/2)^{1/3}]$, where Φ stands for the standard normal distribution function.

We note that, although mild moment assumptions suffice to ensure the validity of our results, the central case $\tau = 1/2$ shows that the ordinary central limit theorem fails to be a corollary to Theorem 1 (ii).

Alternatively, one may estimate ξ_τ from its formulation (7) as a probability-weighted moment. Estimating the expectation by the empirical moment and replacing F by its empirical version $\widehat{F}_n$ leads to the τth sample extremile

$$\widehat{\xi}_\tau^{LM} = \xi_\tau(\widehat{F}_n) = \frac{1}{n}\sum_{i=1}^{n} J_\tau\left(\frac{i}{n}\right) Y_{i,n}.$$

This estimator is another L-statistic, whose asymptotic properties are closely linked to those of $\widehat{\xi}_\tau^L$ since the finite differences built on the function K_τ in the estimator $\widehat{\xi}_\tau^L$ can be approximated by the derivative J_τ when n is large. Note further that an M-estimator $\widehat{\xi}_\tau^M$ of ξ_τ is provided by solving the empirical least squares problem $\min_\theta \sum_{i=1}^n J_\tau\left(\frac{i}{n}\right)|Y_{i,n} - \theta|^2$, which yields the closed form expression

$$\widehat{\xi}_\tau^M = \frac{\sum_{i=1}^n J_\tau\left(\frac{i}{n}\right) Y_{i,n}}{\sum_{i=1}^n J_\tau\left(\frac{i}{n}\right)} \equiv \frac{\widehat{\xi}_\tau^{LM}}{\frac{1}{n}\sum_{i=1}^n J_\tau\left(\frac{i}{n}\right)}.$$

Since the denominator $\frac{1}{n}\sum_{i=1}^n J_\tau\left(\frac{i}{n}\right)$ in the last equality converges to 1 as $n \to \infty$, the L-statistic $\widehat{\xi}_\tau^{LM}$ in the numerator can thus be viewed as a *Linearized* variant of the M-estimator $\widehat{\xi}_\tau^M$. Both $\widehat{\xi}_\tau^{LM}$ and $\widehat{\xi}_\tau^M$ are first-order equivalent with $\widehat{\xi}_\tau^L$. It is also easily seen that the asymptotic distributions of $\widehat{\xi}_\tau^L$, $\widehat{\xi}_\tau^{LM}$ and $\widehat{\xi}_\tau^M$ are identical (see Example A in Serfling (1980), pp. 277–278, and Example 1 in Shorack (1972)).

Of particular interest are the expected minimum $\xi_\tau = \mathbb{E}[\min(Y^1, \ldots, Y^{s(\tau)})]$ and expected maximum $\xi_\tau = \mathbb{E}[\max(Y^1, \ldots, Y^{r(\tau)})]$ which correspond respectively to the cases where $s(\tau)$ and $r(\tau)$ in (8) are positive integers. In these special cases, estimation of ξ_τ might be most conveniently based on unbiased estimates of the probability-weighted moments $M_{1,0,s}$ $(s = 0, 1, 2, \ldots)$ and $M_{1,r,0}$ $(r = 0, 1, 2, \ldots)$ given respectively by (see Landwehr *et al.*, 1979)

$$\widehat{M}_{1,0,s} = \frac{1}{n}\sum_{i=1}^{n-s}\left(\prod_{j=1}^{s}\frac{n-i+1-j}{n-j}\right) Y_{i,n} \quad \text{and} \quad \widehat{M}_{1,r,0} = \frac{1}{n}\sum_{i=r+1}^{n}\left(\prod_{j=1}^{r}\frac{i-j}{n-j}\right) Y_{i,n}.$$

Then, when $s(\tau)$ and $r(\tau)$ in (8) are positive integers, the statistic

$$\widehat{\xi}_{\tau}^{\mathrm{PWM}} = \begin{cases} s(\tau)\widehat{M}_{1,0,s(\tau)-1} & \text{for} \quad 0 < \tau \leqslant \frac{1}{2} \\ r(\tau)\widehat{M}_{1,r(\tau)-1,0} & \text{for} \quad \frac{1}{2} \leqslant \tau < 1 \end{cases}$$

is an unbiased estimate of the extremile ξ_τ. It is straightforward to see from the asymptotic normality of $\widehat{M}_{1,0,s(\tau)-1}$ and $\widehat{M}_{1,r(\tau)-1,0}$ (see Hosking *et al.*, 1985) that $\widehat{\xi}_{\tau}^{\mathrm{PWM}}$ converges to the same normal distribution as the estimators $\widehat{\xi}_{\tau}^{L}$, $\widehat{\xi}_{\tau}^{LM}$ and $\widehat{\xi}_{\tau}^{M}$ do.

Note also that when these empirical extremiles are used to estimate the same quantity as empirical quantiles, our simulation experiments in Supplement B.1 provide Monte-Carlo evidence that the extremile estimators are the most efficient in the case of usual short and light-tailed distributions. However, unlike sample quantiles (single order statistics), ordinary extremile estimators are not robust against heavy-tailed distributions. Yet, they can be 'robustified' by considering trimmed extremiles as discussed below.

3.2 Trimmed extremiles

From a theoretical robustness point of view, extremiles have a similar behavior to expectiles. Indeed, the influence function of the L-functional ξ_τ given by

$$x \mapsto \mathrm{IF}(x; \xi_\tau, F) = \int_{-\infty}^{\infty} \left[F(y) - \mathbb{1}(x \leqslant y)\right] J_\tau\left(F(y)\right) dy$$

(see *e.g.* Serfling, 1980, p.265) is not bounded, indicating the non-robustness of extremiles and their extreme sensitivity to the influence of isolated observations as is the case for expectiles, whilst quantiles have a finite gross-error sensitivity (see *e.g.* Abdous and Remillard, 1995, p.382). A robust alternative would be to use a trimmed L-functional in the sense of Shorack (2000, p.442). Namely, suppose the statistician wants to robustify the estimation procedure by specifying integer trimming numbers k_n and k_n' for which $k_n \wedge k_n' \to \infty$, while $\frac{k_n}{n} \vee \frac{k_n'}{n} \to 0$ with $k_n/k_n' \to 1$. The (k_n, k_n')-trimmed population and empirical τth extremiles can then be defined from (6) as, respectively:

$$\begin{aligned} \xi_\tau(k_n, k_n') &= \int_{\frac{k_n}{n}}^{1-\frac{k_n'}{n}} q_t \, dK_\tau(t), \\ \text{and} \quad \widehat{\xi}_\tau(k_n, k_n') &= \int_{\frac{k_n}{n}}^{1-\frac{k_n'}{n}} \widehat{q}_t \, dK_\tau(t) = \sum_{i=k_n+1}^{n-k_n'} \left\{ K_\tau\left(\frac{i}{n}\right) - K_\tau\left(\frac{i-1}{n}\right) \right\} Y_{i,n}. \end{aligned}$$

The asymptotic normality of $\widehat{\xi}_\tau(k_n, k_n')$, stated in Theorem 2, requires the regular variation of the function $\widetilde{V}_\tau : t \mapsto \widetilde{V}_\tau(t) := V_\tau(t,t)$ at 0 with a non-positive exponent, where

$$V_\tau(t,t') = \int_t^{1-t'} \int_t^{1-t'} (u \wedge v - u\,v)\, J_\tau(u)\, J_\tau(v)\, dF^{-1}(u)\, dF^{-1}(v), \quad t,t' \in [0,1].$$

Theorem 2 *Let k_n and k_n' such that $k_n \wedge k_n' \to \infty$, $\frac{k_n}{n} \vee \frac{k_n'}{n} \to 0$ and $k_n/k_n' \to 1$ as $n \to \infty$. If, for some $\beta \geqslant 0$,*

$$\lim_{t \downarrow 0} \widetilde{V}_\tau(ct)/\widetilde{V}_\tau(t) = c^{-\beta} \quad \textit{for each} \quad c > 0,$$

then $\sqrt{n}\left(\widehat{\xi}_\tau(k_n, k_n') - \xi_\tau(k_n, k_n')\right)/\sigma_\tau(k_n, k_n')$ has an asymptotic standard normal distribution, with $\sigma_\tau^2(k_n, k_n') = V_\tau(k_n/n, k_n'/n)$.

We have undertaken some simulation experiments in Supplement B.2 to evaluate the performance of the empirical trimmed extremile $\widehat{\xi}_\tau(k_n, k_n')$ in comparison with the sample quantile $\widehat{q}_\alpha$ when they estimate the same quantity $\xi_\tau(k_n, k_n') \equiv q_\alpha$. The accuracy of $\widehat{\xi}_\tau(k_n, k_n')$ appeared to be quite respectable with respect to the robust sample quantile.

3.3 Large extremiles

In the important maximum domain of attraction $\mathrm{DA}(\Phi_\gamma)$ of Pareto-type distributions with tail index $\gamma < 1$, which plays in particular a crucial role in financial and actuarial considerations (see *e.g.* Embrechts *et al.*, 1997 and Resnick, 2007), extremiles ξ_τ are more sensitive to the magnitude of heavy right tails than quantiles q_τ are, as $\tau \uparrow 1$. However, the use of sample counterparts such as, for instance, $\widehat{\xi}_\tau^L$ to estimate such large population extremiles may not be appropriate in the extreme region $\tau = \tau_n' \in [1 - 1/n, 1)$ for two reasons. First, in contrast to their population counterparts $\xi_{\tau_n'}$ and $q_{\tau_n'}$, $\widehat{\xi}_{\tau_n'}^L$ is not more spread than $\widehat{q}_{\tau_n'}$ since $\widehat{\xi}_{\tau_n'}^L < Y_{n,n} = \widehat{q}_{\tau_n'}$. Second, $\widehat{\xi}_{\tau_n'}^L$ is not consistent when estimating an extremile $\xi_{\tau_n'}$ that is to the right of all observations, for a τ_n' larger than $1 - 1/n$, which is typically the setting of extreme-value theory and related fields of application: examples include flood risk in Steenbergen *et al.* (2004), medical insurance large claims in Beirlant *et al.* (2004), operational bank losses in Embrechts and Puccetti (2007), loss returns of banks in the US market in Cai *et al.* (2015), the run-up height of ocean waves in de

Valk (2016), excess-of-loss risk measures on automobile insurance data in El Methni and Stupfler (2017), and fire risk for commercial firms in El Methni and Stupfler (2018), to name just a few. To remedy these problems, we construct below two extreme-value type estimators for $\xi_{\tau_n'}$ when $\tau_n' \uparrow 1$ at an arbitrary rate as $n \to \infty$. The first one is based on the use of the asymptotic connection between extremiles and quantiles, while the second one relies directly on asymmetric least squares estimation.

