Extreme M-quantiles as risk measures: From L^1 to L^p optimization

ABDELAATI DAOUIA[1], STÉPHANE GIRARD[2] and GILLES STUPFLER[3]

[1]*Toulouse School of Economics, University of Toulouse Capitole, France.*
E-mail: abdelaati.daouia@tse-fr.eu
[2]*INRIA Grenoble Rhône-Alpes / LJK Laboratoire Jean Kuntzmann, France.*
E-mail: Stephane.Girard@inria.fr
[3]*School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD,*
United Kingdom. E-mail: gilles.stupfler@nottingham.ac.uk

The class of quantiles lies at the heart of extreme-value theory and is one of the basic tools in risk management. The alternative family of expectiles is based on squared rather than absolute error loss minimization. It has recently been receiving a lot of attention in actuarial science, econometrics and statistical finance. Both quantiles and expectiles can be embedded in a more general class of M-quantiles by means of L^p optimization. These generalized L^p-quantiles steer an advantageous middle course between ordinary quantiles and expectiles without sacrificing their virtues too much for $1 < p < 2$. In this paper, we investigate their estimation from the perspective of extreme values in the class of heavy-tailed distributions. We construct estimators of the intermediate L^p-quantiles and establish their asymptotic normality in a dependence framework motivated by financial and actuarial applications, before extrapolating these estimates to the very far tails. We also investigate the potential of extreme L^p-quantiles as a tool for estimating the usual quantiles and expectiles themselves. We show the usefulness of extreme L^p-quantiles and elaborate the choice of p through applications to some simulated and financial real data.

Keywords: asymptotic normality; dependent observations; expectiles; extrapolation; extreme values; heavy tails; L^p optimization; mixing; quantiles; tail risk

1. Introduction

A very important problem in actuarial science, econometrics and statistical finance involves quantifying the "riskiness" implied by the distribution of a non-negative loss variable or a real-valued profit-loss variable X. Greater variability of the random variable X and particularly a heavier tail of its distribution necessitate a higher capital reserve for portfolios or price of the insurance risk. The class of quantiles is one of the basic tools in risk management and lies at the heart of extreme-value theory. A leading quantile-based risk measure in banking and other financial institutions is Value at Risk (VaR capital requirement) with a confidence level $\tau \in (0, 1)$. It is defined as the τth quantile $q(\tau)$ 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. The quantile $q(\tau)$ of X is uniquely defined through the generalized inverse $F_X^{-1}(\tau) = \inf\{x : F_X(x) \geq \tau\}$ of the underlying distribution function F_X. It can also be obtained by minimizing asymmetrically

weighted mean absolute deviations (Koenker and Bassett [49]):

$$q(\tau) = \underset{q \in \mathbb{R}}{\arg\min} \, \mathbb{E}\big(\eta_\tau(X - q; 1) - \eta_\tau(X; 1)\big),$$

where $\eta_\tau(x; 1) = |\tau - \mathbb{1}_{\{x \leq 0\}}| \cdot |x|$ stands for the quantile check function, with $\mathbb{1}_{\{.\}}$ being the indicator function. This property has recently been receiving a lot of attention in the actuarial literature since it corresponds to the existence of a natural backtesting methodology. Gneiting [38] introduced the general notion of elicitability for a functional that is defined by means of the minimization of a suitable asymmetric loss function. The relevance of elicitability in connection with backtesting has been discussed, for instance, by Embrechts and Hofert [34] and Bellini and Di Bernardino [5]. It is generally accepted that elicitability is a desirable property for model selection, estimation, generalized regression, computational efficiency, forecasting and testing algorithms.

Despite their elicitability and strong intuitive appeal, quantiles are not always satisfactory. From the point of view of axiomatic theory, an influential paper in the literature by Artzner *et al.* [3] provides a foundation for coherent risk measures. Quantiles satisfy their requirements of translation invariance, monotonicity and positive homogeneity, but not the property of subadditivity. Hence, quantiles fail to be coherent, while they are elicitable. In contrast to quantiles, the most popular coherent risk measure, referred to as Expected Shortfall, is not elicitable. The relationship of coherency with elicitability has been addressed in *e.g.* Ziegel [72]. From a statistical viewpoint, the asymptotic variance of quantile estimators involves the value of the density function of X at $q(\tau)$ which is notoriously difficult to estimate. From an extreme-value perspective, and perhaps most seriously, quantiles are often criticized for being too liberal or optimistic since they only depend on the frequency of tail losses and not on their values. To reduce this loss of information and other vexing defects of quantiles, Newey and Powell [54] substituted the absolute deviations in the asymmetric loss function of Koenker and Bassett with squared deviations to define the concept of τth expectile

$$\xi(\tau) = \underset{q \in \mathbb{R}}{\arg\min} \, \mathbb{E}\big(\eta_\tau(X - q; 2) - \eta_\tau(X; 2)\big),$$

where $\eta_\tau(x; 2) = |\tau - \mathbb{1}_{\{x \leq 0\}}|x^2$. The special case $\tau = 1/2$ leads to the expectation of X. More generally, by taking the derivative with respect to q in the L^2 criterion and setting it to zero, we get the equation

$$\tau = \frac{\mathbb{E}[|X - \xi(\tau)|\mathbb{1}_{\{X \leq \xi(\tau)\}}]}{\mathbb{E}|X - \xi(\tau)|},$$

that is, the τth expectile specifies the position $\xi(\tau)$ such that the ratio of the average distance from the data to and below $\xi(\tau)$ to the average distance of the data to $\xi(\tau)$ is $100\tau\%$. Thus, the expectile shares an interpretation similar to the quantile, replacing the distance by the number of observations. Jones [47] established that expectiles are precisely the quantiles, not of the original distribution, but of a related transformation. Abdous and Rémillard [1] proved that quantiles and expectiles of the same distribution coincide under the hypothesis of weighted-symmetry. Yao and Tong [70] showed that there exists a unique bijective function $\mathfrak{h} : (0, 1) \to (0, 1)$, depending

on the underlying distribution, such that $q(\tau)$ coincides with $\xi(\mathfrak{h}(\tau))$ for all $\tau \in (0, 1)$. More recently, Zou [73] has derived a class of generic distributions for which $\xi(\tau)$ and $q(\tau)$ coincide for all $\tau \in (0, 1)$. Also, as suggested by many authors including Efron [30], Yao and Tong [70], Schnabel and Eilers [64] and Schulze Waltrup *et al.* [65], quantile estimates and their strong intuitive appeal can be recovered directly from asymmetric least squares estimates of a set of expectiles.

The advantages of expectiles include their computing expedience and their efficient use of the data as the weighted least squares rely on the distance to observations, while the quantile method only uses the information on whether an observation is below or above the predictor. Also, inference on expectiles is much easier than inference on quantiles (see, *e.g.*, Abdous and Rémillard [1]). Most importantly, expectiles depend on both the tail realizations of the loss variable and their probability. This motivated Kuan *et al.* [50] to introduce the expectile-based VaR as $-\xi(\tau)$ for real-valued profit-loss distributions. The key advantage of this new instrument of risk protection is that it defines the only coherent risk measure that is also elicitable (Ziegel [72]). Further theoretical and numerical results obtained by Bellini and Di Bernardino [5] indicate that expectiles are perfectly reasonable alternatives to both classical quantile-based VaR and Expected Shortfall.

A disadvantage of the expectile method is that, by construction, it is not as robust against outliers as the quantiles. This may cause trouble when estimating the tail risk that translates into considering the prudentiality level $\tau = \tau_n \to 0$ or $\tau_n \to 1$ as the sample size n goes to infinity. The behavior of tail expectiles $\xi(\tau_n)$ and the connection with their quantile analogues $q(\tau_n)$ have been elucidated only very recently by Bellini *et al.* [6], Mao *et al.* [51], Bellini and Di Bernardino [5] and Mao and Yang [52], when X belongs to the domain of attraction of a Generalized Extreme Value distribution. The estimation of $\xi(\tau_n)$ in the challenging maximum domain of attraction of Pareto-type distributions, where standard empirical expectiles are often unstable due to data sparsity, has been considered in Daouia *et al.* [18]. In most studies on actuarial and financial data, it has been found that Pareto-type distributions, with tail index $\gamma > 0$, describe quite well the tail structure of losses (see, *e.g.*, Embrechts *et al.* [35], page 9, and Resnick [60], page 1). An intrinsic difficulty with expectiles is that their existence requires $\mathbb{E}|X| < \infty$, which amounts to supposing $\gamma < 1$. Even more seriously, the condition $\gamma < 1/2$ is required to ensure that asymmetric least squares estimators of $\xi(\tau_n)$ are asymptotically Gaussian. Already in the intermediate case, where $n(1 - \tau_n) \to \infty$ as $\tau_n \to 1$, good estimates may require in practice $\gamma < 1/4$. Similar concerns occur with the Expected Shortfall, the so-called Conditional Tail Expectation or certain extreme Wang distortion risk measures (see El Methni *et al.* [31] and El Methni and Stupfler [32,33]). This restricts appreciably the range of potential applications as may be seen in the financial setting from the R package *'CASdatasets'* where realized values of the tail index γ were found to be larger than $1/4$ in several instances.

Instead of the asymmetric square loss, a natural modification of the expectile check function is to use the power loss function

$$\eta_\tau(x; p) = |\tau - \mathbb{1}_{\{x \le 0\}}| \cdot |x|^p, \qquad p \ge 1,$$

leading to

$$q_\tau(p) = \underset{q \in \mathbb{R}}{\arg\min} \, \mathbb{E}\big(\eta_\tau(X - q; p) - \eta_\tau(X; p)\big).$$

These quantities have already been coined as L^p-quantiles by Chen [15]. They define a special case of the generic concept of M-quantiles introduced earlier by Breckling and Chambers [10]. Their existence requires $\mathbb{E}|X|^{p-1} < \infty$. This is a weaker condition, compared with the condition of existence of expectiles, when $p < 2$. The choice of $p < 2$ is also required when the influence of potential outliers is taken into account. The class of L^p-quantiles, with $p \in (1, 2)$, steers an advantageous middle course between the robustness of quantiles ($p = 1$) and the sensitivity of expectiles ($p = 2$) to the magnitude of extreme losses. For fixed levels τ staying away from the distribution tails, inference on $q_\tau(p)$ is straightforward using M-estimation theory. The main purpose of this paper is to extend the estimation of $q_\tau(p)$ and its large sample theory far enough into the upper tail $\tau = \tau_n \to 1$ as $n \to \infty$. There are many important events including big financial losses, high medical costs, large claims in (re)insurance, high bids in auctions, just to name a few, where modeling and estimating the extreme rather than central L^p-quantiles of the underlying distribution is a highly welcome development. We refer to the book of de Haan and Ferreira [23] for a modern formulation of this typical extreme value problem in the case $p = 1$ and to Daouia *et al.* [18] in the case $p = 2$.

More specifically, it is our goal to establish two estimators of $q_{\tau_n}(p)$ for a general p and to unravel their asymptotic behavior for τ_n at an extremely high level that can be even larger than $(1 - 1/n)$, in a framework of weak dependence motivated by the aforementioned financial and actuarial applications. To do so, we first estimate the intermediate tail L^p-quantiles of order $\tau_n \to 1$ such that $n(1 - \tau_n) \to \infty$, and then extrapolate these estimates to the proper extreme L^p-quantile level τ_n which approaches 1 at an arbitrarily fast rate in the sense that $n(1 - \tau_n) \to c$, for some constant c. The main results, established for a strictly stationary and suitably mixing sequence of observations, state the asymptotic normality of our estimators for distributions with tail index $\gamma < [2(p-1)]^{-1}$. As such, unlike expectiles, extreme L^p-quantile estimates cover a larger class of heavy-tailed distributions for $p < 2$. It should also be clear that, in contrast to standard quantiles, generalized L^p-quantiles take into account the whole tail information about the underlying distribution for $p > 1$. These additional benefits raise the following important question: how to elaborate the choice of p in the interval $[1, 2]$? This choice is mainly a practical issue that we first pursue here through some simulation experiments. Although the value of p minimizing the Mean Squared Error of empirical L^p-quantiles depends on the tail index γ, Monte Carlo evidence indicates that the choice of $p \in (1.2, 1.6)$ guarantees a good compromise for Pareto-type distributions with $\gamma < 1/2$. In contrast, when the empirical estimates are extrapolated to properly extreme levels τ_n, the underlying tail L^p-quantiles seem to be estimated more accurately for $p \in [1, 1.3]$ or $p \in [1.7, 2]$. We elaborate further this question from a forecasting perspective, trying to perform extreme L^p-quantile estimation accurately on historical data.