3.3.1 Estimation based on extreme quantiles

A first option to estimate the extremile $\xi_{\tau_n'}$ is by using its asymptotic equivalence $\xi_{\tau_n'} \sim q_{\tau_n'}\, \mathcal{G}(\gamma)$ obtained in Proposition 3 *(i)*, where $\mathcal{G}(s) := \Gamma(1-s)\{\log 2\}^s$, $s < 1$, and γ is the tail index of Y. This suggests to define the quantile-based estimator

$$\widehat{\xi}_{\tau_n'}^{Q,\star} := \widehat{q}_{\tau_n'}^{\star}\, \mathcal{G}(\widehat{\gamma}) \tag{9}$$

by substituting in suitable estimators $\widehat{q}_{\tau_n'}^{\star}$ of $q_{\tau_n'}$ and $\widehat{\gamma}$ of γ. Thenceforth, $\widehat{\xi}_{\tau_n'}^{Q,\star}$ and $\widehat{q}_{\tau_n'}^{\star}$ inherit the same property about the spread of their population counterparts. On the other hand, the naive sample maximum $Y_{n,n}$ will not be a consistent estimator of the extreme quantile $q_{\tau_n'}$. In order to extrapolate outside the range of the available observations, it is most efficient to use the traditional Weissman estimator (Weissman, 1978):

$$\widehat{q}_{\tau_n'}^{\star} \equiv \widehat{q}_{\tau_n'}^{\star}(\tau_n) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{q}_{\tau_n}, \quad \text{where} \;\; \widehat{q}_{\tau_n} := Y_{n-\lfloor n(1-\tau_n)\rfloor,n} \tag{10}$$

with τ_n being an intermediate sequence, *i.e.*, $\tau_n \to 1$ such that $n(1-\tau_n) \to \infty$ as $n \to \infty$. In practice, the intermediate quantile level τ_n is typically set to be $\tau_n = 1 - k/n$, where $k = k(n) \to \infty$ is a sequence of integers with $k(n)/n \to 0$ as $n \to \infty$. According to de Haan and Ferreira (2006, Corollary 1.2.10), the model assumption $F \in \mathrm{DA}(\Phi_\gamma)$ is equivalent to the first-order regular variation condition:

$$\lim_{t\to\infty} \frac{q_{1-(tx)^{-1}}}{q_{1-t^{-1}}} = x^{\gamma} \quad \text{for all} \quad x > 0. \tag{11}$$

Also, following de Haan and Ferreira (2006, Theorem 4.3.8), the asymptotic normality of the normalized Weissman quantile estimator $\widehat{q}_{\tau_n'}^{\star}/q_{\tau_n'}$ requires the extended regular variation assumption that for some second-order parameter $\rho \leqslant 0$ and an auxiliary function

A, with $A(t) \to 0$ as $t \to \infty$,

$$\lim_{t\to\infty} \frac{1}{A(t)} \left\{ \frac{q_{1-(tx)^{-1}}}{q_{1-t^{-1}}} - x^{\gamma} \right\} = x^{\gamma}\, \frac{x^{\rho}-1}{\rho} \quad \text{for all} \quad x > 0. \tag{12}$$

Here and in what follows, $(x^{\rho}-1)/\rho$ is to be understood as $\log x$ when $\rho = 0$. This classical second-order condition controls the rate of convergence in (11). Thorough discussions on the interpretation and the rationale behind this condition can be found in the monographs of Beirlant *et al.* (2004) and de Haan and Ferreira (2006), along with abundant examples of commonly used continuous distributions satisfying (12). Under this same extremal value condition, we show that $\widehat{\xi}^{Q,\star}_{\tau_n'}/\xi_{\tau_n'}$ converges in distribution as well, with the same scaling and limit distribution as $\widehat{q}^{\star}_{\tau_n'}/q_{\tau_n'}$. Our first step to this aim is to analyse the bias incurred by using the estimator $\widehat{\xi}^{Q,\star}_{\tau_n'}$, *i.e.* the approximation error in Proposition 3 *(i)*.

Proposition 4 *Suppose that $\mathbb{E}|Y| < \infty$, and that condition (12) holds with $\gamma < 1$. Recall the notation $\mathcal{G}(s) = \Gamma(1-s)\{\log 2\}^{s}$, for $s < 1$. Then, as $\tau \uparrow 1$:*

$$\frac{\xi_{\tau}}{q_{\tau}} - \mathcal{G}(\gamma) = C_1(\gamma,\rho)A((1-\tau)^{-1}) + C_2(\gamma)(1-\tau) + \mathrm{o}(A((1-\tau)^{-1})) + \mathrm{o}(1-\tau).$$

Here

$$C_1(\gamma,\rho) = \begin{cases} \dfrac{\mathcal{G}(\gamma+\rho)-\mathcal{G}(\gamma)}{\rho} & \text{if } \rho < 0 \\[2ex] (\log 2)^{\gamma} \int_0^{\infty} e^{-t}t^{-\gamma}\left(\log(\log 2) - \log(t)\right) dt & \text{otherwise,} \end{cases}$$

and

$$C_2(\gamma) = -\frac{1}{2}\mathcal{G}(\gamma) + \left[1 + \frac{\log 2}{2}\right]\mathcal{G}(\gamma-1) - \frac{\log 2}{2}\mathcal{G}(\gamma-2).$$

This result makes it possible to obtain the rate of convergence of the estimator $\widehat{\xi}^{Q,\star}_{\tau_n'}$ via standard extrapolation arguments in the spirit of Theorem 4.3.8 in de Haan and Ferreira (2006, p.138). The next result goes in this sense.

Theorem 3 *Suppose that $\mathbb{E}|Y| < \infty$ and:*

(i) condition (12) holds with $\gamma < 1$ and $\rho < 0$;

(ii) $\tau_n \to 1$, $n(1-\tau_n) \to \infty$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda \in \mathbb{R}$ as $n \to \infty$;

(iii) $\sqrt{n(1-\tau_n)}\,(\widehat{\gamma}-\gamma) \xrightarrow{d} Z$*, for a suitable estimator* $\widehat{\gamma}$ *of* γ*, where* Z *is a nondegenerate limiting random variable;*

(iv) $n(1-\tau_n') \to c < \infty$ *and* $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$.

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\xi}^{Q,\star}_{\tau_n'}}{\xi_{\tau_n'}}-1\right) \xrightarrow{d} Z \quad \text{as} \quad n\to\infty.$$

A simple and popular estimator of the tail index γ is the Hill (1975) estimator

$$\widehat{\gamma}_H = \frac{1}{k}\sum_{i=0}^{k-1}\left(\log Y_{n-i,n} - \log Y_{n-k,n}\right), \quad \text{where} \quad k = \lceil n(1-\tau_n)\rceil. \tag{13}$$

Here, $\lceil\cdot\rceil$ denotes the ceiling function. This is a natural estimator in the sense that it is the maximum likelihood estimator for Pareto models above a high threshold; other interpretations such as, for instance, $\widehat{\gamma}_H$ being the sample counterpart of the average log-excess can be found in Beirlant *et al.* (2004). Under conditions *(i)* and *(ii)* of Theorem 3, condition *(iii)* holds for $\widehat{\gamma}_H$ with Z having a normal distribution with mean $\lambda/(1-\rho)$ and variance γ^2, in view of Theorem 3.2.5 in de Haan and Ferreira (2006).

We shall discuss in Section 4.4.2 below a concrete application to medical insurance data using $\widehat{\xi}^{Q,\star}_{\tau_n'}$ in conjunction with the Hill estimator $\widehat{\gamma}_H$ of γ and the Weissman estimator $\widehat{q}^{\star}_{\tau_n'}$ of $q_{\tau_n'}$. Other efficient estimation methods of $q_{\tau_n'}$ and γ, such as the Peaks-Over-Threshold approach and maximum likelihood techniques, can be used as well in (9) to define $\widehat{\xi}^{Q,\star}_{\tau_n'}$. Beirlant *et al.* (2004) and de Haan and Ferreira (2006) give a nice review of these methods and an extensive bibliography.

3.3.2 Estimation based on intermediate extremiles

Here, we first consider estimating a high extremile ξ_{τ_n} of intermediate level τ_n satisfying $\tau_n \to 1$ and $n(1-\tau_n) \to \infty$ as $n \to \infty$. Then, we extrapolate the resulting estimate to the very extreme level τ_n' which approaches 1 at an arbitrarily fast rate in the sense that $n(1-\tau_n') \to c$, for some nonnegative constant c.