Yet, the L^p-quantile approach is not without disadvantages. It does not have an intuitive interpretation as direct as ordinary L^1-quantiles. More precisely, the generalized quantile $q_\tau(p)$ exists, is unique and satisfies

$$\tau = \frac{\mathbb{E}[|X - q_\tau(p)|^{p-1} \mathbb{1}_{\{X \leq q_\tau(p)\}}]}{\mathbb{E}[|X - q_\tau(p)|^{p-1}]}. \tag{1}$$

It can thus be interpreted only in terms of the average distance from X in the (nonconvex when $1 < p < 2$) space L^{p-1}. This should not be considered to be a serious disadvantage however,

since one can recover the usual quantiles $q(\alpha_n) \equiv q_{\alpha_n}(1)$ of extreme order $\alpha_n \to 1$ and their strong intuitive appeal from tail L^p-quantiles $q_{\tau_n}(p)$, $\tau_n \to 1$, that coincide with $q_{\alpha_n}(1)$. Indeed, given a relative frequency of interest α_n, the level τ_n such that $q_{\tau_n}(p) \equiv q_{\alpha_n}(1)$ can be written in closed form as

$$\tau_n = \frac{\mathbb{E}[|X - q_{\alpha_n}(1)|^{p-1} \mathbb{1}_{\{X \leq q_{\alpha_n}(1)\}}]}{\mathbb{E}[|X - q_{\alpha_n}(1)|^{p-1}]} \tag{2}$$

in view of (1). One can then estimate τ_n via extrapolation techniques before calculating the corresponding L^p-quantile estimators. In this way, we perform tail L^p-quantile estimation as a main tool when the ultimate interest is in estimating the intuitive L^1-quantiles themselves.

From the point of view of the axiomatic theory of risk measures, the L^p-quantile method can be criticized for not being coherent for all values of p. According to Bellini *et al.* [6] and Ziegel [72], the only L^p-quantiles that are actually coherent risk measures are the expectiles, or L^2-quantiles. This disadvantage does not prevent the investigator, however, to employ tail L^p-quantiles $q_{\tau_n}(p)$ as a tool for estimating extreme expectiles $\xi(\alpha_n) \equiv q_{\alpha_n}(2)$ by applying again (1) in conjunction with similar considerations to the above in extreme quantile estimation. Built on the presented extreme L^p-quantile estimators, we construct three different tail expectile estimators and derive their asymptotic normality. Two among these new estimators appear to be appreciably more efficient relatively to the rival expectile estimators of Daouia *et al.* [18] in the important case of profit-loss distributions with long tails.

The paper is organized as follows. Section 2 describes in some detail how population L^p-quantiles $q_\tau(p)$ are linked to standard quantiles $q_\tau(1)$ as $\tau \to 1$. Section 3 deals with estimation of intermediate and extreme L^p-quantiles $q_{\tau_n}(p)$ for $p > 1$. Estimators of the extreme level τ_n in (2) are discussed in Section 4, with implications for recovering composite estimators of high quantiles $q_{\alpha_n}(1)$. Extrapolated high expectile ($p = 2$) estimation is discussed in Section 5. The theory in these sections is derived in the general case of stationary and dependent data satisfying a mixing condition. A detailed simulation study and a concrete application to the S&P500 Index are given, respectively, in Section 6 and Section 7 to illustrate the usefulness of extremal L^p-quantiles. Proofs and further simulation results are deferred to a supplementary material [19].

2. Extremal population L^p-quantiles

This section describes in detail what happens for large population L^p-quantiles and how they are linked to large standard quantiles. We denote in the sequel the cumulative distribution function of X by F, that we suppose to be continuous, and its survival function by $\overline{F} = 1 - F$. We first assume that X has a heavy right-tail or, equivalently, that $\overline{F}$ satisfies the following regular variation condition:

$\mathcal{C}_1(\gamma)$ The function $\overline{F}$ is regularly varying in a neighborhood of $+\infty$ with index $-1/\gamma < 0$, that is,

$$\lim_{t \to +\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma} \qquad \text{for all } x > 0.$$

This is equivalent to the standard first-order condition

$$\lim_{t \to +\infty} \frac{U(tx)}{U(t)} = x^{\gamma} \qquad \text{for all } x > 0,$$

by Theorem 1.2.1 in de Haan and Ferreira [23], where $U(t) = (1/\overline{F})^{\leftarrow}(t)$ is the left-continuous inverse of $1/\overline{F}$. In contrast to many situations in extreme value analysis, we do not assume here that X is positive or even bounded below. In particular X may have a heavy left-tail as well, a case that we shall discuss in what follows.

Under this condition, the asymptotic properties (for $\tau \to 1$) of the usual quantile $q_\tau(1)$ have been extensively studied in the literature as may be seen from, for example, de Haan and Ferreira [23]. Here, we focus on the less discussed generalized quantiles $q_\tau(p)$ with $p > 1$. Denoting by $X_- = \max(-X, 0)$ the negative part of X, we first have the following asymptotic connection between $\overline{F}(q_\tau(p))$ and $\overline{F}(q_\tau(1)) \equiv 1 - \tau$.

Proposition 1. *Assume that the survival function $\overline{F}$ satisfies condition $\mathcal{C}_1(\gamma)$. For any $p > 1$, whenever $\mathbb{E}(X_-^{p-1}) < \infty$ and $\gamma < 1/(p-1)$, we have*

$$\lim_{\tau \uparrow 1} \frac{\overline{F}(q_\tau(p))}{1 - \tau} = \frac{\gamma}{B(p, \gamma^{-1} - p + 1)},$$

where $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} \, dt$ stands for the Beta function.

Note that when the survival function $\overline{F}$ satisfies condition $\mathcal{C}_1(\gamma)$ and $\gamma < 1/(p-1)$, we have $\mathbb{E}(X_+^{p-1}) < \infty$ with $X_+ = \max(X, 0)$. This entails together with condition $\mathbb{E}(X_-^{p-1}) < \infty$ that $\mathbb{E}|X|^{p-1} < \infty$, and hence the L^p-quantiles of X are indeed well-defined. Even more strongly, we get the following direct asymptotic connection between $q_\tau(p)$ and $q_\tau(1)$ themselves.

Corollary 1. *Under the conditions of Proposition 1, we have*

$$\lim_{\tau \uparrow 1} \frac{q_\tau(p)}{q_\tau(1)} = \left[\frac{\gamma}{B(p, \gamma^{-1} - p + 1)} \right]^{-\gamma}.$$

Accordingly, extreme L^p-quantiles are asymptotically proportional to extreme usual quantiles, for all $p > 1$. The evolution of the proportionality constant

$$C(\gamma; p) := \left[\frac{\gamma}{B(p, \gamma^{-1} - p + 1)} \right]^{-\gamma}$$

with respect to $\gamma \in (0, 1/2]$ is visualized in Figure 1, for some values of $p \in [1, 2]$. It can be seen that the usual quantile $q_\tau(1)$ is more spread (conservative) than the L^p-quantile $q_\tau(p)$ as the level $\tau \to 1$. This property is of particular interest in actuarial risk theory, where loss distributions typically belong to the maximum domain of attraction of Pareto-type distributions with tail index $\gamma < 1/2$. Indeed, when the ordinary quantile breaks down at an extremely high tail
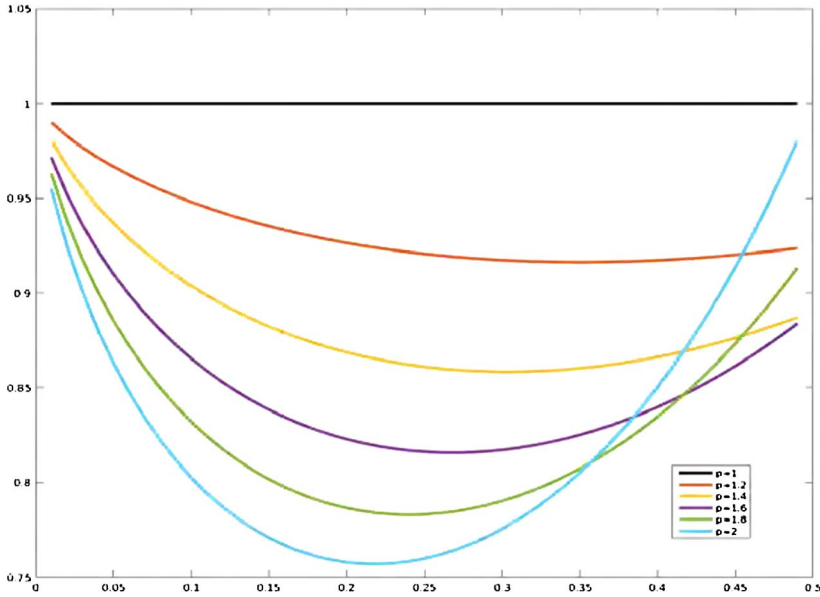
Figure 1. Behavior of $\gamma \in (0, 1/2] \mapsto C(\gamma; p)$ for some values of $p \in [1, 2]$.

probability τ (and hence the underlying VaR changes drastically the order of magnitude of the capital requirement), its generalized L^p-quantile analogue remains definitely more liberal. The latter would result in less excessive amounts of required capital reserve, which might be good news to actuarial institutions.

In the particular case of integers p, we get the next corollary immediately from the identities

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)} \quad \text{and} \quad \Gamma(x + 1) = x\Gamma(x) \qquad \text{for all } x, y > 0,$$

where $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} \, dt$ denotes Euler's Gamma function.

Corollary 2. *Assume that the survival function $\overline{F}$ satisfies condition $\mathcal{C}_1(\gamma)$. Assume that $p = k + 1$ where k is a positive integer. Whenever $\mathbb{E}(X_-^k) < \infty$ and $\gamma < 1/k$, we have*

$$\lim_{\tau \uparrow 1} \frac{\overline{F}(q_\tau(k+1))}{1 - \tau} = \frac{\prod_{j=1}^k (1 - j\gamma)}{\gamma^k k!}.$$

Note that for $p = 2$, we find that

$$\lim_{\tau \uparrow 1} \frac{\overline{F}(q_\tau(2))}{1 - \tau} = \gamma^{-1} - 1$$

which was already shown in Daouia *et al.* [18].

Next, we shall derive some asymptotic expansions of L^p-quantiles, which shall be very useful when it comes to establish the asymptotic normality of extreme L^p-quantile estimators in the next section. As is customary in the case of ordinary quantiles, this requires the extra condition:

$\mathcal{C}_2(\gamma, \rho, A)$ The function $\overline{F}$ is second-order regularly varying in a neighborhood of $+\infty$ with index $-1/\gamma < 0$, second-order parameter $\rho \leq 0$ and an auxiliary function A having constant sign and converging to 0 at infinity, that is,

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

where the right-hand side should be read as $\frac{x^{-1/\gamma} \log x}{\gamma^2}$ when $\rho = 0$.

This classical second-order condition controls the rate of convergence in $\mathcal{C}_1(\gamma)$: in particular, the function $|A|$ is regularly varying with index $\rho \leq 0$, and therefore, the larger $|\rho|$ is, the faster the function $|A|$ converges to 0 and the smaller the error in the approximation of the right tail of X by a Pareto tail will be. Further elements of interpretation of the extreme value condition $\mathcal{C}_2(\gamma, \rho, A)$ can be found in Beirlant *et al.* [4] and de Haan and Ferreira [23] along with a list of examples of commonly used continuous distributions satisfying this assumption: for instance, the (Generalized) Pareto, Burr, Fréchet, Student, Fisher and Inverse-Gamma distributions all satisfy this condition. More generally, so does any distribution whose distribution function F satisfies

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

where $a > 0$, $b \in \mathbb{R} \setminus \{0\}$ and $c > 0$ are constants. This contains in particular the Hall–Weiss class of models (see Hua and Joe [45]), and it is straightforward to see that in any such case, condition $\mathcal{C}_2(\gamma, \rho, A)$ is met with $\rho = -c\gamma$ and $A(t) = -a^{-c\gamma-1}bc\gamma^2 t^{-c\gamma}$.

Besides, as can be seen from Theorem 2.3.9 in de Haan and Ferreira [23], condition $\mathcal{C}_2(\gamma, \rho, A)$ is equivalent to the perhaps more usual extremal assumption on the tail quantile function U that

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

From now on, we denote by $\overline{F}_-$ the survival function of $-X$. Also, a survival function S will be said to be *light-tailed* (and by convention, we shall say it has tail index 0) if it satisfies $x^a S(x) \to 0$ as $x \to +\infty$, for all $a > 0$. The following second-order based refinement of Proposition 1 is the key element in order to obtain the desired asymptotic expansion of L^p-quantiles.

Proposition 2. *Assume that $p > 1$ and*:

- $\overline{F}$ *satisfies condition* $\mathcal{C}_2(\gamma_r, \rho, A)$;
- $\overline{F}_-$ *is either light-tailed or satisfies condition* $\mathcal{C}_1(\gamma_\ell)$;
- $\gamma_r < 1/(p-1)$, *and* $\gamma_\ell < 1/(p-1)$ *in case* $\overline{F}_-$ *is heavy-tailed.*

Then

$$\frac{\overline{F}(q_\tau(p))}{1 - \tau} = \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} (1 + R(\tau, p)),$$

where

$$R(\tau, p) = -\frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)}$$

$$\times \left((1 - \tau)(1 + o(1)) + K(p, \gamma_r, \rho) A\left(\frac{1}{1 - \tau}\right)(1 + o(1)) \right)$$

$$- (p - 1)\left(\left[\frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \right]^{\min(\gamma_r, 1)} R_r\big(q_\tau(1), p, \gamma_r\big) \right.$$

$$\left. - \left[\frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \right]^{\gamma_r / \max(\gamma_\ell, 1)} R_\ell\big(q_\tau(1), p, \gamma_\ell\big) \right)$$

as $\tau \uparrow 1$, *with*

$$K(p, \gamma_r, \rho)$$

$$= \begin{cases} \frac{1}{\gamma_r^2 \rho}\left[\frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \right]^{-\rho} \\ \quad \times \left[(1 - \rho)B\big(p, (1 - \rho)\gamma_r^{-1} - p + 1\big) - B\big(p, \gamma_r^{-1} - p + 1\big) \right] & \text{if } \rho < 0, \\ \frac{p - 1}{\gamma_r^2} \int_1^{+\infty} (x - 1)^{p-2} x^{-1/\gamma_r} \log(x)\, dx & \text{if } \rho = 0, \end{cases}$$

$$R_r(q, p, \gamma_r)$$

$$= \begin{cases} \frac{\mathbb{E}(X \mathbb{1}_{\{0 < X < q\}})}{q}(1 + o(1)) & \text{if } \gamma_r \leq 1, \\ \overline{F}(q)B\big(p - 1, 1 - \gamma_r^{-1}\big)(1 + o(1)) & \text{if } \gamma_r > 1, \end{cases}$$

and

$$R_\ell(q, p, \gamma_\ell)$$

$$= \begin{cases} -\frac{\mathbb{E}(X \mathbb{1}_{\{-q < X < 0\}})}{q}(1 + o(1)) & \text{if } \gamma_\ell \leq 1 \text{ or } \overline{F}_- \text{ is light-tailed,} \\ F(-q)B\big(\gamma_\ell^{-1} - p + 1, 1 - \gamma_\ell^{-1}\big)(1 + o(1)) & \text{if } \gamma_\ell > 1. \end{cases}$$

When X is integrable, and in particular when expectiles of X can be computed, the asymptotic expansion of L^p-quantiles reduces to the following.

Corollary 3. *Under the conditions of Proposition* 2, *if* $\mathbb{E}|X| < \infty$, *then*

$$\frac{\overline{F}(q_\tau(p))}{1 - \tau} = \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)}(1 + r(\tau, p))$$

as $\tau \uparrow 1$, where

$$r(\tau, p) = -(p-1)\left[\frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)}\right]^{\gamma_r} \frac{1}{q_\tau(1)}\left(\mathbb{E}(X) + \mathrm{o}(1)\right)$$

$$-\frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} K(p, \gamma_r, \rho) A\left(\frac{1}{1-\tau}\right)(1 + \mathrm{o}(1)).$$

Finally, we get the following refined asymptotic expansion of $q_\tau(p)$ itself with respect to the ordinary quantile $q_\tau(1)$.

Proposition 3. *Under the conditions of Proposition 2, if in addition $\overline{F}$ is strictly decreasing*:

$$\frac{q_\tau(p)}{q_\tau(1)} = C(\gamma_r; p)\left(1 - \gamma_r R(\tau, p) + \left\{\frac{1}{\rho}\left[\left[\frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)}\right]^{-\rho} - 1\right] + \mathrm{o}(1)\right\} A\left(\frac{1}{1-\tau}\right)\right)$$

as $\tau \uparrow 1$.

3. Estimation of high L^p-quantiles

Suppose, as will be the case in the remainder of this paper, that we observe a random sample $(X_1, \ldots, X_n)$ from a strictly stationary sequence $(X_1, X_2, \ldots)$, in the sense that for any positive integers k and l, the k-tuples $(X_1, \ldots, X_k)$ and $(X_l, \ldots, X_{l+k-1})$ have the same distribution. Suppose further that the common marginal distribution of that sequence is that of X and denote by $X_{1,n} \leq \cdots \leq X_{n,n}$ the ascending order statistics of the observed sample $(X_1, \ldots, X_n)$. The overall objective in this section is to estimate extreme L^p-quantiles $q_{\tau_n}(p)$ of X, where $\tau_n \to 1$ as $n \to \infty$. Here τ_n may approach one at any rate, covering the special cases of intermediate L^p-quantiles with $n(1 - \tau_n) \to \infty$ and extreme L^p-quantiles with $n(1 - \tau_n) \to c$, where c is some constant.

In order to do so, we need to specify the dependence framework we shall be working in. Dependence frameworks and time series models have been used for a long time in statistical and econometric considerations when estimating nonextreme (*i.e.*, central) quantities, including regression contexts, by employing well-established theoretical arguments; we refer in particular to Boente and Fraiman [7], Honda [43], Zhao *et al.* [71], Kuan *et al.* [50] and references therein in the case of quantiles and Yao and Tong [70], Cai [13] and references therein in the case of expectiles. Let us emphasize that this is arguably not, however, the case in statistical treatments of extreme value theory, even when considering the kind of financial or actuarial applications this paper focuses on. The earliest theoretical development in this context is Hsing [44], who worked on the asymptotic properties of the Hill estimator (Hill [42]) of the tail index γ for strongly mixing (or α-mixing) sequences. Related studies are Resnick and Stărică [57–59], although they worked in a different dependence framework. An important theoretical advance was made by Drees [26–28], who in a series of papers obtained tools making it possible to examine the asymptotic properties of a wide class of statistical indicators of extremes of strictly stationary and dependent observations through a general approximation result for the tail quantile process

by a Gaussian process. These papers were written for absolutely regular (or β-mixing) sequences and influenced a sizeable part of very recent research on the extremes of a time series: we refer, among others, to Davis and Mikosch [20–22], Robert [61,62], Rootzén [63], Drees and Rootzén [29] and de Haan *et al.* [24], which, in their respective contexts, worked under mixing conditions or under assumptions that can be embedded in a mixing framework.

Due to the flexibility of the results of Drees [28], and the necessity here to extrapolate beyond the available data and therefore to use an estimator of the tail index, we also elect to work in such a mixing framework, which we introduce hereafter. For any positive integer m, let $\mathcal{F}_{1,m}$ and $\mathcal{F}_{m,\infty}$ denote the past and future σ-fields generated by the sequence (X_n):

$$\mathcal{F}_{1,m} = \sigma(X_1, \ldots, X_m) \quad \text{and} \quad \mathcal{F}_{m,\infty} = \sigma(X_m, X_{m+1}, \ldots).$$

Define then the ϕ-mixing coefficients of the sequence (X_n) by:

$$\forall n \in \mathbb{N} \setminus \{0\}, \qquad \phi(n) = \sup_{\substack{m \in \mathbb{N} \setminus \{0\} \\ A \in \mathcal{F}_{1,m} \\ B \in \mathcal{F}_{m+n,\infty}}} \left| \mathbb{P}(B|A) - \mathbb{P}(B) \right|.$$

The sequence (X_n) is said to be ϕ-mixing if $\phi(n) \to 0$ as $n \to \infty$, and this is precisely the notion of mixing we shall work with to obtain our theoretical results. Intuitively, this condition means that the sequence (X_n) is asymptotically memoryless: an event that happened in the past has a vanishingly small influence on current and future events as the time elapsed since this past event increases. While this notion of mixing is not the β-mixing condition introduced in Drees [28], the rationale behind this choice is twofold:

(i) First, ϕ-mixing is stronger than β-mixing, which shall be used in our extrapolation step. See Bradley [9].

(ii) Second, the ϕ-mixing condition implies a ρ-mixing condition in the following sense: let, for any σ-field $\mathcal{A}$, $L^2(\mathcal{A})$ denote the space of square-integrable random variables which are $\mathcal{A}$-measurable. If $(X_1, X_2, \ldots)$ is ϕ-mixing, then the ρ-mixing coefficients

$$\forall n \in \mathbb{N} \setminus \{0\}, \qquad \rho(n) = \sup_{\substack{m \in \mathbb{N} \setminus \{0\} \\ Y \in L^2(\mathcal{F}_{1,m}) \\ Z \in L^2(\mathcal{F}_{m+n,\infty})}} \left| \frac{\mathrm{Cov}(Y, Z)}{\sqrt{\mathrm{Var}(Y)}\sqrt{\mathrm{Var}(Z)}} \right|$$

must satisfy $\rho(n) \to 0$, see Bradley [9]. If moreover the positive quadrant dependence of any pair (X_1, X_k) (for $k \geq 2$) is assumed, in the sense that

$$\forall x_1, x_k \in \mathbb{R}, \qquad \mathbb{P}(X_1 > x_1, X_k > x_k) \geq \mathbb{P}(X_1 > x_1)\mathbb{P}(X_k > x_k),$$

then the ρ-mixing condition, which by definition is adapted to variance and correlation considerations, shall make it easy to compute the exact rate of growth of the variance of a wide class of sums of square-integrable, $\sigma(X_i)$-measurable random variables, of which our empirical least asymmetrically weighted L^p estimator at the intermediate level introduced in Section 3.1 below is precisely an element. This will then be used in conjunction

with limit theory from Utev [68], valid in our ϕ-mixing framework, to obtain the asymptotic normality of the aforementioned estimator. We refer the reader to Lemmas 7 and 8 in Appendix B of our supplementary material document [19] for the full technical developments. Let us highlight that ρ-mixing alone is in general *not sufficient* to compute the exact rate of convergence of a sum of strictly stationary random variables, see Ibragimov [46], Peligrad [55] and Bradley [8].

It should, finally, be noted that there is in general no relationship between β-mixing and ρ-mixing, and that ϕ-mixing is the least restrictive of the widely used mixing conditions that imply both β- and ρ-mixing, see again Bradley [9]. The ϕ-mixing framework therefore seems to be convenient and reasonable for our purpose.

All in all, we shall work under the following dependence condition on the sequence (X_n):

$\mathcal{S}(\phi)$ The sequence $(X_1, X_2, \ldots)$ is a strictly stationary and ϕ-mixing sequence with positive quadrant dependent bivariate margins.

Note that the positive quadrant dependence of bivariate margins is itself a fairly weak assumption, see Nelsen [53], page 200. It is satisfied if and only if the copula function C_k of the pair (X_1, X_k) satisfies $C_k(u, v) \geq uv$ for any $u, v \in [0, 1]$ (the function C_k always exists by Sklar's theorem; see Sklar [66]). This, in turn, contains the case of extreme-value copulas, which are particularly adapted to the description of the joint extremes of a random pair, see *e.g.* Gudendorf and Segers [41]. Finally, let us mention that condition $\mathcal{S}(\phi)$ allows for the case of an independent and identically distributed sequence $(X_1, X_2, \ldots)$, and we shall specifically highlight the particular form of our results in this case.

3.1. Intermediate levels

We define the empirical least asymmetrically weighted L^p estimator of $q_{\tau_n}(p)$ as

$$
\begin{aligned}
\widehat{q}_{\tau_n}(p) &= \underset{u \in \mathbb{R}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \eta_{\tau_n}(X_i - u; p) \\
&= \underset{u \in \mathbb{R}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} |\tau_n - \mathbb{1}_{\{X_i \leq u\}}| |X_i - u|^p.
\end{aligned}
\tag{3}
$$

Clearly

$$
\sqrt{n(1-\tau_n)}\left(\frac{\widehat{q}_{\tau_n}(p)}{q_{\tau_n}(p)} - 1\right) = \underset{u \in \mathbb{R}}{\arg\min} \, \psi_n(u; p),
\tag{4}
$$

where

$$
\psi_n(u; p) := \frac{1}{p[q_{\tau_n}(p)]^p} \sum_{i=1}^{n} \eta_{\tau_n}\big(X_i - q_{\tau_n}(p) - u q_{\tau_n}(p)/\sqrt{n(1-\tau_n)}; p\big) - \eta_{\tau_n}\big(X_i - q_{\tau_n}(p); p\big).
$$

Since this empirical criterion is a convex function of u, the asymptotic properties of the minimizer follow directly from those of the criterion itself by the convexity lemma of Geyer [37]; see also Theorem 5 in Knight [48]. For this, we require the second-order condition $\mathcal{C}_2(\gamma, \rho, A)$ or, alternatively, the following refined first-order condition:

$\mathcal{H}_1(\gamma)$ For x large enough, the survival function $\overline{F}$ verifies

$$\overline{F}(x) = x^{-1/\gamma} \left\{ c(x) \exp\left(\int_{x_0}^{x} \frac{\Delta(u)}{u}\, du \right) \right\},$$

where $\gamma > 0$, c is a differentiable function such that $c(x) \to c_\infty > 0$ and $xc'(x) \to 0$ as $x \to +\infty$, $x_0 > 0$ and Δ is a measurable function converging to 0 at $+\infty$.