A natural estimator of the intermediate extremile ξ_{τ_n} follows from the solution of the asymmetric least squares problem as

$$\widehat{\xi}^M_{\tau_n} = \frac{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right) Y_{i,n}}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)} = \frac{\int_0^1 J_{\tau_n}\left(\frac{\lceil nt\rceil}{n}\right) Y_{\lceil nt\rceil,n}\, dt}{\frac{1}{n}\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)}.$$

This is the M-estimator $\widehat{\xi}^M_\tau$ presented in Section 3.1 when $\tau = \tau_n$. Alternative estimators can be defined by plugging $\tau = \tau_n$ in our L-estimator and LM-estimator, resulting in the following estimators:

$$\begin{aligned}
\widehat{\xi}^L_{\tau_n} &= \sum_{i=1}^n \left\{K_{\tau_n}\left(\frac{i}{n}\right) - K_{\tau_n}\left(\frac{i-1}{n}\right)\right\} Y_{i,n} = \int_0^1 J_{\tau_n}(t) Y_{\lceil nt\rceil,n}\, dt \\
\text{and}\quad \widehat{\xi}^{LM}_{\tau_n} &= \frac{1}{n}\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right) Y_{i,n} = \int_0^1 J_{\tau_n}\left(\frac{\lceil nt\rceil}{n}\right) Y_{\lceil nt\rceil,n}\, dt.
\end{aligned}$$

The asymptotic normality of these estimators requires, in addition to (12), to control the empirical quantile function $t \mapsto Y_{\lceil nt\rceil,n}$ in the central part of the distribution of Y. This is why we introduce the following extra condition:

($\mathcal{H}$) The support of Y is an interval, and on its interior F is twice differentiable with positive probability density function f and

$$\sup_{0<t<1} t(1-t)\frac{f'(q_t)}{[f(q_t)]^2} < \infty.$$

Condition ($\mathcal{H}$) makes it possible to approximate efficiently the empirical quantile process $t \mapsto Y_{\lceil nt\rceil,n}$ by a sequence of standard Brownian bridges in the central part of the interval $(0, 1)$, as well as in the far left tail due to the geometrically strong penalization of left tail quantiles by the weighting function J_{τ_n}. It can be found in Theorem 6.2.1 of Csörgő and Horváth (1993) and Proposition 2.4.9 in de Haan and Ferreira (2006), among others. Finally, to elaborate our asymmetric least squares estimation based on the highest values in the sample, we shall slightly strengthen the condition that $F \in \mathrm{DA}(\Phi_\gamma)$ by assuming:

$$\lim_{t\to+\infty} t\frac{f(t)}{1-F(t)} = \gamma. \tag{14}$$

This von Mises condition indeed implies the model assumption $F \in \mathrm{DA}(\Phi_\gamma)$ as shown in Theorem 1.11 of de Haan and Ferreira (2006, p.17) and Proposition 2.1 of Beirlant *et al.* (2004, p.60). While the necessary and sufficient condition $F \in \mathrm{DA}(\Phi_\gamma)$ for F to be a Pareto-type distribution is sometimes difficult to verify, the sufficient von Mises condition (14) may be more helpful for absolutely continuous distributions. All commonly used Pareto-type distributions satisfy (14), see Beirlant *et al.* (2004, p.60). Next, under these assumptions, we unravel the common limit distribution of the normalized estimators $\widehat{\xi}^L_{\tau_n}/\xi_{\tau_n}$, $\widehat{\xi}^{LM}_{\tau_n}/\xi_{\tau_n}$ and $\widehat{\xi}^M_{\tau_n}/\xi_{\tau_n}$.

Theorem 4 *Suppose that* $\mathbb{E}|Y| < \infty$ *and:*

- *conditions* (12), *($\mathcal{H}$) and* (14) *hold with* $\gamma < 1/2$*;*
- *the sequence* $\tau_n \uparrow 1$ *is such that* $n(1-\tau_n) \to \infty$ *and* $\sqrt{n(1-\tau_n)}\, A((1-\tau_n)^{-1}) = \mathrm{O}(1)$.

Then, if $\widehat{\xi}_{\tau_n}$ *is either* $\widehat{\xi}^L_{\tau_n}$, $\widehat{\xi}^{LM}_{\tau_n}$ *or* $\widehat{\xi}^M_{\tau_n}$, *we have*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \frac{\gamma\sqrt{\log 2}}{\Gamma(1-\gamma)} \int_0^\infty e^{-s} s^{-\gamma-1} W(s) ds$$

where W is a standard Brownian motion. In other words, the above limit distribution is Gaussian centered with variance

$$V(\gamma) = \left(\frac{\gamma}{\Gamma(1-\gamma)}\right)^2 (\log 2) \int_0^\infty \int_0^\infty e^{-s-t} (st)^{-\gamma-1} (s \wedge t)\, ds\, dt.$$

Turning now to the extreme level τ_n', we have under the model assumption of heavy-tailed distributions $F \in \mathrm{DA}(\Phi_\gamma)$ that

$$\frac{\xi_{\tau_n'}}{q_{\tau_n'}} \sim \mathcal{G}(\gamma) \sim \frac{\xi_{\tau_n}}{q_{\tau_n}} \quad \text{and hence} \quad \frac{\xi_{\tau_n'}}{\xi_{\tau_n}} \sim \frac{q_{\tau_n'}}{q_{\tau_n}} \quad \text{as} \quad n \to \infty,$$

in view of Proposition 3 *(i)*. On the other hand, the assumption $F \in \mathrm{DA}(\Phi_\gamma)$ or equivalently the first-order regular variation condition (11) leads to the quantile approximation

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

This final extrapolation motivates the alternative purely extremile-based estimator

$$\widehat{\xi}^{M,\star}_{\tau_n'} := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}} \widehat{\xi}^M_{\tau_n}. \tag{15}$$

This is still a Weissman-type device which, in contrast to $\widehat{\xi}^{Q,\star}_{\tau'_n}$ in (9), relies crucially on our intermediate asymmetric least squares estimator $\widehat{\xi}^{M}_{\tau_n}$. Another option would be to replace the intermediate M-estimator $\widehat{\xi}^{M}_{\tau_n}$ in (15) by either the L-estimator $\widehat{\xi}^{L}_{\tau_n}$ or the LM-estimator $\widehat{\xi}^{LM}_{\tau_n}$. The resulting extrapolated L- and LM-estimators can easily be shown to share the asymptotic properties of $\widehat{\xi}^{M,\star}_{\tau'_n}$ stated below. However, experiments with simulated data indicate that these estimators perform no better than $\widehat{\xi}^{M,\star}_{\tau'_n}$. We therefore restrict our attention to the latter estimator.

Theorem 5 *Suppose that* $\mathbb{E}|Y| < \infty$ *and:*

- *conditions* (12), *($\mathcal{H}$) and* (14) *hold with* $\gamma < 1/2$ *and* $\rho < 0$;
- *the sequence* $\tau_n \uparrow 1$ *is such that* $n(1-\tau_n) \to \infty$ *and*

$$\sqrt{n(1-\tau_n)}\max(A((1-\tau_n)^{-1}), 1-\tau_n) = \mathrm{O}(1);$$

- $\sqrt{n(1-\tau_n)}\,(\widehat{\gamma}-\gamma) \xrightarrow{d} Z$, *for a suitable estimator* $\widehat{\gamma}$ *of* γ, *where* Z *is a nondegenerate limiting random variable;*
- $n(1-\tau'_n) \to c < \infty$ *and* $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau'_n)] \to \infty$.

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\widehat{\xi}^{M,\star}_{\tau'_n}}{\xi_{\tau'_n}} - 1\right) \xrightarrow{d} Z \quad \text{as} \quad n \to \infty.$$

Like the quantile-based estimator $\widehat{\xi}^{Q,\star}_{\tau'_n}$, the extrapolated M-estimator $\widehat{\xi}^{M,\star}_{\tau'_n}$ inherits the limit distribution of $\widehat{\gamma}$ with a slightly slower rate of convergence.

3.3.3 Some simulation evidence

To investigate the finite sample performance of the two rival estimators $\widehat{\xi}^{Q,\star}_{\tau'_n}$ in (9) and $\widehat{\xi}^{M,\star}_{\tau'_n}$ in (15), we have considered simulated samples from the Student $t_{1/\gamma}$ distribution, the Pareto distribution $F(y) = 1 - y^{-1/\gamma}$, $y \geqslant 1$, and the Fréchet distribution $F(y) = e^{-y^{-1/\gamma}}$, $y > 0$. All the experiments have tail index $\gamma \in \{1/4, 1/3, 2/5\}$, sample size $n = 1000$, extreme level $\tau'_n = 1 - 1/n$, and intermediate level $\tau_n = 1 - k/n$, where the integer k can actually be viewed as the effective sample size for tail extrapolation. We used in all our simulations the Hill estimator $\widehat{\gamma}_H$ described in (13) to estimate γ.

Figures 1, 2 and 3 give the relative Mean-Squared Error (MSE) in top panels and bias in bottom panels of the estimators $\widehat{\xi}^{Q,\star}_{\tau'_n}$ and $\widehat{\xi}^{M,\star}_{\tau'_n}$, computed over $10,000$ replications, for $\gamma = 1/4, 1/3$ and $2/5$, respectively. In terms of MSE, it may be seen that $\widehat{\xi}^{M,\star}_{\tau'_n}$ has a very similar behavior to $\widehat{\xi}^{Q,\star}_{\tau'_n}$ in the Fréchet model, but is clearly the winner in both Pareto and Student models, for all values of γ. It may also be seen that $\widehat{\xi}^{M,\star}_{\tau'_n}$ is superior in terms of bias in all scenarios. Although either might be used in practice, we would therefore have a particular preference for the purely extremile-based estimator $\widehat{\xi}^{M,\star}_{\tau'_n}$.

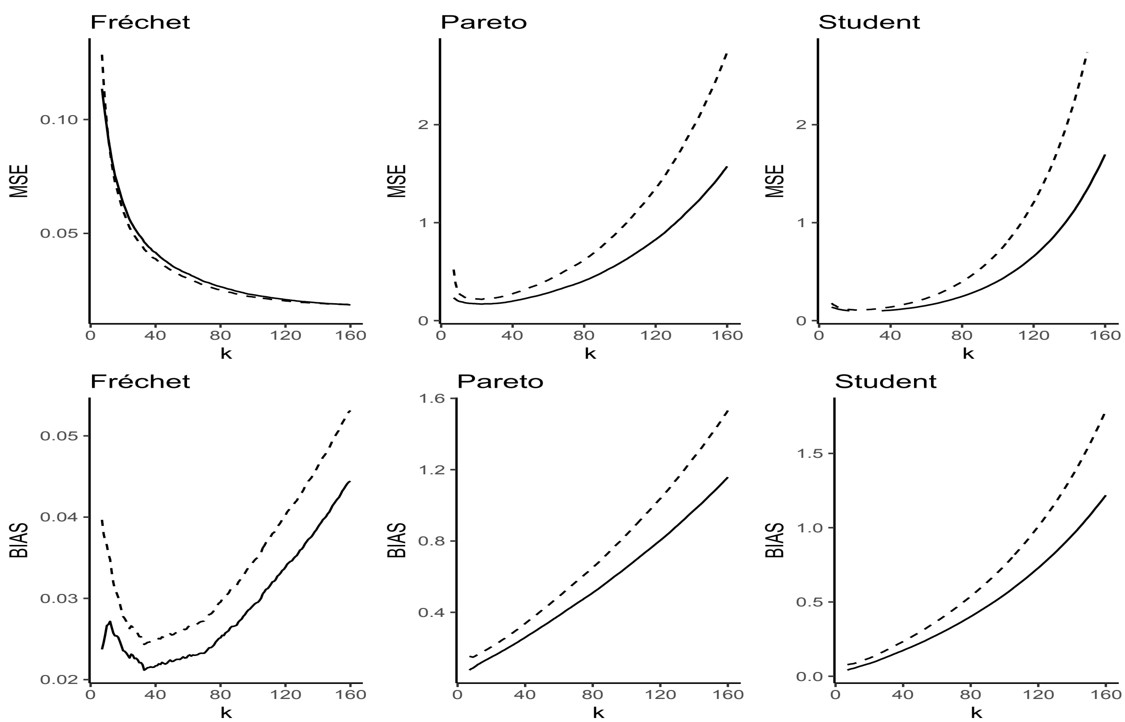

Figure 1: *Results for $\gamma = 1/4$ – MSE estimates (top panels) and bias estimates (bottom panels) of $\widehat{\xi}^{M,\star}_{\tau'_n}/\xi_{\tau'_n}$ (solid line) and $\widehat{\xi}^{Q,\star}_{\tau'_n}/\xi_{\tau'_n}$ (dashed line) as functions of k, for the Fréchet, Pareto and Student distributions, respectively, from left to right.*

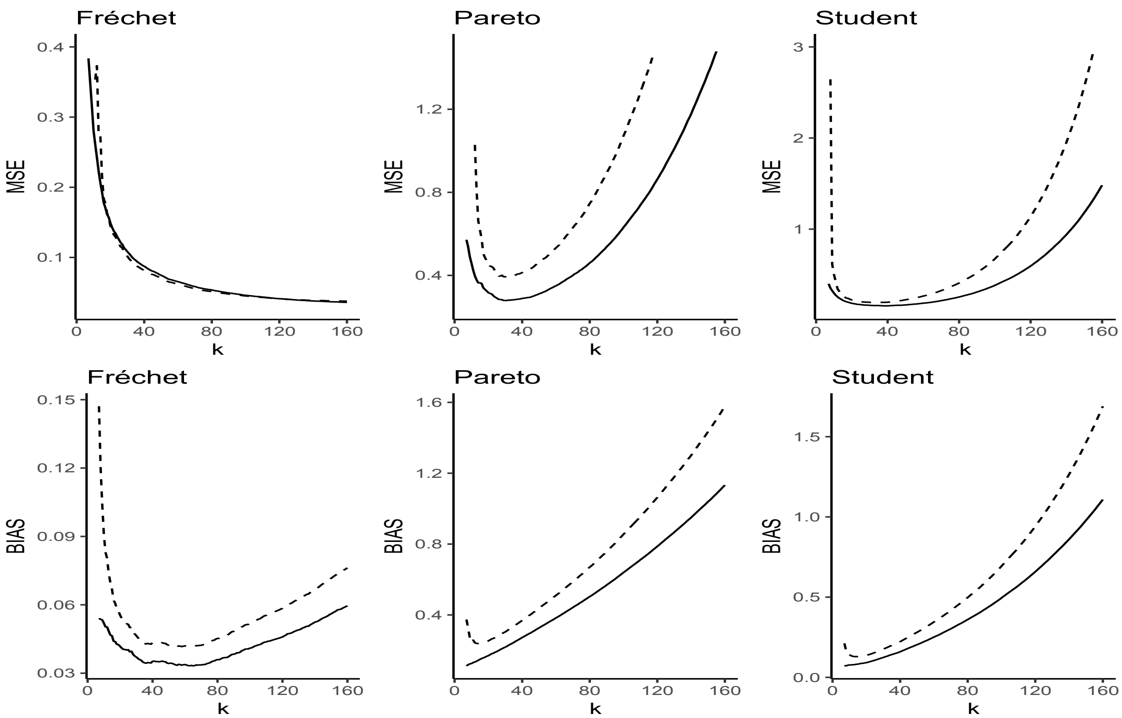

Figure 2: *As in Figure 1 with* $\gamma = 1/3$.

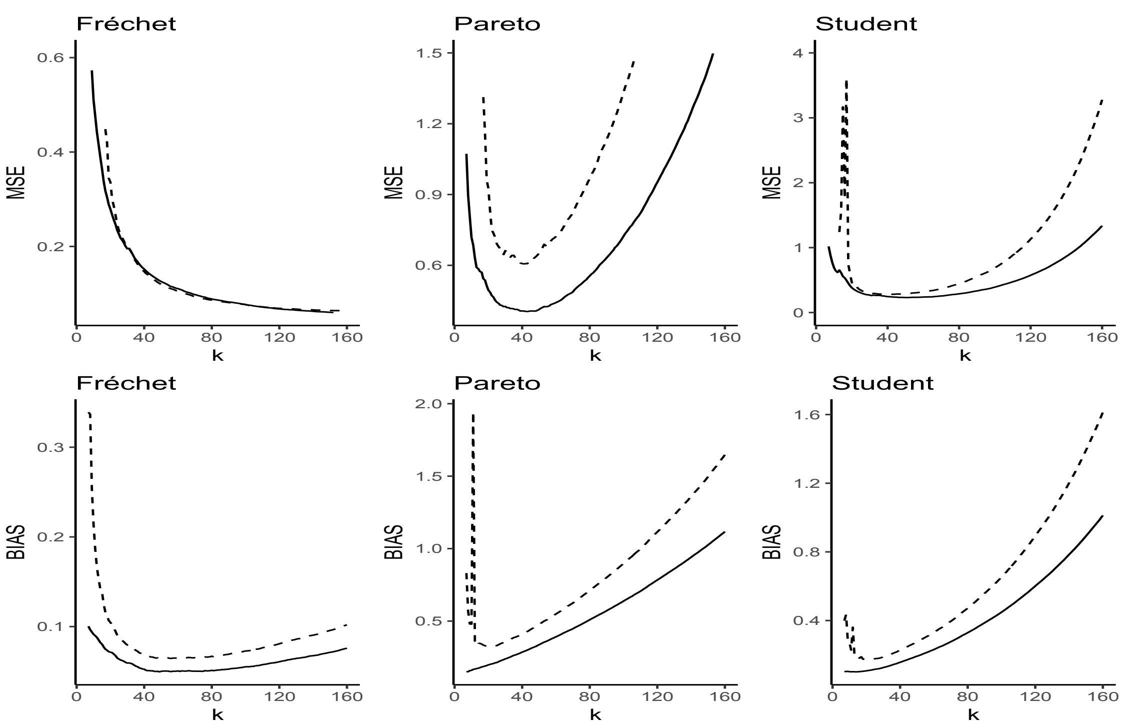

Figure 3: *As in Figure 1 with* $\gamma = 2/5$.

4 Extremiles as risk measures

A very important actuarial and financial problem involves quantifying the "riskiness" implied by the distribution of a non-negative loss variable or a real-valued profit-loss variable. Greater variability of the random variable under consideration and particularly a heavier tail of its distribution necessitate a higher capital reserve for portfolios or price of the insurance risk. A leading risk measure in banking and other financial institutions is Value at Risk (VaR) with a confidence level $\tau \in (0,1)$. It is defined as the τth quantile q_τ of the non-negative loss distribution with τ being close to one, and as $-q_\tau$ for the real-valued profit-loss distribution with τ being close to zero. An undesirable property of the VaR measure is that it is insensitive to the magnitude of extreme losses since it only depends on the frequency of tail losses and not on their values. A solution to this problem is to use the conditional tail mean $\nu_\tau = \mathbb{E}[Y|Y > q_\tau]$, for non-negative loss distributions, and $\nu_\tau = -\mathbb{E}[Y|Y < q_\tau]$ for real-valued profit-loss distributions [Rockafellar and Uryasev (2002)]. An often convenient, equivalent definition of ν_τ is as a spectral risk measure called Expected Shortfall (ES) [Acerbi (2002)]. Practitioners who are more concerned with the risk exposure to a catastrophic event, that may wipe out an investment in terms of the size of potential losses, favor the use of the ES ν_τ. The latter may, however, be criticized for being too pessimistic since it only depends on the tail event. The quantile-based VaR and ES may therefore be considered as too liberal or too conservative, depending on the tail shape of the underlying distribution. The use of expectiles as an alternative measure of risk has recently attracted a lot of interest thanks to their asymmetric least squares nature [see, *e.g.*, Kuan *et al.* (2009) and Daouia *et al.* (2018)]. This proposal was criticized though for its lack of comonotonic additivity [Acerbi and Szekely (2014)].

We will investigate in the sequel the properties of the extremile functional from the point of view of the axiomatic theory of risk measures. The discussion in Section 4.1 pertains to non-negative loss distributions, while Section 4.2 is concerned with real-valued profit-loss random variables. Section 4.3 examines the connection between tail extremiles, expectiles and ES. Section 4.4 considers two motivating examples on inflation-adjusted hurricanes and medical insurance data.

4.1 Coherency, regularity and pessimism

Given that extremiles depend on both the tail losses and their probability, they steer an advantageous middle course between the potential excessive optimism or pessimism of the VaR and ES. Taking ξ_τ as a margin (amount of capital as a hedge against extreme risks), a larger ξ_τ is then a more prudential margin requirement and results in larger τ. As such, the index τ reflects the level of prudentiality to be typically set by regulators and/or the management level. When $\tau = 1/2$, ξ_τ is the net expected loss. When $\tau < 1/2$, the values of the risk measure ξ_τ are less than $\mathbb{E}(Y)$, which is not appropriate when $Y \geqslant 0$ is a loss random variable. Hence, a natural range for the security level τ is given by $[1/2, 1]$. The use of ξ_τ as a proper risk measure can be justified further as follows. In actuarial terms, a risk measure $\Pi[\cdot]$ is defined as a mapping from the set of all bounded loss random variables to the set of all non-negative real numbers, with $\Pi[Y]$ being the price associated with a loss variable $Y \geqslant 0$. A variety of coherent risk measures have been specified in the literature as the expected value of Y with respect to a distortion function $g(\cdot)$, that is,

$$\Pi[Y] = \int_0^\infty g\left(1 - F(y)\right) dy, \tag{16}$$

where g is a nondecreasing and concave function with $g(0) = 0$ and $g(1) = 1$. The class of such measures is referred to as Wang's (1996) class of distortion risk measures. The review article by Wirch and Hardy (1999) discusses a number of important distortion functions including the Proportional Hazard transform $g(t) = \mathrm{PH}_\lambda(t) := t^\lambda$, for the so-called risk-aversion index $\lambda \in (0, 1]$, and the Dual Power transform $g(t) = \mathrm{DP}_r(t) := 1 - (1 - t)^r$ for $r \geqslant 1$. The ES can be expressed in terms of the distortion function $ES_\tau(t) := (t/(1-\tau))\mathbb{I}(0 \leqslant t < 1 - \tau) + \mathbb{I}(1 - \tau \leqslant t \leqslant 1)$. While the quantile-VaR and ES use only a small part of the loss distribution, both PH and DP distortion functions utilize the whole loss distribution and are more reliable for the purpose of differentiating between more and less risky distributions.