This is a slightly more stringent assumption than the usual first-order condition $\mathcal{C}_1(\gamma)$ and its related Karamata representation (see Theorem B.1.6 in de Haan and Ferreira [23], page 365).

Theorem 1. *Assume that $p > 1$ and*:

- *condition $\mathcal{S}(\phi)$ holds, with $\sum_{n=1}^{\infty} \sqrt{\phi(n)} < \infty$;*
- *there is $\delta > 0$ such that $\mathbb{E}(X_-^{(2+\delta)(p-1)}) < \infty$;*
- *$\overline{F}$ satisfies either condition $\mathcal{H}_1(\gamma)$ or $\mathcal{C}_2(\gamma, \rho, A)$, with $\gamma < 1/[2(p-1)]$;*
- *$\tau_n \uparrow 1$ is such that $n(1 - \tau_n) \to \infty$;*
- *under condition $\mathcal{C}_2(\gamma, \rho, A)$, we have $\sqrt{n(1 - \tau_n)} A((1 - \tau_n)^{-1}) = O(1)$.*

Then there is $\sigma^2 \geq 0$ such that

$$\sqrt{n(1 - \tau_n)} \left(\frac{\widehat{q}_{\tau_n}(p)}{q_{\tau_n}(p)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2 V(\gamma; p)(1 + \sigma^2)\right) \qquad as\ n \to \infty,$$

with

$$V(\gamma; p) = \frac{B(p - 1, \gamma^{-1} - 2p + 2)}{B(p - 1, p)} = \frac{\Gamma(2p - 1)\Gamma(\gamma^{-1} - 2p + 2)}{\Gamma(p)\Gamma(\gamma^{-1} - p + 1)}.$$

In the particular case of an independent sequence $(X_1, X_2, \ldots)$, then $\sigma^2 = 0$, that is,

$$\sqrt{n(1 - \tau_n)} \left(\frac{\widehat{q}_{\tau_n}(p)}{q_{\tau_n}(p)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2 V(\gamma; p)\right) \qquad as\ n \to \infty.$$

Note that the condition $\gamma < 1/[2(p-1)]$ implies $\gamma < 1/(p-1)$ and hence $\mathbb{E}(X_+^{p-1}) < \infty$. Moreover, the condition $\mathbb{E}(X_-^{(2+\delta)(p-1)}) < \infty$ implies $\mathbb{E}(X_-^{p-1}) < \infty$. Hence $\mathbb{E}|X|^{p-1} < \infty$, and thus the L^p-quantiles exist and are finite. Note also that conditions $\gamma < 1/[2(p-1)]$ and $\mathbb{E}(X_-^{(2+\delta)(p-1)}) < \infty$ ensure the convergence of the (convex) empirical criterion $\psi_n(u; p)$, which entails the convergence of its minimizer. Finally, the estimator $\widehat{q}_{\tau_n}(p)$ has the same rate of convergence under the dependence condition $\mathcal{S}(\phi)$ as it has for independent observations, the price to pay for allowing our dependence setup being an enlarged asymptotic variance.
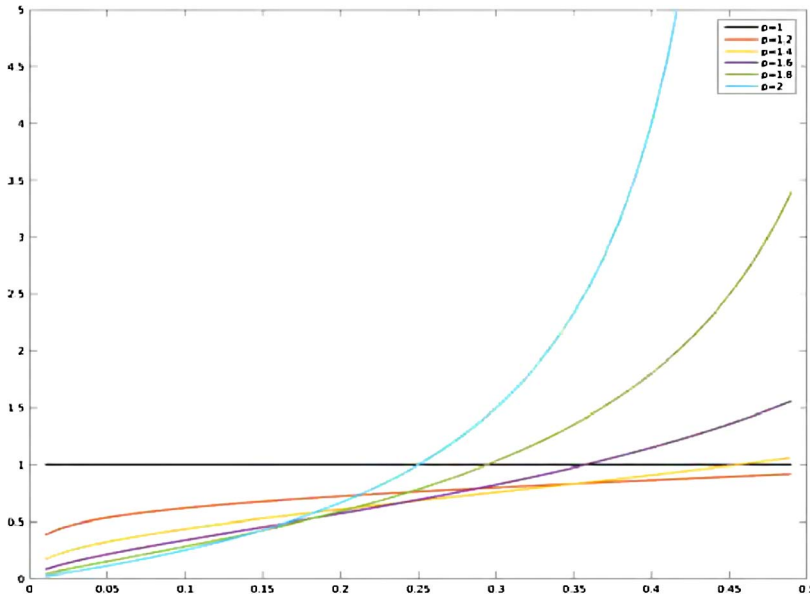
Figure 2. Asymptotic variance $\gamma \in (0, 1/2] \mapsto V(\gamma; p)$ for some values of $p \in [1, 2]$. Black line: $p = 1$; Red line: $p = 1.2$; Yellow line: $p = 1.4$; Purple line: $p = 1.6$; Green line: $p = 1.8$; Blue line: $p = 2$.

When the sequence $(X_1, X_2, \ldots)$ is independent and in the special cases $p \downarrow 1$ and $p = 2$, we recover the asymptotic normality of intermediate sample quantiles and expectiles, respectively, with asymptotic variances

$$V(\gamma; 1) = \frac{\Gamma(1)\Gamma(\gamma^{-1})}{\Gamma(1)\Gamma(\gamma^{-1})} = 1 \quad \text{and} \quad V(\gamma; 2) = \frac{\Gamma(3)\Gamma(\gamma^{-1} - 2)}{\Gamma(2)\Gamma(\gamma^{-1} - 1)} = \frac{2\gamma}{1 - 2\gamma}.$$

The behavior of the variance $\gamma \mapsto V(\gamma; p)$ is visualized in Figure 2 for some values of $p \in [1, 2]$, with $\gamma \in (0, 1/2]$. It can be seen in this Figure that for values of p close to but larger than 1, the asymptotic variance of the intermediate sample L^p-quantile is appreciably smaller than the asymptotic variance of the traditional sample quantile. In particular, values of p between 1.2 and 1.4 seem to yield estimators who may be more precise than the sample quantile in all usual applications (for which $\gamma \in (0, 1/2]$).

3.2. Extreme levels

We now discuss how to extrapolate intermediate L^p-quantile estimates of order $\tau_n \uparrow 1$, such that $n(1 - \tau_n) \to \infty$, to very extreme levels $\tau'_n \uparrow 1$ with $n(1 - \tau'_n) \to c < \infty$ as $n \to \infty$. The basic idea is to first use the regular variation condition $\mathcal{C}_1(\gamma)$ that entails the following classical

Weissman extrapolation formula for ordinary quantiles:

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

as τ_n and τ'_n approach 1 (Weissman [69]). The key argument is then to use the asymptotic equivalence

$$q_\tau(p) \sim C(\gamma; p) \cdot q_\tau(1) \qquad \text{as } \tau \uparrow 1, \tag{5}$$

shown in Corollary 1, to get the purely L^p-quantile approximation

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

This motivates us to define the estimator

$$\widehat{q}^{W}_{\tau'_n}(p) := \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(p) \tag{6}$$

for some $\sqrt{n(1-\tau_n)}$-consistent estimator $\widehat{\gamma}_n$ of $\gamma \equiv \gamma_r$, with $\widehat{q}_{\tau_n}(p)$ being the empirical least asymmetrically weighted L^p estimator of $q_{\tau_n}(p)$.

Theorem 2. *Assume that $p > 1$ and:*

- *condition $\boldsymbol{\mathcal{S}(\phi)}$ holds, with $\sum_{n=1}^{\infty} \sqrt{\phi(n)} < \infty$;*
- $\overline{F}$ *is strictly decreasing and satisfies $\boldsymbol{\mathcal{C}_2(\gamma_r, \rho, A)}$ with $\gamma_r < 1/[2(p-1)]$ and $\rho < 0$;*
- $\overline{F}_-$ *is either light-tailed or satisfies condition $\boldsymbol{\mathcal{C}_1(\gamma_\ell)}$;*
- $\gamma_\ell < 1/[2(p-1)]$ *in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- τ_n *and* $\tau'_n \uparrow 1$, *with* $n(1-\tau_n) \to \infty$ *and* $n(1-\tau'_n) \to c < \infty$;
- $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$, *for a suitable estimator $\widehat{\gamma}_n$ of γ_r and ζ a nondegenerate limiting random variable;*
- $\sqrt{n(1-\tau_n)} \max\{1-\tau_n,\ A((1-\tau_n)^{-1}),\ R_r(q_{\tau_n}(1), p, \gamma_r),\ R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\} = \mathrm{O}(1)$ *(in this bias condition the notation of Proposition 2 is used);*
- $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau'_n)] \to \infty$.

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\widehat{q}^{W}_{\tau'_n}(p)}{q_{\tau'_n}(p)} - 1\right) \xrightarrow{d} \zeta \qquad \text{as } n \to \infty.$$

Another option for estimating $q_{\tau'_n}(p)$ is by using directly its asymptotic connection (5) with $q_{\tau'_n}(1)$ to define the plug-in estimator

$$\widetilde{q}^{W}_{\tau'_n}(p) := C(\widehat{\gamma}_n; p)\widehat{q}^{W}_{\tau'_n}(1), \tag{7}$$

obtained by substituting in a $\sqrt{n(1-\tau_n)}$-consistent estimator $\widehat{\gamma}_n$ of γ and the traditional Weissman estimator

$$\widehat{q}_{\tau_n'}^W(1) = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(1), \qquad (8)$$

of the extreme quantile $q_{\tau_n'}(1)$, where $\widehat{q}_{\tau_n}(1) = X_{n-\lfloor n(1-\tau_n)\rfloor,n}$ and $\lfloor\cdot\rfloor$ denotes the floor function.

Theorem 3. *Assume that $p > 1$ and:*

- *$\overline{F}$ is strictly decreasing and satisfies $\mathcal{C}_2(\gamma_r, \rho, A)$ with $\gamma_r < 1/(p-1)$ and $\rho < 0$;*
- *$\overline{F}_-$ is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$;*
- *$\gamma_\ell < 1/(p-1)$ in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- *τ_n and $\tau_n' \uparrow 1$, with $n(1-\tau_n) \to \infty$ and $n(1-\tau_n') \to c < \infty$;*
- *$\sqrt{n(1-\tau_n)}(X_{n-\lfloor n(1-\tau_n)\rfloor,n}/q_{\tau_n}(1) - 1) = O_{\mathbb{P}}(1)$;*
- *$\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma_r) \overset{d}{\longrightarrow} \zeta$, for a suitable estimator $\widehat{\gamma}_n$ of γ_r and ζ a nondegenerate limiting random variable;*
- *$\sqrt{n(1-\tau_n)}\max\{1-\tau_n, A((1-\tau_n)^{-1}), R_r(q_{\tau_n}(1), p, \gamma_r), R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\} = O(1)$ (in this condition the notation of Proposition 2 is used);*
- *$\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$.*

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{q}_{\tau_n'}^W(p)}{q_{\tau_n'}(p)} - 1\right) \overset{d}{\longrightarrow} \zeta \qquad \text{as } n \to \infty.$$

Both these results, as well as the extrapolation results of the upcoming Sections 4 and 5, require a tail index estimator $\widehat{\gamma}_n$ such that $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma_r) \overset{d}{\longrightarrow} \zeta$ with ζ a limiting random variable with nondegenerate distribution. Under our dependence condition $\mathcal{S}(\phi)$, the sequence $(X_1, X_2, \dots)$ is in particular β-mixing, and it then follows from Drees [28] that, under further conditions on the β-mixing coefficients as well as regularity conditions on the tail of the underlying distribution and on the joint tail of bivariate margins, there exists a wide class of estimators $\widehat{\gamma}_n$ satisfying this convergence condition. In particular, it is mentioned in Drees [28], pages 625–626, that the Hill estimator (Hill [42]), the Pickands estimator (Pickands [56]), the moment estimator (Dekkers *et al.* [25]) and the maximum likelihood estimator in a generalized Pareto model are all part of this class; later, de Haan *et al.* [24] proved that a bias-reduced version of the Hill estimator is also $\sqrt{n(1-\tau_n)}$-consistent in this sense. Theorem 3 further requires that the empirical estimator $X_{n-\lfloor n(1-\tau_n)\rfloor,n}$ of $q_{\tau_n}(1)$ be $\sqrt{n(1-\tau_n)}$-consistent; under the regularity conditions of Drees [28], this is also true and $X_{n-\lfloor n(1-\tau_n)\rfloor,n}$ is in fact asymptotically Gaussian, see Theorem 2.1 therein. Finally, let us mention that this convergence condition on $X_{n-\lfloor n(1-\tau_n)\rfloor,n}$ is clearly satisfied for independent observations, see Theorem 2.4.1 in de Haan and Ferreira [23], page 50.