Extremiles clearly belong to the class of risk measures of the form (16), with the alternative concave distortion function

$$g_\tau(t) := 1 - K_\tau(1 - t) = 1 - (1 - t)^{r(\tau)}, \quad \frac{1}{2} \leqslant \tau \leqslant 1.$$

It turns out that this function is very closely related to the DP transform in the sense that $g_\tau = \mathrm{DP}_{r(\tau)}$. Its formulation in terms of the asymmetry parameter $\tau \in [1/2, 1]$ allows, like the tail probability τ in the VaR q_τ and ES $\nu_\tau = \mathbb{E}[Y|Y > q_\tau]$, for better interpretability compared to the distortion parameter $r \in [1, \infty)$ in the DP transform. Let us highlight in particular that the asymmetric least squares formulation makes the comparison of extremiles (and, incidentally, of DP transforms) with both VaR and ES easier and more insightful than the purely distortion-based interpretation: being least squares analogues of quantiles, extremiles rely on the distance to observations, making thus more efficient use of the available data, whereas quantiles only use the information on whether an observation is below or above the predictor. Also, extremiles depend by construction on both the tail losses and their probability, while VaR only depends on the frequency of tail losses and ES only depends on the tail event.

On the other hand, the extremile distortion function g_τ results in a more widely applicable risk measure than the popular PH-measure, at least in the following respect: the empirical estimators of both measures allow one to determine the price of an insurance risk without recourse to any fitting of a parametric model. However, as shown by Jones and Zitikis (2003), the asymptotic normality of the empirical PH-measure, $\int_0^\infty (1 - \widehat{F}_Y(y))^\lambda dy$, does not cover the range $0 < \lambda \leqslant 1/2$, and necessitates the assumption that $\mathbb{E}|Y|^\kappa < \infty$, for some $\kappa > 1/(\lambda - 1/2)$, when $1/2 < \lambda \leqslant 1$. Hence, the number κ of required finite moments tends to $+\infty$ as $\lambda \downarrow 1/2$. By contrast, as shown in Theorem 1, inference on ξ_τ is applicable for any index τ, under the weaker assumption that $\mathbb{E}|Y|^\kappa < \infty$ for some $\kappa > 2$.

It is straightforward to see that the extremile-based risk measure $\Pi_\tau[Y] := \xi_{Y,\tau}$, for $\tau \in (1/2, 1)$, satisfies the following two natural properties related to risk loading:

(A1) Positive loading and no ripoff: $\mathbb{E}(Y) < \Pi_\tau[Y] < y_u$, $\lim_{\tau\to 1/2} \Pi_\tau[Y] = \mathbb{E}(Y)$, and $\lim_{\tau\to 1} \Pi_\tau[Y] = y_u$.

(A2) No unjustified risk-loading: if $P(Y = b) = 1$ for some constant b then $\Pi_\tau[Y] = b$.

More importantly, the extremile risk measure also satisfies the requirements for being a coherent risk measure, in the sense of the influential paper by Artzner *et al.* (1999):

(A3) Translation and scale invariance: $\Pi_\tau[a + bY] = a + b\Pi_\tau[Y]$, for any $a \in \mathbb{R}$ and $b > 0$.

(A4) Subadditivity: $\Pi_\tau[Y + \widetilde{Y}] \leqslant \Pi_\tau[Y] + \Pi_\tau[\widetilde{Y}]$, for any loss variables Y and $\widetilde{Y}$.

(A5) Preserving of stochastic order: $\Pi_\tau[Y] \leqslant \Pi_\tau[\widetilde{Y}]$ if $Y \leqslant \widetilde{Y}$ with probability 1.

Finally, $\Pi_\tau[\cdot]$ is a regular risk measure since it fulfills the following additional fundamental conditions [as imposed *e.g.* in Definition 5 of Bassett *et al.* (2004)]:

(A6) Law invariance: $\Pi_\tau[Y] = \Pi_\tau[\widetilde{Y}]$ if Y and $\widetilde{Y}$ have the same distribution.

(A7) Comonotonic additivity: $\Pi_\tau[Y + \widetilde{Y}] = \Pi_\tau[Y] + \Pi_\tau[\widetilde{Y}]$ if Y and $\widetilde{Y}$ are comonotone.

According to Bassett *et al.* (2004, Theorem 1), being coherent and regular, $\Pi_\tau[\cdot]$ is then a pessimistic risk measure in the sense that the corresponding distortion function g_τ acts to depress the implicit likelihood of the most favorable outcomes, and to accentuate the likelihood of the least favorable ones (see Definition 4 in Bassett *et al.* (2004) for a formal characterization of pessimistic risk measures). As such, the index τ becomes a natural measure of the degree of pessimism. Note also that equation (6) makes $\Pi_\tau[\cdot]$ a spectral risk measure [Acerbi (2002)].

4.2 Real-valued profit-loss random variables

The extremile-based risk measure $\Pi_\tau[Y]$ can still be defined when Y is an asset return, which can take any real value, by using the more general expression (5), that is,

$$\xi_\tau = \int_{-\infty}^{0} \left[g_\tau\left(1 - F(y)\right) - 1\right] dy + \int_{0}^{\infty} g_\tau\left(1 - F(y)\right) dy,$$

or its various equivalent formulations described in Sections 2.1-2.2, where

$$g_\tau(t) := t^{s(\tau)}\, \mathbb{I}(0 \leqslant \tau < 1/2) + \left[1 - (1-t)^{r(\tau)}\right] \mathbb{I}(1/2 \leqslant \tau \leqslant 1).$$

A number of studies, including Jones and Zitikis (2003), Bassett *et al.* (2004) and the references therein, have recognized the usefulness of such Choquet integrals in actuarial and financial applications as well as in the area of measuring economic inequality. The coherency axioms [Translation invariance, Monotonicity, Subadditivity, and Positive homogeneity] of the risk measure $\Pi_\tau[Y] := -\xi_{Y,\tau}$ for asset returns can easily be checked by making use of the alternative formulation (7) of ξ_τ as a probability-weighted moment. The key argument is that the special weight-generating function $J_\tau(\cdot)$ is an admissible risk spectrum [Acerbi (2002)].

Proposition 5 *For the profit-loss Y of a given portfolio and for any $0 < \tau < 1/2$,*

$$\Pi_\tau[Y] = -\xi_{Y,\tau} \equiv -\mathbb{E}\left[Y J_\tau\left(F(Y)\right)\right]$$

is a coherent spectral risk measure.

In the special case where the power $s(\tau)$ in the distortion function g_τ is an integer, we recover the particularly attractive and very intuitive interpretation of the so-called *MINVAR* risk measure introduced in Cherny and Madan (2009), namely the negative of the expected minimum $\Pi_\tau[Y] \equiv -\mathbb{E}[\min(Y^1, \ldots, Y^{s(\tau)})]$ of $s(\tau)$ independent observations from Y (see also Föllmer and Knispel, 2013).

Despite their statistical virtues and all their nice axiomatic properties as spectral risk measures and concave distortion risk measures, extremiles have an apparent limitation when applied to distributions with infinite mean. This should not be considered to be a serious disadvantage however, at least in financial and actuarial applications, since the definition of a coherent risk measure for distributions with an infinite first moment is not clear, see the discussion in Section 3 of Nešlehová *et al.* (2006). Whether operational risk models in which losses have an infinite mean make sense in the first place has also recently been questioned by Cirillo and Taleb (2016).

4.3 Connection between extremiles, expectiles and ES

We consider both cases of non-negative loss distributions and real-valued profit-loss distributions with heavy right tails. In the case of a profit-loss distribution, the financial position Y stands for the negative of the asset return so that the right-tail of F corresponds to the negative of extreme losses. For most studies in risk management, it has been found that Pareto-type distributions $F \in \mathrm{DA}(\Phi_\gamma)$, with $\gamma < 1/2$, describe quite well the tail structure of actuarial and financial data. We refer for instance to the R package 'CASdatasets' which contains a large variety of data examples where realized values of γ often vary between 1/4 and 1/2. Accordingly, as established in Proposition 3, the corresponding extremile-based risk measure ξ_τ is more pessimistic, from the risk management viewpoint, than the traditional quantile-VaR q_τ for large values of τ. Next, we show that

ξ_τ is more pessimistic than the expectile-based VaR e_τ as well, in the standard case of finite-variance distributions, but it is always less pessimistic than the ES ν_τ, for large τ. The key ingredients to show this are the following asymptotic connections.

Proposition 6 *Suppose that* $\mathbb{E}|Y| < \infty$ *and* $F \in DA(\Phi_\gamma)$ *with* $0 < \gamma < 1$*. Then, as* $\tau \to 1$*,*

$$\frac{\xi_\tau}{e_\tau} \sim \Gamma(1-\gamma)\{(\gamma^{-1}-1)\log 2\}^\gamma \quad \textit{and} \quad \frac{\xi_\tau}{\nu_\tau} \sim \Gamma(2-\gamma)\{\log 2\}^\gamma.$$

We now show how Proposition 6 entails that high extremiles are more conservative than high expectiles at the same level, when $\gamma \in (0, 1/2)$. The idea is simply to note that for $\gamma \in (0, 1/2)$, we have $\gamma^{-1} - 1 > 1$, and therefore $\Gamma(1-\gamma)\{(\gamma^{-1}-1)\log 2\}^\gamma > \Gamma(1-\gamma)\{\log 2\}^\gamma > 1$, establishing thus that $\xi_\tau > e_\tau$ for τ large enough provided the underlying distribution has a finite variance. This is, however, no longer valid for the heaviest tails since the proportionality constant $\Gamma(1-\gamma)\{(\gamma^{-1}-1)\log 2\}^\gamma$ tends to $\log 2 < 1$ as $\gamma \to 1$. More precisely, a numerical study shows that high extremiles shall be more conservative than high expectiles if and only if $0 < \gamma < \gamma_0$, with $\gamma_0 \approx 0.8729$. By contrast, that high extremiles are always less conservative than their ES analogues is a consequence of the inequality $\Gamma(2-\gamma)\{\log 2\}^\gamma < \{\log 2\}^\gamma < 1$, for all $\gamma \in (0, 1)$.

Finally, we would like to stress why our finding in Proposition 6 that $q_\tau < \xi_\tau < \nu_\tau$, as $\tau \to 1$, is not a contradiction to Theorem 13 in Delbaen (2002). This theorem states that any coherent, law-invariant risk measure satisfying the Fatou property and greater than or equal to q_τ (for every $Y \in L^\infty$) must also be greater than or equal to ν_τ. The extremile ξ_τ defines a law-invariant coherent, and hence convex risk measure, satisfying thus the Fatou property [see Theorem 2.1 in Jouini *et al.* (2006) and Theorem 2.2 in Tsukahara (2009)]. However, the key condition of Delbaen's theorem that $q_\tau \leqslant \xi_\tau$ is not fulfilled for short-tailed distributions with a large $|\gamma|$, as discussed below our Proposition 3.

4.4 Data examples

To illustrate the ideas discussed above, we explore in this section two real data examples.

4.4.1 Trended hurricane losses

We first consider a dataset on trended or, equivalently, inflation-adjusted (to 1981 using the U.S. Residential Construction Index) hurricane losses that occurred between 1949 and 1980. Figure 4 (left panel) displays the histogram and scatterplot of the recorded $n = 35$ trended hurricane losses in excess of \$5 Million (in units of \$1,000). An analysis of this data set conducted from a central point of view is presented in Jones and Zitikis (2003); our focus here is rather on the right tail of the observations. We treat the 35 amounts as the outcomes of i.i.d. non-negative loss random variables $Y_1, \ldots, Y_{35}$. The corresponding sample mean and standard deviation are 199,900 and 325,807, respectively. The empirical estimates of the expected shortfall, quantile, expectile and extremile risk measures are graphed in Figure 4 (right panel) against the security level τ. The first impression to be gained from this figure is the lack of smoothness and stability of both sample quantile-VaR $\widehat{q}_\tau$ (blue) and ES $\widehat{\nu}_\tau$ (orange): their discreteness as piecewise constant functions of the argument τ is a serious defect, especially in the important upper tail. Indeed, a small change in τ can trigger a (severe) jump in the values of the estimated VaR and ES. Moreover, the fact that the "steps" result in the same or similar measures for significantly different risk levels is itself risky. By contrast, the sample extremile $\widehat{\xi}_\tau = \widehat{\xi}_\tau^L$ (red) and expectile $\widehat{e}_\tau$ (green) have the benefit to be very stable and to change continuously and increasingly without recourse to any smoothness procedure.

When comparing the four estimated risk measures at the same level τ, it can be seen that the ES $\widehat{\nu}_\tau$, in orange, is much larger and hence is more conservative than the extremile $\widehat{\xi}_\tau$ in red. In contrast, the quantile $\widehat{q}_\tau$, in blue, remains less alert to extreme risks than $\widehat{\xi}_\tau$ until it breaks down at $\tau = (n-1)/n = 0.9714$. Thenceforth, for all $\tau > 0.9714$, $\widehat{q}_\tau$ becomes identical to $\widehat{\nu}_\tau$ which in turn coincides with the maximum catastrophic loss $Y_{n,n} = 1{,}633{,}000$, whereas $\widehat{\xi}_\tau$ provides less pessimistic risk measure values. Finally, although the expectile $\widehat{e}_\tau$, in green, exhibits a smooth evolution, it diverges from $\widehat{\xi}_\tau$ in the region $\tau \in [0.8, 0.975]$ and becomes less alert to infrequent disasters. The extremile in red seems to afford a middle course between the pessimistic ES in orange and the optimistic expectile in green.

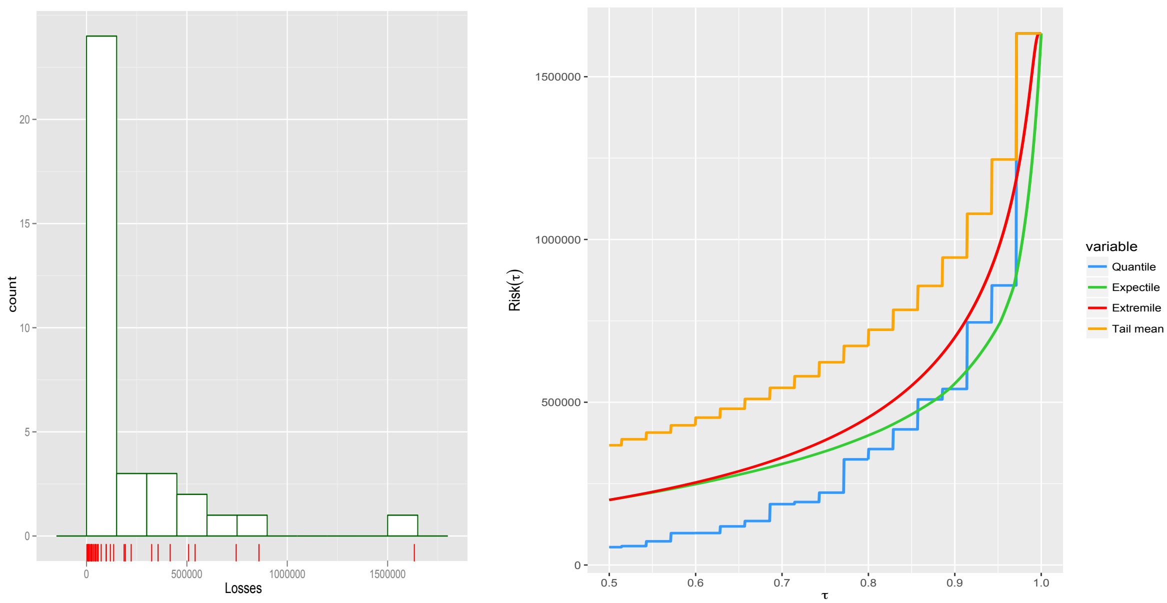

Figure 4: *Trended Hurricane Losses data. (Left) Histogram and scatterplot; (Right) Empirical expected-shortfall, extremile, expectile and quantile τ-risk measures.*

Of particular interest is the deviation between the estimated extremile $\widehat{\xi}_\tau$ and quantile $\widehat{q}_\tau$. Being the estimates, respectively, of the mean and the median of the same asymmetric distribution $K_\tau(F)$ of the transformation $\phi_\tau(Y)$, a significant deviation between their values diagnoses a heavier right tail of Y. Thereby the comparison above with the same level τ may be viewed as an explanatory tool for quantifying the “riskiness” implied by the distribution of Y, rather than as a method for final analysis, especially since we know that non-extrapolated sample extremiles and quantiles will be inconsistent in the typical extreme range $\tau \geqslant 1 - 1/n$. Moreover, moving from the standard quantile-based VaR to the extremile-based risk measure requires in practice the use of different asymmetry levels τ, say, τ_q and τ_ξ. For example, if we consider a pre-specified extremile level τ_ξ, then for the comparison between the two estimators $\widehat{\xi}_{\tau_\xi}$ and $\widehat{q}_{\tau_q}$ to be insightful, they must estimate the same theoretical risk measure $\xi_{\tau_\xi} \equiv q_{\tau_q}$. This latter equality implies that $\tau_q = F(\xi_{\tau_\xi})$, whose empirical counterpart is $\widehat{\tau}_q = \widehat{F}_n(\widehat{\xi}_{\tau_\xi})$. This leads to a rival composite quantile estimator $\widehat{q}_{\widehat{\tau}_q}$ of the direct extremile estimator $\widehat{\xi}_{\tau_\xi}$. As is to be expected, the latter estimator of an L^2-nature appears to be more alert to big losses than its L^1 antecedent $\widehat{q}_{\widehat{\tau}_q}$, for any pre-specified level τ_ξ.

4.4.2 Medical insurance data

The Society of Actuaries 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) and Daouia *et al.* (2018), we deal here only with the 75,789 claims for 1991. The scatterplot and histogram shown in Figure 5 (a) give evidence of a considerable right-skewness. Also, it has been found in both Beirlant *et al.* (2004, p.123) and Daouia *et al.* (2018) that the loss severity distribution is heavy-tailed with Hill's estimates $\widehat{\gamma}_H$ around 0.359, where $\widehat{\gamma}_H = \frac{1}{k}\sum_{i=0}^{k-1}(\log Y_{n-i,n} - \log Y_{n-k,n})$ as described in (13) with $k = \lceil n(1-\tau_n)\rceil$ being an intermediate sequence of integers. Then nothing guarantees that the future does not hold some unexpected higher claim amounts. A usual way to assess the magnitude of such infrequent claim amounts from the extreme-value perspective is by using the Weissman quantile estimate $\widehat{q}^{\star}_{\tau_n'} = Y_{n-k,n}\,(k/np_n)^{\widehat{\gamma}_H}$, as described in (10) with $\tau_n' = 1-p_n$. By construction, this tail estimate is expected to be capable of extrapolating outside the range of the available observations when $p_n < 1/n$. Following Beirlant *et al.* (2004, p.123), insurance companies typically are interested in $p_n = 1/100{,}000 < 1/n$, that is, in an estimate of the claim amount that will be exceeded (on average) only once in 100,000 cases. Figure 5 (b) shows the estimate $\widehat{q}^{\star}_{\tau_n'}$ against the sample fraction k (solid black curve). In practice, a commonly used heuristic approach for choosing an appropriate estimate is to pick out a suitable k corresponding to the first stable part of the plot [see, *e.g.*, Section 3 in de Haan and Ferreira (2006)]. A stable region appears for k from 150 up to 500, leading to estimates between 3.73 and 4.12 million, with an averaged estimate around 3.90 million. This tail risk estimate does not exceed the sample maximum $Y_{n,n} = 4.51$ million (indicated by the horizontal pink line).