Our experience with simulated and real data indicates that, for non-negative loss distributions, the plug-in Weissman estimator $\widetilde{q}_{\tau_n'}^W(p)$ in (7) tends to be more efficient relative to the least

asymmetrically weighted L^p estimator $\widehat{q}_{\tau_n'}^W(p)$ in (6), for all values of $p > 1$. However, for real-valued profit-loss random variables, the least asymmetrically weighted L^p estimator $\widehat{q}_{\tau_n'}^W(p)$ is sometimes the winner following the values of p and γ. In particular, $\widehat{q}_{\tau_n'}^W(p = 2)$ appears to be superior to $\widetilde{q}_{\tau_n'}^W(p = 2)$ for all values of γ.

4. Recovering extreme quantiles from L^p-quantiles

The generalized L^p-quantiles do not have, for $p > 1$, an intuitive interpretation as direct as the original L^1-quantiles. If the statistician wishes to estimate tail L^p-quantiles $q_{\tau_n'}(p)$ that have the same probabilistic interpretation as a quantile $q_{\alpha_n}(1)$, with a given relative frequency α_n, then the extreme level τ_n' can be specified by the closed form expression (2), that is,

$$\tau_n'(p, \alpha_n; 1) = \frac{\mathbb{E}[|X - q_{\alpha_n}(1)|^{p-1}\mathbb{1}_{\{X \leq q_{\alpha_n}(1)\}}]}{\mathbb{E}[|X - q_{\alpha_n}(1)|^{p-1}]},$$

or equivalently

$$1 - \tau_n'(p, \alpha_n; 1) = \frac{\mathbb{E}[|X - q_{\alpha_n}(1)|^{p-1}\mathbb{1}_{\{X > q_{\alpha_n}(1)\}}]}{\mathbb{E}[|X - q_{\alpha_n}(1)|^{p-1}]}. \tag{9}$$

In order to manage extreme events, financial institutions and insurance companies are typically interested in tail probabilities $\alpha_n \to 1$ with $n(1 - \alpha_n) \to c$, a finite constant, as the sample size $n \to \infty$. For example, in the context of medical insurance data with 75,789 claims, Beirlant *et al.* [4], page 123, estimate the claim amount that will be exceeded on average only once in 100,000 cases. In the context of systemic risk measurement, Acharya *et al.* [2] handle once-in-a-decade events with one year of data, while Brownlees and Engle [11] and Cai *et al.* [12] consider once-per-decade systemic events with a data time horizon of ten years. In the context of the backtesting problem, which is crucial in the current Basel III regulatory framework, Chavez-Demoulin *et al.* [14] and Gong *et al.* [40] estimate quantiles exceeded on average once every 100 cases with sample sizes of the order of hundreds. Such examples are abundant especially in the extreme value literature.

The statistical problem is now to estimate the unknown extreme level $\tau_n'(p, \alpha_n; 1)$ from the available historical data. To this end, we first note that under condition $\mathcal{C}_1(\gamma_r)$ and if $\gamma_r < 1/(p - 1)$, then Proposition 1 entails

$$\frac{\overline{F}(q_{\tau_n'}(p))}{1 - \tau_n'} \sim \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \qquad \text{as } n \to \infty.$$

It then follows from $q_{\tau_n'}(p) \equiv q_{\alpha_n}(1)$ and $\overline{F}(q_{\alpha_n}(1)) = 1 - \alpha_n$ that

$$\frac{1 - \alpha_n}{1 - \tau_n'} \sim \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \qquad \text{as } n \to \infty.$$

Therefore $\tau_n' \equiv \tau_n'(p, \alpha_n; 1)$ satisfies the following asymptotic equivalence:

$$1 - \tau_n'(p, \alpha_n; 1) \sim (1 - \alpha_n) \frac{1}{\gamma_r} B\left(p, \frac{1}{\gamma_r} - p + 1\right) \qquad \text{as } n \to \infty.$$

Interestingly, $1 - \tau_n'(p, \alpha_n; 1)$ in (9) then asymptotically depends on the tail index γ_r but not on the actual value $q_{\alpha_n}(1)$ of the quantile itself. A natural estimator of $1 - \tau_n'(p, \alpha_n; 1)$ can now be defined by replacing, in its asymptotic approximation, the tail index γ_r by a $\sqrt{n(1 - \tau_n)}$-consistent estimator $\widehat{\gamma}_n$ as above, to get

$$\widehat{\tau}_n'(p, \alpha_n; 1) = 1 - (1 - \alpha_n) \frac{1}{\widehat{\gamma}_n} B\left(p, \frac{1}{\widehat{\gamma}_n} - p + 1\right). \tag{10}$$

Next, we derive the limiting distribution of $\widehat{\tau}_n'(p, \alpha_n; 1)$.

Theorem 4. *Assume that $p > 1$ and*:

- $\overline{F}$ *satisfies condition* $\mathcal{C}_2(\gamma_r, \rho, A)$;
- $\overline{F}_-$ *is either light-tailed or satisfies condition* $\mathcal{C}_1(\gamma_\ell)$;
- $\gamma_r < 1/(p - 1)$, *and* $\gamma_\ell < 1/(p - 1)$ *in case* $\overline{F}_-$ *is heavy-tailed.*

Assume further that

- τ_n *and* $\alpha_n \uparrow 1$, *with* $n(1 - \tau_n) \to \infty$;
- $\sqrt{n(1 - \tau_n)}(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$, *for a suitable estimator* $\widehat{\gamma}_n$ *of* γ_r *and* ζ *a nondegenerate limiting random variable*;
- $\sqrt{n(1 - \tau_n)} \max\{1 - \alpha_n, A((1 - \alpha_n)^{-1}), R_r(q_{\alpha_n}(1), p, \gamma_r), R_\ell(q_{\alpha_n}(1), p, \gamma_\ell)\} = O(1)$ (*in this condition the notation of Proposition* 2 *is used*).

Then:

$$\sqrt{n(1 - \tau_n)} \left(\frac{1 - \widehat{\tau}_n'(p, \alpha_n; 1)}{1 - \tau_n'(p, \alpha_n; 1)} - 1\right) = O_\mathbb{P}(1)$$

as $n \to \infty$. If actually

$$\sqrt{n(1 - \tau_n)} \max\left\{1 - \alpha_n, A\left((1 - \alpha_n)^{-1}\right), R_r\left(q_{\alpha_n}(1), p, \gamma_r\right), R_\ell\left(q_{\alpha_n}(1), p, \gamma_\ell\right)\right\} \to 0$$

then:

$$\sqrt{n(1 - \tau_n)} \left(\frac{1 - \widehat{\tau}_n'(p, \alpha_n; 1)}{1 - \tau_n'(p, \alpha_n; 1)} - 1\right)$$

$$\xrightarrow{d} -\left\{1 + \frac{1}{\gamma_r}\left[\Psi\left(\frac{1}{\gamma_r} - p + 1\right) - \Psi\left(\frac{1}{\gamma_r} + 1\right)\right]\right\} \frac{\zeta}{\gamma_r}$$

as $n \to \infty$, where $\Psi(x) = \Gamma'(x)/\Gamma(x)$ denotes the digamma function.

In practice, given a tail probability α_n and a power $p \in (1, 2]$, the extreme quantile $q_{\alpha_n}(1)$ can be estimated from the generalized L^p-quantile estimators $\widehat{q}_{\tau_n'}^W(p)$ and $\widetilde{q}_{\tau_n'}^W(p)$ in two steps: first,

estimate $\tau'_n \equiv \tau'_n(p, \alpha_n; 1)$ by $\widehat{\tau}'_n(p, \alpha_n; 1)$ and, second, use the estimators $\widehat{q}^W_{\tau'_n}(p)$ and $\widetilde{q}^W_{\tau'_n}(p)$ as if τ'_n were known, by substituting the estimated value $\widehat{\tau}'_n(p, \alpha_n; 1)$ in place of τ'_n, yielding the following two extreme quantile estimators:

$$\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p) = \left(\frac{1 - \widehat{\tau}'_n(p, \alpha_n; 1)}{1 - \tau_n} \right)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(p)$$

and

$$\widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p) = C(\widehat{\gamma}_n; p)\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(1).$$

This is actually a two-stage estimation procedure in the sense that the intermediate level τ_n used in the first stage to compute $\widehat{\tau}'_n(p, \alpha_n; 1)$ need not be the same as the intermediate levels used in the second stage to compute the extrapolated L^p-quantile estimators $\widehat{q}^W_{\tau'_n}(p)$ and $\widetilde{q}^W_{\tau'_n}(p)$. Detailed practical guidelines to implement efficiently the final composite estimates $\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p)$ and $\widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p)$ are provided in Section 7 through a real data example. For the sake of simplicity, we do not emphasize in the asymptotic results below the distinction between the intermediate level used in the first stage and those used in the second stage. It should be, however, noted that when the estimation procedure is carried out in one single step instead, that is, with the same intermediate level in both $\widehat{\tau}'_n(p, \alpha_n; 1)$ and the extrapolated L^p-quantile estimators, then the composite version $\widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p)$ is nothing but the Weissman quantile estimator $\widehat{q}^W_{\alpha_n}(1)$. Indeed, in that case, we have by (8) and the definition of $C(\cdot, \cdot)$ below Corollary 1 that

$$\widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p) = C(\widehat{\gamma}_n; p)\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(1)$$

$$= \left[\frac{\widehat{\gamma}_n}{B(p, \widehat{\gamma}_n^{-1} - p + 1)} \right]^{-\widehat{\gamma}_n} \left(\frac{1 - \widehat{\tau}'_n(p, \alpha_n; 1)}{1 - \tau_n} \right)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(1)$$

$$= \left[\frac{\widehat{\gamma}_n}{B(p, \widehat{\gamma}_n^{-1} - p + 1)} \cdot \frac{(1 - \alpha_n)^{\frac{1}{\widehat{\gamma}_n}} B(p, \frac{1}{\widehat{\gamma}_n} - p + 1)}{1 - \tau_n} \right]^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(1)$$

$$= \left[\frac{1 - \alpha_n}{1 - \tau_n} \right]^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(1) \equiv \widehat{q}^W_{\alpha_n}(1).$$

Our next two convergence results examine the asymptotic properties of the two composite estimators $\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p)$ and $\widetilde{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p)$. We first consider the estimator $\widehat{q}^W_{\widehat{\tau}'_n(p,\alpha_n;1)}(p)$.

Theorem 5. *Assume that $p > 1$ and:*

- *condition $\mathcal{S}(\phi)$ holds, with $\sum_{n=1}^{\infty} \sqrt{\phi(n)} < \infty$;*
- *$\overline{F}$ is strictly decreasing and satisfies $\mathcal{C}_2(\gamma_r, \rho, A)$ with $\gamma_r < 1/[2(p-1)]$ and $\rho < 0$;*
- *F_- is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$;*
- *$\gamma_\ell < 1/[2(p-1)]$ in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- τ_n and $\alpha_n \uparrow 1$, with $n(1 - \tau_n) \to \infty$ and $n(1 - \alpha_n) \to c < \infty$;
- $\sqrt{n(1 - \tau_n)}(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$, *for a suitable estimator* $\widehat{\gamma}_n$ *of* γ_r *and* ζ *a nondegenerate limiting random variable*;
- $\sqrt{n(1 - \tau_n)} \max\{1 - \tau_n, A((1 - \tau_n)^{-1}), R_r(q_{\tau_n}(1), p, \gamma_r), R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\} = \mathrm{O}(1)$ (*in this condition the notation of Proposition 2 is used*);
- $\sqrt{n(1 - \tau_n)}/\log[(1 - \tau_n)/(1 - \alpha_n)] \to \infty$.

Then

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \alpha_n)]}\left(\frac{\widehat{q}^W_{\tau_n'(p,\alpha_n;1)}(p)}{q_{\alpha_n}(1)} - 1\right) \xrightarrow{d} \zeta \qquad as\ n \to \infty.$$

As regards the alternative extrapolated estimator $\widetilde{q}^W_{\tau_n'(p,\alpha_n;1)}(p)$, we have the following asymptotic result.

Theorem 6. *Assume that $p > 1$ and*:

- $\overline{F}$ *is strictly decreasing and satisfies* $\mathcal{C}_2(\boldsymbol{\gamma_r}, \boldsymbol{\rho}, \boldsymbol{A})$ *with* $\gamma_r < 1/(p - 1)$ *and* $\rho < 0$;
- $\overline{F}_-$ *is either light-tailed or satisfies condition* $\mathcal{C}_1(\boldsymbol{\gamma_\ell})$;
- $\gamma_\ell < 1/(p - 1)$ *in case* $\overline{F}_-$ *is heavy-tailed.*

Assume further that

- τ_n and $\alpha_n \uparrow 1$, with $n(1 - \tau_n) \to \infty$ and $n(1 - \alpha_n) \to c < \infty$;
- $\sqrt{n(1 - \tau_n)}(X_{n-\lfloor n(1-\tau_n)\rfloor, n}/q_{\tau_n}(1) - 1) = \mathrm{O}_\mathbb{P}(1)$;
- $\sqrt{n(1 - \tau_n)}(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$, *for a suitable estimator* $\widehat{\gamma}_n$ *of* γ_r *and* ζ *a nondegenerate limiting random variable*;
- $\sqrt{n(1 - \tau_n)} \max\{1 - \tau_n, A((1 - \tau_n)^{-1}), R_r(q_{\tau_n}(1), p, \gamma_r), R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\} = \mathrm{O}(1)$ (*in this condition the notation of Proposition 2 is used*);
- $\sqrt{n(1 - \tau_n)}/\log[(1 - \tau_n)/(1 - \alpha_n)] \to \infty$.

Then

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \alpha_n)]}\left(\frac{\widetilde{q}^W_{\tau_n'(p,\alpha_n;1)}(p)}{q_{\alpha_n}(1)} - 1\right) \xrightarrow{d} \zeta \qquad as\ n \to \infty.$$

Just as in the previous section, we note that Theorems 4, 5 and 6 hold true in the dependence framework $\mathcal{S}(\boldsymbol{\phi})$, for a wide class of estimators $\widehat{\gamma}_n$, under further conditions on the β-mixing coefficients as well as regularity conditions on the tail of the underlying distribution and on the joint tail of bivariate margins, see Drees [28].

The analysis above is concerned with the heavy right-tails ($\alpha_n \to 1$) of non-negative loss distributions as well as real-valued profit-loss random variables. Similar considerations evidently apply when the focus is on the heavy left-tail ($\alpha_n \to 0$) of a series of financial returns. In this case, the problem translates into estimating the quantile $-q_{1-\alpha_n}(1)$ with the sign convention for values of X as the negative of returns. A comparison and validation on financial time series in Section 7 shows that the two-stage estimation procedure may afford more accurate estimates $\widehat{q}^W_{\tau_n'(p,\alpha_n;1)}(p)$ and $\widetilde{q}^W_{\tau_n'(p,\alpha_n;1)}(p)$ of $q_{\alpha_n}(1)$ than the traditional Weissman estimator $\widehat{q}^W_{\alpha_n}(1)$ defined in (8).

5. Recovering extreme expectiles from L^p-quantiles

In this section, we focus on L^2-quantiles, or equivalently expectiles, which define the only M-quantiles that are coherent risk measures, and we assume therefore that $\mathbb{E}(X_-) < \infty$ and $\gamma_r < 1$ to guarantee their existence. We consider extrapolated estimation of tail expectiles $q_{\alpha_n}(2)$, where $\alpha_n \uparrow 1$ and $n(1 - \alpha_n) \to c < \infty$ as $n \to \infty$. The first presented asymmetric least squares estimator $\widehat{q}_{\alpha_n}^W(2)$ in (6) reads as

$$\widehat{q}_{\alpha_n}^W(2) = \left(\frac{1 - \alpha_n}{1 - \tau_n}\right)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}(2), \tag{11}$$

where $\widehat{q}_{\tau_n}(2)$ is defined in (3) with $p = 2$. The second plug-in Weissman estimator $\widetilde{q}_{\alpha_n}^W(2)$, described in (7), translates into

$$\widetilde{q}_{\alpha_n}^W(2) := C(\widehat{\gamma}_n; 2)\widehat{q}_{\alpha_n}^W(1) \equiv \left(\widehat{\gamma}_n^{-1} - 1\right)^{-\widehat{\gamma}_n} \widehat{q}_{\alpha_n}^W(1), \tag{12}$$

where $\widehat{q}_{\alpha_n}^W(1)$ is the classical Weissman quantile estimator given in (8). The asymptotic properties of both extreme expectile estimators $\widehat{q}_{\alpha_n}^W(2)$ and $\widetilde{q}_{\alpha_n}^W(2)$ had been already established in Daouia *et al.* [18] for independent observations. It was also found there that $\widehat{q}_{\alpha_n}^W(2)$ is superior to $\widetilde{q}_{\alpha_n}^W(2)$ in the case of real-valued profit-loss random variables, while $\widetilde{q}_{\alpha_n}^W(2)$ essentially is the winner in the case of non-negative loss distributions. Here, we suggest novel extrapolated estimators that might be more efficient than $\widehat{q}_{\alpha_n}^W(2)$ and $\widetilde{q}_{\alpha_n}^W(2)$ themselves. The first basic tool is the following asymptotic connection between the extreme expectile $q_{\alpha_n}(2)$ and its L^p-quantile analogue $q_{\alpha_n}(p)$:

$$q_{\alpha_n}(2) \sim C(\gamma_r; 2) \cdot q_{\alpha_n}(1) \qquad \text{as } \alpha_n \uparrow 1$$
$$\sim C(\gamma_r; 2) \cdot C^{-1}(\gamma_r; p) \cdot q_{\alpha_n}(p) \qquad \text{as } \alpha_n \uparrow 1,$$

when $p > 1$ is such that $\gamma_r < 1/(p - 1)$ [in particular, this is true for any $p \in (1, 2]$ since it is assumed here that $\gamma_r < 1$]. This asymptotic equivalence follows immediately by applying Corollary 1 twice. One may then define the alternative estimator

$$\breve{q}_{\alpha_n}^p(2) := C(\widehat{\gamma}_n; 2) \cdot C^{-1}(\widehat{\gamma}_n; p) \cdot \widehat{q}_{\alpha_n}^W(p) \tag{13}$$

$$\equiv \left(\widehat{\gamma}_n^{-1} - 1\right)^{-\widehat{\gamma}_n} \left[\frac{\widehat{\gamma}_n}{B(p, \widehat{\gamma}_n^{-1} - p + 1)}\right]^{\widehat{\gamma}_n} \widehat{q}_{\alpha_n}^W(p),$$

obtained by substituting in a $\sqrt{n(1 - \tau_n)}$-consistent estimator $\widehat{\gamma}_n$ of γ_r and the extrapolated version $\widehat{q}_{\alpha_n}^W(p)$ of the least asymmetrically weighted L^p-quantile estimator, given in (6). The idea is therefore to exploit the accuracy of the asymptotic connection between population L^p-quantiles and traditional quantiles in conjunction with the superiority of sample L^p-quantiles in terms of finite-sample performance. Note that by replacing $\widehat{q}_{\alpha_n}^W(p)$ in (13) with the plug-in estimator $\widetilde{q}_{\alpha_n}^W(p)$ introduced in (7), we recover the estimator $\widetilde{q}_{\alpha_n}^W(2)$ described in (12).

Theorem 7. *Pick $p \in (1, 2]$. Assume that*:

- *condition $\mathcal{S}(\phi)$ holds, with $\sum_{n=1}^{\infty} \sqrt{\phi(n)} < \infty$*;
- $\overline{F}$ *is strictly decreasing and satisfies $\mathcal{C}_2(\gamma_r, \rho, A)$ with $\gamma_r < 1/\max[1, 2(p-1)]$ and $\rho < 0$*;
- $\overline{F}_-$ *is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$*;
- $\gamma_\ell < 1/\max[1, 2(p-1)]$ *in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- τ_n *and* $\alpha_n \uparrow 1$*, with* $n(1 - \tau_n) \to \infty$ *and* $n(1 - \alpha_n) \to c < \infty$;
- $\sqrt{n(1 - \tau_n)}(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$*, for a suitable estimator $\widehat{\gamma}_n$ of γ_r and ζ a nondegenerate limiting random variable*;
- $\sqrt{n(1 - \tau_n)} \max\{1 - \tau_n, A((1 - \tau_n)^{-1}), R_r(q_{\tau_n}(1), p, \gamma_r), R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\} = O(1)$ *(in this condition the notation of Proposition 2 is used)*;
- $\sqrt{n(1 - \tau_n)}/\log[(1 - \tau_n)/(1 - \alpha_n)] \to \infty$.

Then

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \alpha_n)]}\left(\frac{\breve{q}^p_{\alpha_n}(2)}{q_{\alpha_n}(2)} - 1\right) \xrightarrow{d} \zeta \qquad \text{as } n \to \infty.$$

As a matter of fact, $\breve{q}^p_{\alpha_n}(2)$ approaches $\widehat{q}^W_{\alpha_n}(2)$ when p tends to 2, whereas it approaches $\widetilde{q}^W_{\alpha_n}(2)$ when p tends to 1. In practice, as suggested by our experiments with simulated data in Appendix C.1 of the supplementary material document [19], we favor the use of $\breve{q}^p_{\alpha_n}(2)$ with p very close to 1 for non-negative loss distributions, and with p very close to 2 for real-valued profit-loss random variables. It is in the latter case that $\breve{q}^p_{\alpha_n}(2)$ may appear to be appreciably more efficient relatively to both estimators $\widehat{q}^W_{\alpha_n}(2)$ and $\widetilde{q}^W_{\alpha_n}(2)$, especially for profit-loss distributions with long tails.

Another way of recovering expectiles from L^p-quantiles is by proceeding as in the previous section in the case of ordinary quantiles. To estimate the extreme expectile $q_{\alpha_n}(2)$, the idea is to use a tail L^p-quantile $q_{\tau'_n}(p)$ which coincides with (and therefore has the same interpretation as) $q_{\alpha_n}(2)$. Given α_n and the power p, the level τ'_n such that $q_{\tau'_n}(p) \equiv q_{\alpha_n}(2)$ has the explicit expression

$$\tau'_n(p, \alpha_n; 2) = \frac{\mathbb{E}[|X - q_{\alpha_n}(2)|^{p-1}\mathbb{1}_{\{X \leq q_{\alpha_n}(2)\}}]}{\mathbb{E}[|X - q_{\alpha_n}(2)|^{p-1}]} \tag{14}$$

in view of (1). This closed form of $\tau'_n \equiv \tau'_n(p, \alpha_n; 2)$ depends heavily on $q_{\alpha_n}(2)$, but for any $p > 1$ such that $\gamma_r < 1/(p-1)$, condition $\mathcal{C}_1(\gamma_r)$ and Proposition 1 entail that

$$\frac{\overline{F}(q_{\tau'_n}(p))}{1 - \tau'_n} \sim \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \qquad \text{as } n \to \infty.$$

It follows from $q_{\tau'_n}(p) \equiv q_{\alpha_n}(2)$ that

$$\frac{\overline{F}(q_{\alpha_n}(2))}{1 - \tau'_n} \sim \frac{\gamma_r}{B(p, \gamma_r^{-1} - p + 1)} \qquad \text{as } n \to \infty.$$

We also have by Theorem 11 in Bellini *et al.* [6] that

$$\overline{F}\big(q_{\alpha_n}(2)\big) \sim (1 - \alpha_n)\big(\gamma_r^{-1} - 1\big) \qquad \text{as } n \to \infty.$$

Therefore τ_n' in (14) satisfies the asymptotic equivalence

$$1 - \tau_n'(p, \alpha_n; 2) \sim (1 - \alpha_n)\big(\gamma_r^{-1} - 1\big)\frac{1}{\gamma_r}B\left(p, \frac{1}{\gamma_r} - p + 1\right) \qquad \text{as } n \to \infty.$$

By substituting a $\sqrt{n(1 - \tau_n)}$-consistent estimator $\widehat{\gamma}_n$ in place of the tail index γ_r, we obtain the following estimator of $\tau_n'(p, \alpha_n; 2)$:

$$\widehat{\tau}_n'(p, \alpha_n; 2) := 1 - (1 - \alpha_n)\big(\widehat{\gamma}_n^{-1} - 1\big)\frac{1}{\widehat{\gamma}_n}B\left(p, \frac{1}{\widehat{\gamma}_n} - p + 1\right). \tag{15}$$

Finally, one may estimate the extreme expectile $q_{\alpha_n}(2) \equiv q_{\tau_n'(p,\alpha_n;2)}(p)$ by the following composite L^p-quantile estimators

$$\widehat{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(p) = \left(\frac{1 - \widehat{\tau}_n'(p, \alpha_n; 2)}{1 - \tau_n}\right)^{-\widehat{\gamma}_n}\widehat{q}_{\tau_n}(p) \tag{16}$$

and

$$\widetilde{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(p) = C(\widehat{\gamma}_n; p)\widehat{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(1), \tag{17}$$

obtained by replacing τ_n' in $\widehat{q}^W_{\tau_n'}(p)$ and $\widetilde{q}^W_{\tau_n'}(p)$ with $\widehat{\tau}_n'(p, \alpha_n; 2)$. It is remarkable that these two estimators are intimately linked to those of Section 4, since

$$\widehat{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(p) = C(\widehat{\gamma}_n; p)\widehat{q}^W_{\widehat{\tau}_n'(p,\alpha_n;1)}(p) \quad \text{and} \quad \widetilde{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(p) = C(\widehat{\gamma}_n; p)\widetilde{q}^W_{\widehat{\tau}_n'(p,\alpha_n;1)}(p).$$

We first unravel the limit distribution of the extrapolated estimator $\widehat{q}^W_{\widehat{\tau}_n'(p,\alpha_n;2)}(p)$.