A more conservative estimator of the same extreme quantile-VaR $q_{\tau_n'}$ has been recently derived by Daouia *et al.* (2018) as

$$\widehat{e}^{\star}_{\widehat{\tau}_n'} = \left(\frac{1-\widehat{\tau}_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_H}\widehat{e}_{\tau_n} \quad \text{where} \quad \widehat{\tau}_n' = 1 - \frac{1-\tau_n'}{\widehat{\gamma}_H^{-1}-1} = 1 - \frac{p_n}{\widehat{\gamma}_H^{-1}-1}$$

and $\widehat{e}_{\tau_n}$ stands for the empirical counterpart of the expectile e_{τ_n}, with $\tau_n = 1-k/n$. The plot of the composite expectile estimator $\widehat{e}^{\star}_{\widehat{\tau}_n'}$ against k (dashed gray curve) indicates

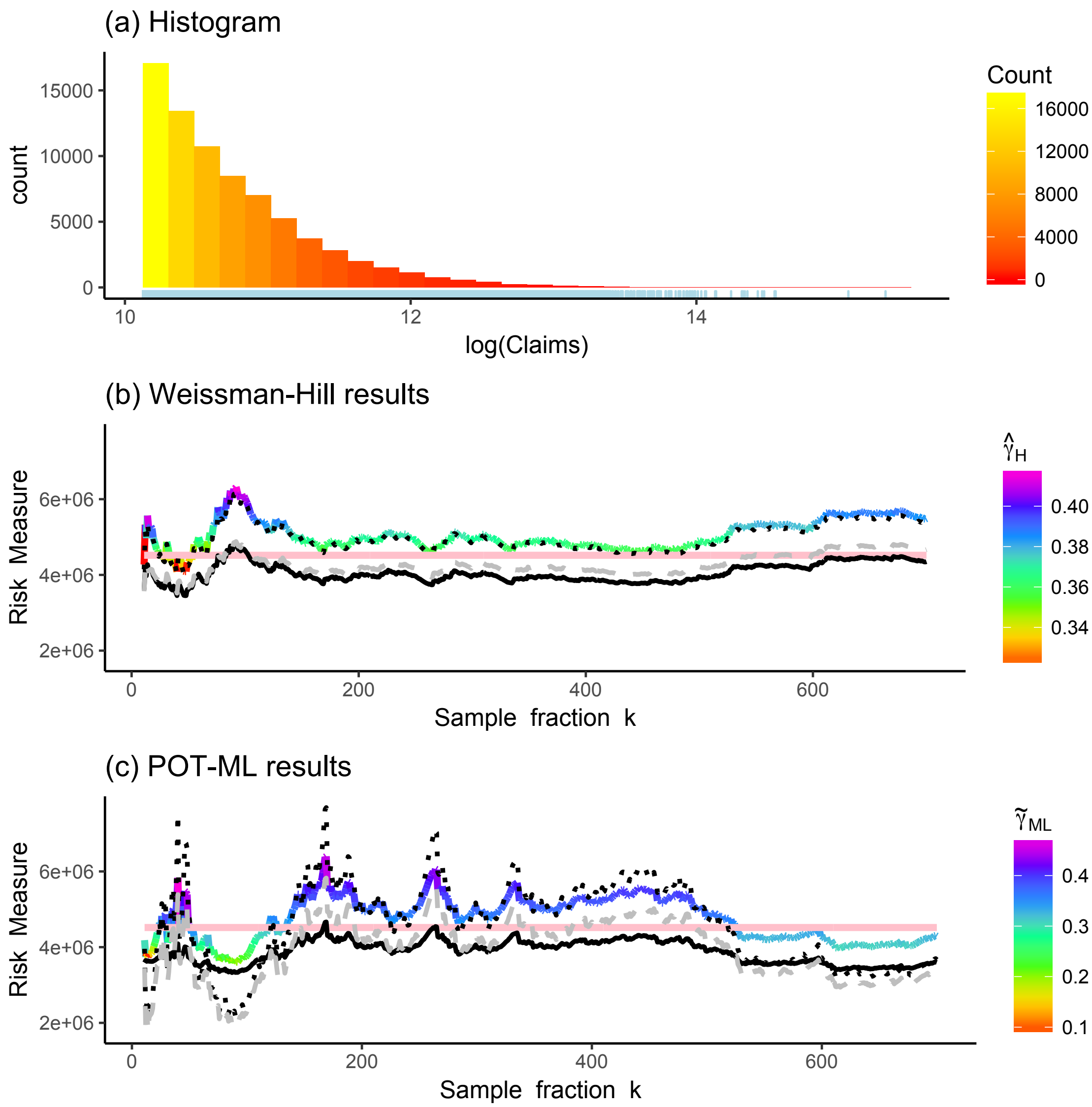

Figure 5: *SOA Group Medical Insurance data. (a) Histogram and scatterplot of the log-claim amounts; (b) Using the Weissman extrapolation method and the Hill estimator: The extremile plots* $k \mapsto \widehat{\xi}^{Q,\star}_{\tau'_n}$ *(rainbow) and* $k \mapsto \widehat{\xi}^{M,\star}_{\tau'_n}$ *(dotted black), along with the quantile plot* $k \mapsto \widehat{q}^{\star}_{\tau'_n}$ *(solid black), the expectile plot* $k \mapsto \widehat{e}^{\star}_{\widehat{\tau}'_n}$ *(dashed gray) and the sample maximum* $Y_{n,n}$ *(pink line); (c) Using the POT approach and the maximum likelihood estimator of* γ*: The extremile plots* $k \mapsto \widetilde{\xi}^{Q,\star}_{\tau'_n}$ *(rainbow) and* $k \mapsto \widetilde{\xi}^{M,\star}_{\tau'_n}$ *(dotted black), along with the quantile plot* $k \mapsto \widetilde{q}^{\star}_{\tau'_n}$ *(solid black), the expectile plot* $k \mapsto \widetilde{e}^{\star}_{\widetilde{\tau}'_n}$ *(dashed gray) and the maximum* $Y_{n,n}$ *(pink line).*

an averaged estimate of around 4.13 million for $k \in [150, 500]$. In contrast to $\widehat{q}^{\star}_{\tau_n'}$ which relies on a single order statistic $Y_{n-k,n}$, the extrapolated expectile estimator $\widehat{e}^{\star}_{\widehat{\tau}_n'}$ is based on the least asymmetrically weighted squares estimator $\widehat{e}_{\tau_n}$, and hence is more sensitive to the magnitude of infrequent large claims. Yet, both $\widehat{e}^{\star}_{\widehat{\tau}_n'}$ and $\widehat{q}^{\star}_{\tau_n'}$ actually estimate the median $q_{\tau_n'}$ of the asymmetric heavy-tailed distribution $K_{\tau_n'}(F)$ of $Z_{\tau_n'} \stackrel{d}{=} \phi_{\tau_n'}(Y)$, see Proposition 1. Accordingly, when faced with very long tails of $Z_{\tau_n'}$, the burden of representing a pessimistic risk measure is thwarted by the robustness properties of the median. The mean of $Z_{\tau_n'}$, which is nothing but the extremile $\xi_{\tau_n'}$, bears naturally much better this burden. The plots of its two estimators $\widehat{\xi}^{Q,\star}_{\tau_n'}$ (rainbow curve) and $\widehat{\xi}^{M,\star}_{\tau_n'}$ (dotted black curve), defined respectively in (9) and (15), clearly afford more pessimistic risk information than the plots of $\widehat{q}^{\star}_{\tau_n'}$ and $\widehat{e}^{\star}_{\widehat{\tau}_n'}$. The final results based on averaging these extremile estimates from the stable region $k \in [150, 500]$ are 4.83 million for $\widehat{\xi}^{Q,\star}_{\tau_n'}$ and 4.78 million for $\widehat{\xi}^{M,\star}_{\tau_n'}$. These estimates deserve indeed to be qualified as 'pessimistic' since they do lie beyond the range of the data, but not by much. This might be good news to practitioners whose concern is to contrast 'pessimistic' and 'optimistic' judgments as in the duality between the mean and the median. Besides this duality, it should also be clear that the resulting extremile estimates have their own intuitive interpretation. Indeed, since $\tau_n' = (1/2)^{1/r}$ with $r \approx 69{,}314$, then $\xi_{\tau_n'} \equiv \mathbb{E}[\max(Y^1, \ldots, Y^r)]$ gives the expected maximum claim amount among a fixed number of 69,314 potential claims. Finally, note that the effect of Hill's estimator $\widehat{\gamma}_H$ of the tail index γ on $\widehat{\xi}^{Q,\star}_{\tau_n'}$ is highlighted by a color-scheme, ranging from dark red (high $\widehat{\gamma}_H$) to dark violet (low $\widehat{\gamma}_H$). This effect closely parallels the influence of $\widehat{\gamma}_H$ on the other extrapolated risk estimates $\widehat{\xi}^{M,\star}_{\tau_n'}$, $\widehat{q}^{\star}_{\tau_n'}$ and $\widehat{e}^{\star}_{\widehat{\tau}_n'}$.