Theorem 8. *Pick $p \in (1, 2]$. Assume that*:

- *condition $\mathcal{S}(\phi)$ holds, with $\sum_{n=1}^{\infty}\sqrt{\phi(n)} < \infty$;*
- *$\overline{F}$ is strictly decreasing and satisfies $\mathcal{C}_2(\gamma_r, \rho, A)$ with $\gamma_r < 1/\max[1, 2(p-1)]$ and $\rho < 0$;*
- *$\overline{F}_-$ is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$;*
- *$\gamma_\ell < 1/\max[1, 2(p-1)]$ in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- *τ_n and $\alpha_n \uparrow 1$, with $n(1 - \tau_n) \to \infty$ and $n(1 - \alpha_n) \to c < \infty$;*
- *$\sqrt{n(1 - \tau_n)}(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$, for a suitable estimator $\widehat{\gamma}_n$ of γ_r and ζ a nondegenerate limiting random variable;*
- *$\sqrt{n(1 - \tau_n)}\max\{1 - \tau_n, A((1 - \tau_n)^{-1}), R_r(q_{\tau_n}(1), p, \gamma_r), R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\} = O(1)$ (in this condition the notation of Proposition 2 is used);*

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

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widehat{q}^W_{\tau'_n(p,\alpha_n;2)}(p)}{q_{\alpha_n}(2)}-1\right) \xrightarrow{d} \zeta \qquad as \ n \to \infty.$$

Next, we derive the asymptotic distribution of the composite estimator $\widetilde{q}^W_{\tau'_n(p,\alpha_n;2)}(p)$.

Theorem 9. *Pick $p \in (1, 2]$. Assume that*:

- $\overline{F}$ *is strictly decreasing and satisfies $\mathcal{C}_2(\gamma_r, \rho, A)$ with $\gamma_r < 1$ and $\rho < 0$;*
- $\overline{F}_-$ *is either light-tailed or satisfies condition $\mathcal{C}_1(\gamma_\ell)$;*
- $\gamma_\ell < 1$ *in case $\overline{F}_-$ is heavy-tailed.*

Assume further that

- τ_n *and $\alpha_n \uparrow 1$, with $n(1-\tau_n) \to \infty$ and $n(1-\alpha_n) \to c < \infty$;*
- $\sqrt{n(1-\tau_n)}(X_{n-\lfloor n(1-\tau_n)\rfloor,n}/q_{\tau_n}(1) - 1) = O_{\mathbb{P}}(1)$;
- $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n - \gamma_r) \xrightarrow{d} \zeta$, *for a suitable estimator $\widehat{\gamma}_n$ of γ_r and ζ a nondegenerate limiting random variable;*
- $\sqrt{n(1-\tau_n)}\max\{1-\tau_n, A((1-\tau_n)^{-1}), R_r(q_{\tau_n}(1), p, \gamma_r), R_\ell(q_{\tau_n}(1), p, \gamma_\ell)\} = O(1)$ (*in this condition the notation of Proposition 2 is used*);
- $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\alpha_n)] \to \infty$.

Then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\widetilde{q}^W_{\tau'_n(p,\alpha_n;2)}(p)}{q_{\alpha_n}(2)}-1\right) \xrightarrow{d} \zeta \qquad as \ n \to \infty.$$

Our experience with simulated data in Appendix C.2 of the supplementary material document [19] indicates that $\widehat{q}^W_{\tau'_n(p,\alpha_n;2)}(p)$ in (16) behaves very similarly to $\breve{q}^p_{\alpha_n}(2)$ in (13). In particular, $\widehat{q}^W_{\tau'_n(p,\alpha_n;2)}(p)$ exhibits better accuracy relative to both rival estimators $\widehat{q}^W_{\alpha_n}(2)$ and $\widetilde{q}^W_{\alpha_n}(2)$ in the important case of profit-loss distributions with heavier tails. By contrast, the second composite estimator $\widetilde{q}^W_{\tau'_n(p,\alpha_n;2)}(p)$ in (17) does not bring Monte Carlo evidence of any added value with respect to the benchmark estimators $\widehat{q}^W_{\alpha_n}(2)$ and $\widetilde{q}^W_{\alpha_n}(2)$.

6. Some simulation evidence

To evaluate the finite-sample performance of the L^p-quantile estimators described above, we have undertaken some simulation experiments. The experiments all employ the Pareto distribution $F(x) = 1 - x^{-1/\gamma}$, $x > 1$, the Fréchet distribution $F(x) = e^{-x^{-1/\gamma}}$, $x > 0$, and the Student t-distribution with degree of freedom $1/\gamma$. The accuracy of the estimators is assessed through the Relative Mean-Squared Error (RMSE) and the bias computed over 3000 replications. Most of the error is due to variance, the squared bias being typically much smaller. We present mainly

the RMSE estimates to save space. All the experiments here have sample size $n = 200$. Further simulation results about extreme expectile estimation are discussed in Appendices C.1 and C.2 of the supplementary material document [19]. We also investigate the normality of some presented extreme L^p-quantile estimators in Supplement C.3 in [19], where the QQ-plots indicate that our limit theorems provide adequate approximations for finite sample sizes.

6.1. Which L^p-quantiles can accurately be estimated?

To answer this first question we compare the least asymmetrically weighted L^p estimators $\widehat{q}_\tau(p)$ of $q_\tau(p)$, with two intermediate levels $\tau = 0.9$ and $\tau = 0.95$, for different values of p. The obtained Monte Carlo estimates are graphed in Figure 3 for $p \in \{1, 1.05, \ldots, 2\}$ and $\gamma \in \{0.1, 0.15, \ldots, 0.45\}$. Not surprisingly, the quality of the estimation deteriorates when γ increases. In particular, for large values of γ (say $\gamma \geq 0.2$), the expectile estimation appears to be the worst as the RMSE achieves its maximum at $p = 2$. In contrast, for these particularly large values of γ (although this seems to be true for smaller γ as well), the estimation accuracy is clearly higher for small values of p, say $p \leq 1.45$. Also, we see that the value of p minimizing the RMSE depends heavily on γ. Yet, the choice of $p \in (1.2, 1.6)$ seems to be a good global compromise.
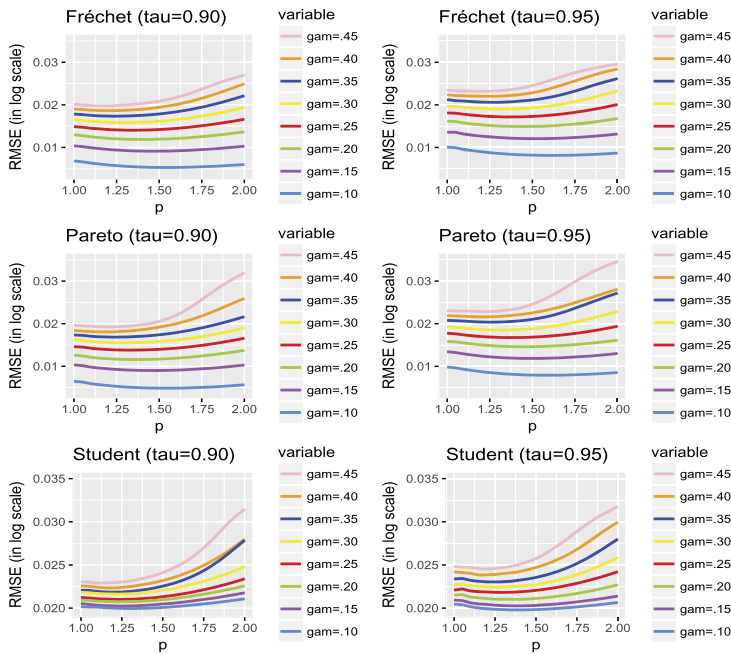

Figure 3. Relative MSE (in log scale) as a function of p, for different values of γ. From left to right, $\tau = 0.9, 0.95$. From top to bottom, Fréchet, Pareto and Student distributions.

This conclusion is, however, no longer valid when it comes to estimate extreme L^p-quantiles $q_{\tau_n'}(p)$ with, for instance, $\tau_n' \in [1 - \frac{1}{n}, 1)$. To see this, we compare the extrapolated least asymmetrically weighted L^p estimators $\widehat{q}_{\tau_n'}^W(p)$ in (6) and the plug-in Weissman estimators $\widetilde{q}_{\tau_n'}^W(p)$ in (7). The experiments all employ $\tau_n' = 1 - 1/n$ and various values of p in $[1, 2]$. We also used here the intermediate level $\tau_n = 1 - k/n$ and the Hill estimator $\widehat{\gamma}_n = \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n-i+1,n}}{X_{n-k,n}}$ of the tail index γ (see Hill [42]). The number k can be viewed as the effective sample size for tail extrapolation.