The risk estimates $\widehat{q}^{\star}_{\tau_n'}$, $\widehat{e}^{\star}_{\widehat{\tau}_n'}$, $\widehat{\xi}^{M,\star}_{\tau_n'}$ and $\widehat{\xi}^{Q,\star}_{\tau_n'}$, graphed in Figure 5 (b), are all based on the Hill estimator $\widehat{\gamma}_H$ of γ and the Weissman estimator $\widehat{q}^{\star}_{\tau_n'}$ of $q_{\tau_n'}$. Instead, one may choose to use the maximum likelihood (ML) and Peaks-Over-Threshold (POT) estimators of γ and $q_{\tau_n'}$. The POT estimator of $q_{\tau_n'}$ has the form

$$\widetilde{q}^{\star}_{\tau_n'} = Y_{n-k,n} + \frac{\widetilde{\sigma}_{\mathrm{ML}}}{\widetilde{\gamma}_{\mathrm{ML}}} \left(\left(\frac{np_n}{k} \right)^{-\widetilde{\gamma}_{\mathrm{ML}}} - 1 \right),$$

where $\widetilde{\sigma}_{\mathrm{ML}}$ and $\widetilde{\gamma}_{\mathrm{ML}}$ are chosen here to be the ML estimates for the parameters σ and γ of the Generalized Pareto approximation [see, *e.g.*, Beirlant *et al.* (2004), p.158]. Replacing

$\widehat{\gamma}_H$ and $\widehat{q}^{\star}_{\tau'_n}$ by their respective analogues $\widetilde{\gamma}_{\mathrm{ML}}$ and $\widetilde{q}^{\star}_{\tau'_n}$ in the expectile and extremile risk estimates $\widehat{e}^{\star}_{\widehat{\tau}'_n}$, $\widehat{\xi}^{M,\star}_{\tau'_n}$ and $\widehat{\xi}^{Q,\star}_{\tau'_n}$, we get the alternative versions

$$\widetilde{e}^{\star}_{\widetilde{\tau}'_n} := \left(\frac{1-\widetilde{\tau}'_n}{1-\tau_n}\right)^{-\widetilde{\gamma}_{\mathrm{ML}}} \widehat{e}_{\tau_n} \quad \text{where} \quad \widetilde{\tau}'_n = 1 - \frac{1-\tau'_n}{\widetilde{\gamma}_{\mathrm{ML}}^{-1}-1} = 1 - \frac{p_n}{\widetilde{\gamma}_{\mathrm{ML}}^{-1}-1},$$
$$\widetilde{\xi}^{M,\star}_{\tau'_n} := \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widetilde{\gamma}_{\mathrm{ML}}} \widehat{\xi}^{M}_{\tau_n} \quad \text{and} \quad \widetilde{\xi}^{Q,\star}_{\tau'_n} := \widetilde{q}^{\star}_{\tau'_n}\, \mathcal{G}\left(\widetilde{\gamma}_{\mathrm{ML}}\right).$$

The plots of the alternative risk estimates $\widetilde{q}^{\star}_{\tau'_n}$, $\widetilde{e}^{\star}_{\widetilde{\tau}'_n}$, $\widetilde{\xi}^{M,\star}_{\tau'_n}$ and $\widetilde{\xi}^{Q,\star}_{\tau'_n}$ are graphed in Figure 5 (c). These plots seem to be more volatile than those obtained via the Weissman extrapolation method in Figure 5 (b). A stable region appears, however, for k from 200 up to 500, leading to the averaged estimates $\widetilde{q}^{\star}_{\tau'_n} = 4.10$ million, $\widetilde{e}^{\star}_{\widetilde{\tau}'_n} = 4.52$ million, $\widetilde{\xi}^{M,\star}_{\tau'_n} = 5.32$ million and $\widetilde{\xi}^{Q,\star}_{\tau'_n} = 5.17$ million. These POT and ML-based risk values are larger and hence more conservative than their Weissman and Hill-based analogues $\widehat{q}^{\star}_{\tau'_n} = 3.90$ million, $\widehat{e}^{\star}_{\widehat{\tau}'_n} = 4.13$ million, $\widehat{\xi}^{M,\star}_{\tau'_n} = 4.78$ million and $\widehat{\xi}^{Q,\star}_{\tau'_n} = 4.83$ million. The most substantial difference is 0.54 million, between the rival extremile estimators $\widetilde{\xi}^{M,\star}_{\tau'_n}$ and $\widehat{\xi}^{M,\star}_{\tau'_n}$, followed by a difference of 0.39 million between the expectile estimators $\widetilde{e}^{\star}_{\widetilde{\tau}'_n}$ and $\widehat{e}^{\star}_{\widehat{\tau}'_n}$. Taking a closer look to the construction of these competing asymmetric least squares estimators, we see that their substantial deviation is mainly due to the use of the Hill estimator $\widehat{\gamma}_H$ in the Weissman extrapolation method and the ML estimator $\widetilde{\gamma}_{\mathrm{ML}}$ in the POT method. The volatility of the POT plots can thus be explained, on the one hand, by an averaged estimate $\widetilde{\gamma}_{\mathrm{ML}} = 0.38$ slightly higher than $\widehat{\gamma}_H = 0.36$ (these estimates were averaged over $k \in [200, 500]$ and $k \in [150, 500]$, respectively), and most importantly, on the other hand, by ML estimates of γ being far more volatile than their Hill counterparts as functions of k. The advantageous stability of $\widehat{\gamma}_H$ relative to $\widetilde{\gamma}_{\mathrm{ML}}$ becomes clear by comparing the color-schemes in Figures 5 (b) and (c). Thanks to the POT method however, one may be able to obtain profile likelihood confidence intervals by applying the elegant device in Chapter 5 of McNeil *et al.* (2015).

5 Discussion

Extremiles have several merits which deserve to be studied in more detail. They are attractive because of their conceptual simplicity, the easy implementation of their estimators and their good properties. They can be defined for a wide range of distributions and summarize a distribution in a similar way as quantiles and expectiles do. They are specified as a least squares analogue of quantiles but, unlike expectiles, they benefit from various equivalent explicit formulations and straightforward interpretations.

As is the case in the duality between the mean and the median, the choice between extremiles and quantiles will usually depend on the application at hand. Quantiles are appealing because of their conventional probabilistic interpretation in terms of relative frequency and their inherent robustness to extreme observations. By contrast, extremiles are more appropriate in any tail analysis where sensitivity to the magnitude of extremes is of prime importance. They afford more valuable information about how spread a distribution is and can serve as a more efficient instrument of risk protection than quantiles and expectiles in actuarial and portfolio allocation problems. Their use could be of genuine interest in any other decision problem where "realistic" and "optimistic" judgments are contrasted such as, for instance, in survival analysis and medical decision making where pessimistic patients might favor the consideration of extremiles rather than quantiles to measure the realistic performance of a hypothetical medical treatment.

In risk management, quantiles do not provide a coherent risk measure because they are not subadditive. By contrast, extremiles define a coherent spectral risk measure thanks to the nice attributes of the special weight-generating function J_τ. Expectiles and the Expected Shortfall are coherent as well, but they have serious disadvantages, related to the absence of comonotonic additivity for expectiles and to the dependency of the Expected Shortfall solely on the tail event. By contrast, extremiles are comonotonically additive and depend on both the tail realizations and their probability. The key advantage of expectiles over extremiles and Expected Shortfall is their property of *elicitability* that corresponds to the existence of a natural forecast verification and comparison methodology (see *e.g.* Ziegel, 2016). Yet, it is known that competing forecasting methods of

Expected Shortfall may be compared and assessed thanks to Fissler and Ziegel's (2016) important result that certain spectral risk measures are jointly elicitable with quantiles. Although this result does not apply automatically to extremiles, these special spectral risk measures can already be recognized as reasonable alternatives to expectiles and traditional quantile-based risk measures. Further comparisons and practical experimentations will undoubtedly yield new refinements.

Supplementary Material

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

Acknowledgement

The authors are grateful to the Editor, an Associate Editor, and two referees for their very valuable comments, which led to an improved presentation of the paper.

References

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

Abramovitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series.

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

Acerbi, C. and Szekely, B. (2014). Backtesting Expected Shortfall. *MSCI*, **December 2014**, 1–34.

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

Bassett, G.W, Koenker, J.R. and Kordas, G. (2004). Pessimistic Portfolio Allocation and Choquet Expected Utility. *Journal of Financial Econometrics*, **4**, 477–492.

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

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

Cheng, C. and Parzen, E. (1997). Unified estimators of smooth quantile and quantile density functions. *Journal of Statistical Planning and Inference*, **2**, 291–307.

Cherny, A. and Madan, D. (2009). New measures for performance evaluation. *Review of Financial Studies*, **22**, 2571–2606.

Cirillo, P. and Taleb, N.N. (2016). Expected shortfall estimation for apparently infinite-mean models of operational risk. *Quantitative Finance*, **16**, 1485–1494.

Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*, John Wiley & Sons, Chichester.

Daouia, A. and Gijbels, I. (2011). Estimating frontier cost models using extremiles. In *Festschrift in honor of Leopold Simar*, Editors: P. Wilson, I. Van Keilegom. Springer-Verlag: Berlin, Heidelberg, ISBN 978-3-7908-2348-6. Chapter 4, 65–81.

Daouia, A., Girard, S. and Stupfler, G. (2017). Extreme M-quantiles as risk measures: From L^1 to L^p optimization, *Bernoulli* (to appear). Available at https://hal.inria.fr/hal-01585215.

Daouia, A., Girard, S. and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B*, **80**, 263–292.

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

Delbaen, F. (2002). *Coherent risk measures – Lectures given at the Cattedra Galileiana at the Scuola Normale Superiore di Pisa, March 2000*, Publications of the Scuola Normale Superiore, Pisa.

de Valk, C. (2016). Approximation and estimation of very small probabilities of multivariate extreme events. *Extremes*, **19**, 687–717.

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

El Methni, J. and Stupfler, G. (2018). Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions. *Econometrics and Statistics*, **6**, 129–148.

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

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

Föllmer, H. and Knispel, T. (2013). Convex risk measures: Basic facts, law-invariance and beyond, asymptotics for large portfolios. In *Handbook of the Fundamentals of Financial Decision Making: Part II*, pp. 507–554, World Scientific.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, **15**, 1049–1054.

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

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

Hosking, J.R.M., Wallis, J.R. and Wood, F. (1985). Estimation of the generalized extreme-value distribution by the method of probability weighted moments. *Technometrics*, **27**, 251–261.

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

Jones, B.L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, **7**, 44–54.

Jouini, E., Schachermayer, W. and Touzi, N. (2006). Law invariant risk measures have the Fatou property. *Advances in Mathematical Economics*, **9**, 49–71.

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

Kuan, C-M., Yeh, J-H. and Hsu, Y-C. (2009). Assessing value at risk with CARE, the Conditional Autoregressive Expectile models. *Journal of Econometrics*, **2**, 261–270.

Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research* **15**, 1055–1064.

McNeil, A.J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management,* revised edition, New Jersey: Princeton University Press.

Nešlehová, J., Embrechts, P. and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. *Journal of Operational Risk*, **1**, 3–25.

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

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

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

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Shorack, G.R. (1972). Functions of order statistics. *The Annals of Mathematical Statistics*, **43**, 412–427.

Shorack, G.R. (2000). *Probability for Statisticians*, Springer, New York.

Steenbergen, H.M.G.M., Lassing, B.L., Vrouwenvelder, A.C.W.M. and Waarts, P.H. (2004). Reliability analysis of flood defence systems. *Heron*, **49**, 51–73.

Tsukahara, H. (2009). Some properties of distortion risk measures. *Advances in Mathematical Economics*, **12**, 153–166.

Wang, S.S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, **26**, 71–92.

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

Wirch, J.L. and Hardy, M.R. (1999). A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, **25**, 337–347.