The evolution of the RMSE of the two classes of estimators $\{\widehat{q}_{\tau_n'}^W(p)\}_p$ and $\{\widetilde{q}_{\tau_n'}^W(p)\}_p$ in terms of the value k is displayed in Figures 4, 5 and 6 for the Fréchet, Pareto and Student distributions, respectively. To save space, we show only the Monte Carlo estimates obtained for the tail index
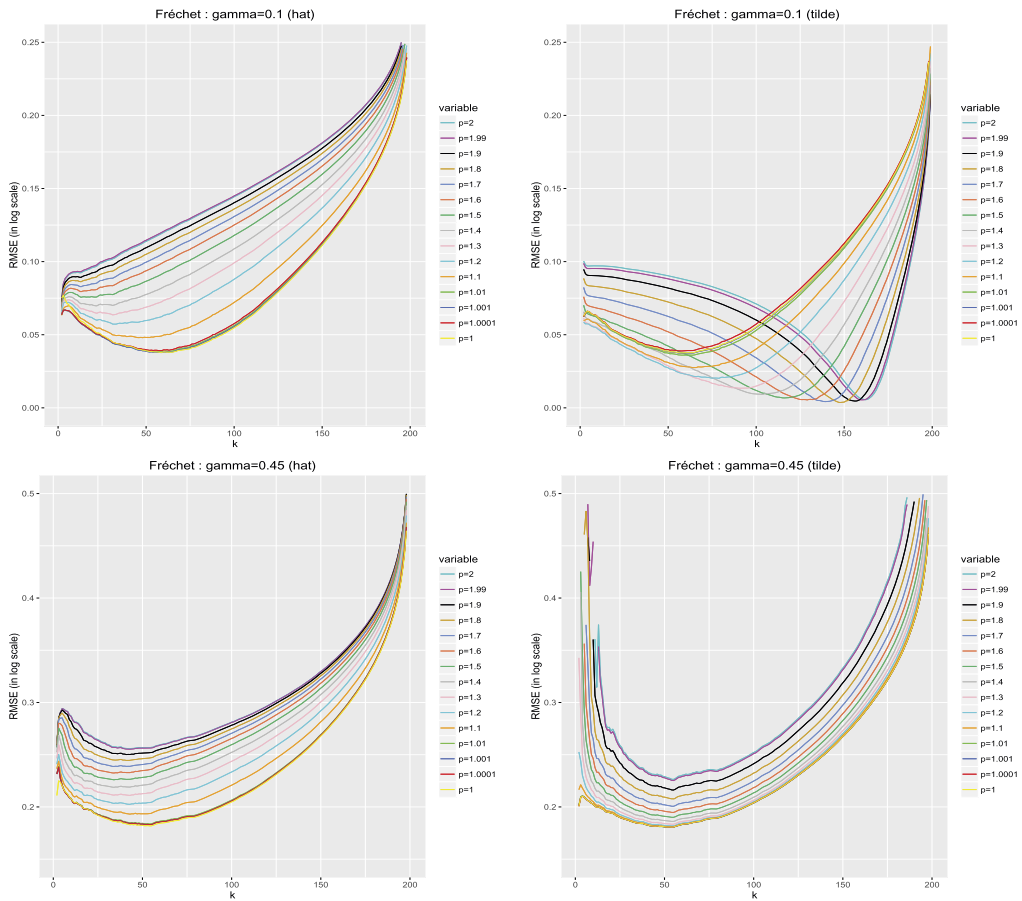
Figure 4. Fréchet distribution–RMSE (in log scale) of $\widehat{q}_{\tau_n'}^W(p)$ in left panels and $\widetilde{q}_{\tau_n'}^W(p)$ in right panels. From top to bottom: $\gamma = 0.1, 0.45$.
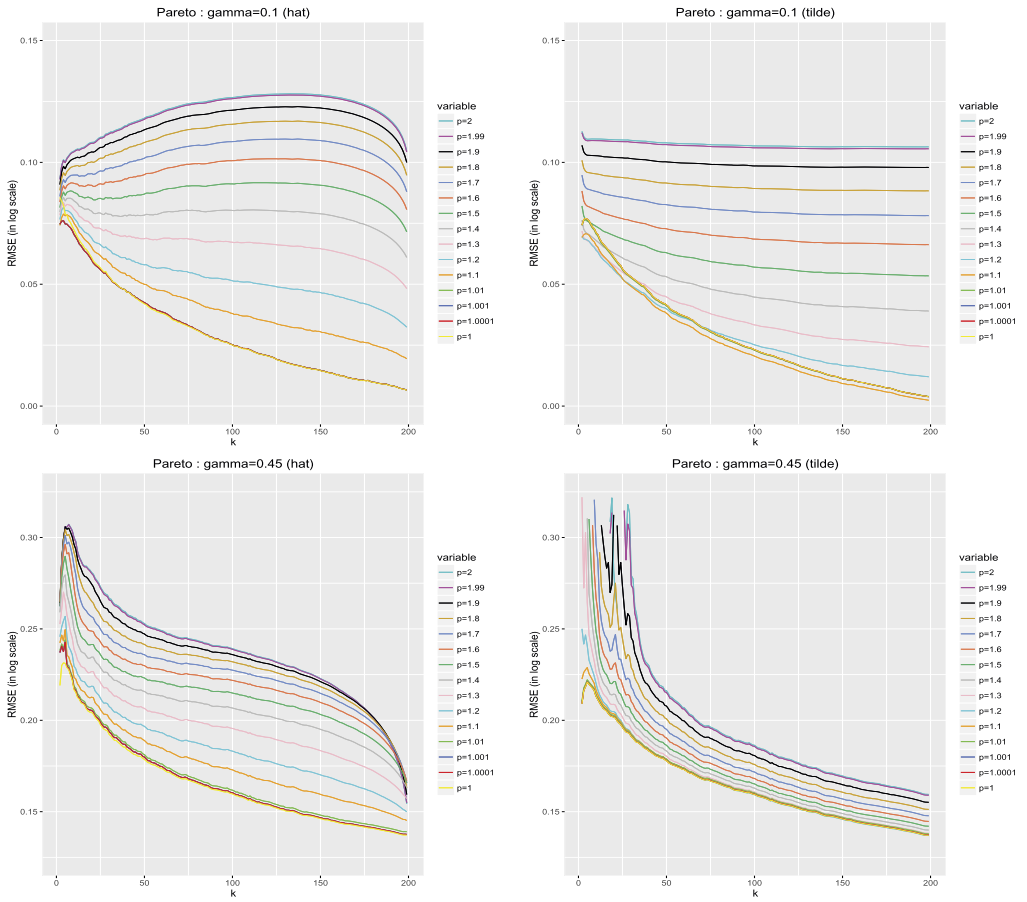
Figure 5. Pareto distribution–RMSE (in log scale) of $\widehat{q}^{W}_{\tau'_n}(p)$ in left panels and $\widetilde{q}^{W}_{\tau'_n}(p)$ in right panels. From top to bottom: $\gamma = 0.1, 0.45$.

values $\gamma = 0.1$ (top panels) and $\gamma = 0.45$ (bottom panels). It may be seen that both extreme L^p-quantile estimators $\{\widehat{q}^{W}_{\tau'_n}(p)\}_p$, in the left panels, and $\{\widetilde{q}^{W}_{\tau'_n}(p)\}_p$, in the right panels, attain more accuracy for $p \in [1, 1.3]$ or $p \in [1.7, 2]$. This can also be observed from Figure 12 where the RMSE is graphed as function of the power p in dashed red for $\widehat{q}^{W}_{\tau'_n}(p)$ and in dashed blue for $\widetilde{q}^{W}_{\tau'_n}(p)$, with k being chosen optimally so as to minimize the RMSE of each estimator.

It should also be emphasized that, in 8 cases among the 12 pictures in Figures 4–6, the best accuracy is not achieved at $p = 1$ or $p = 2$, but at in between values: a zoom in on some pictures where the best accuracy is achieved with values of $p \notin \{1, 2\}$ is given in Figure 7. We shall discuss below the important question of how to pick out p in practice in order to get the most accurate extreme L^p-quantile estimates from a forecasting perspective.
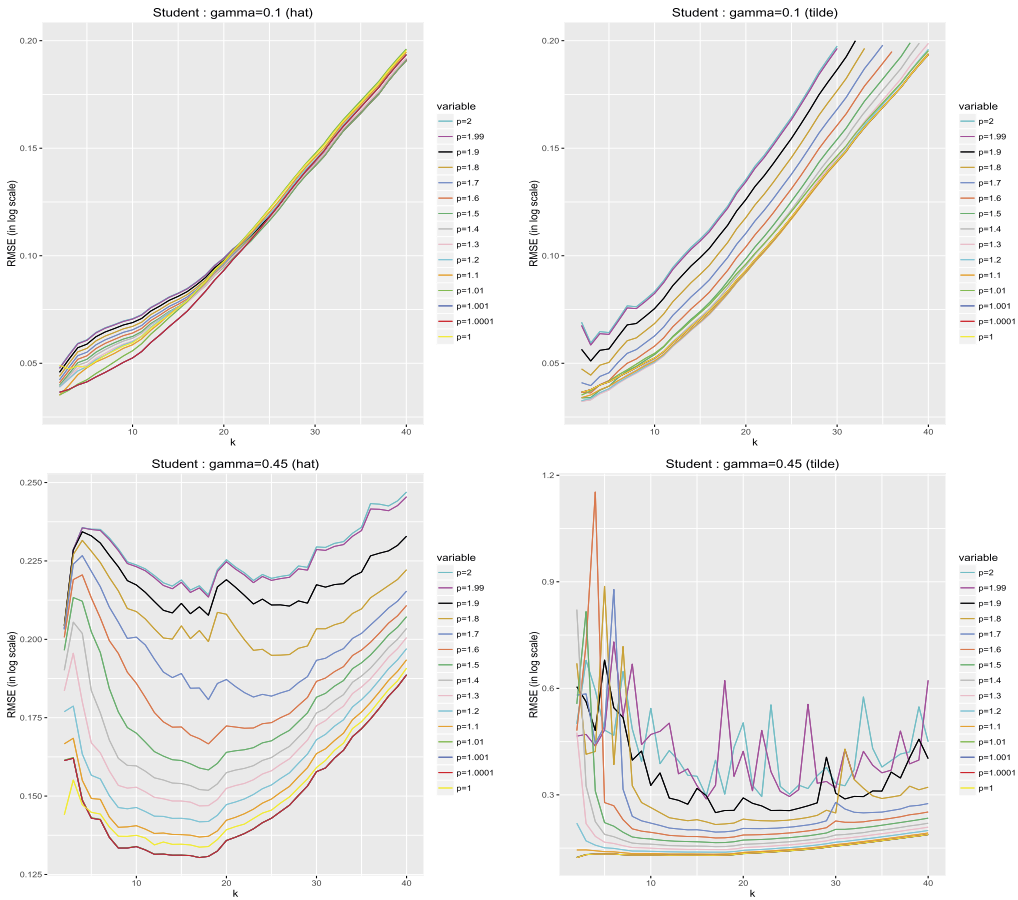
Figure 6. Student distribution–RMSE (in log scale) of $\widehat{q}_{\tau_n'}^{W}(p)$ in left panels and $\widetilde{q}_{\tau_n'}^{W}(p)$ in right panels. From top to bottom: $\gamma = 0.1, 0.45$.

6.2. Which extreme L^p-quantile estimator: $\widehat{q}_{\tau_n'}^{W}(p)$ or $\widetilde{q}_{\tau_n'}^{W}(p)$?

Based on the experiments above, we would like to comment here on the performance of the least asymmetrically weighted L^p estimator $\widehat{q}_{\tau_n'}^{W}(p)$ in comparison with the plug-in Weissman estimator $\widetilde{q}_{\tau_n'}^{W}(p)$, for each fixed value of $p \in (1, 2]$.

In the Fréchet and Pareto cases that correspond to non-negative random variables, it may be seen from Figures 4 and 5 that $\widetilde{q}_{\tau_n'}^{W}(p)$, in the right panels, behaves almost overall better than $\widehat{q}_{\tau_n'}^{W}(p)$ in the left panels. This can be visualized more clearly in Figures 8 and 9 for three chosen values of $p \in \{1.2, 1.5, 1.8\}$. This may also be seen from Figure 12 where the RMSE is plotted against p (in dashed lines) for k chosen to minimize the RMSE.
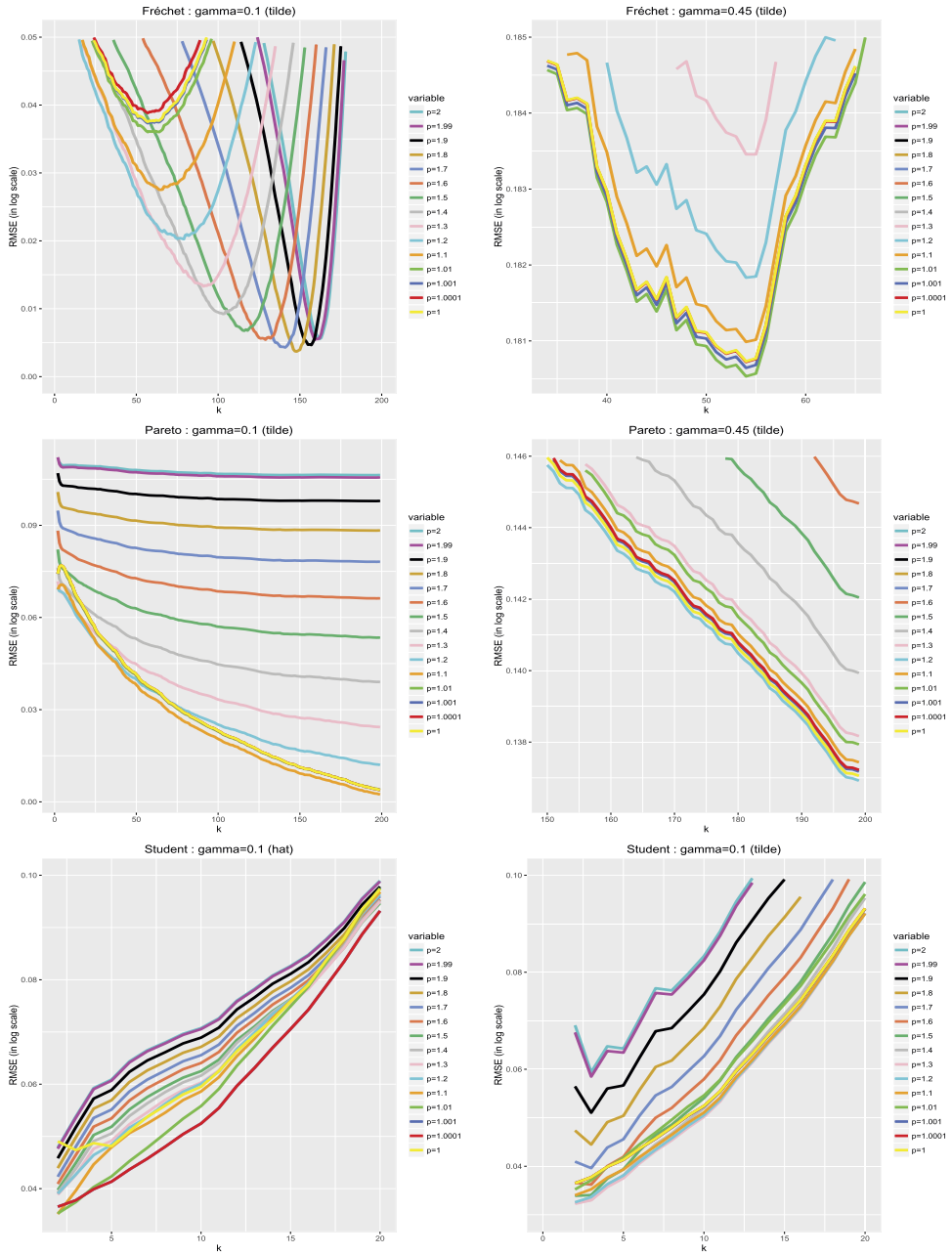
Figure 7. Zoom in on some pictures in Figures 4, 5 and 6, where the best accuracy corresponds to values of $p \notin \{1, 2\}$.
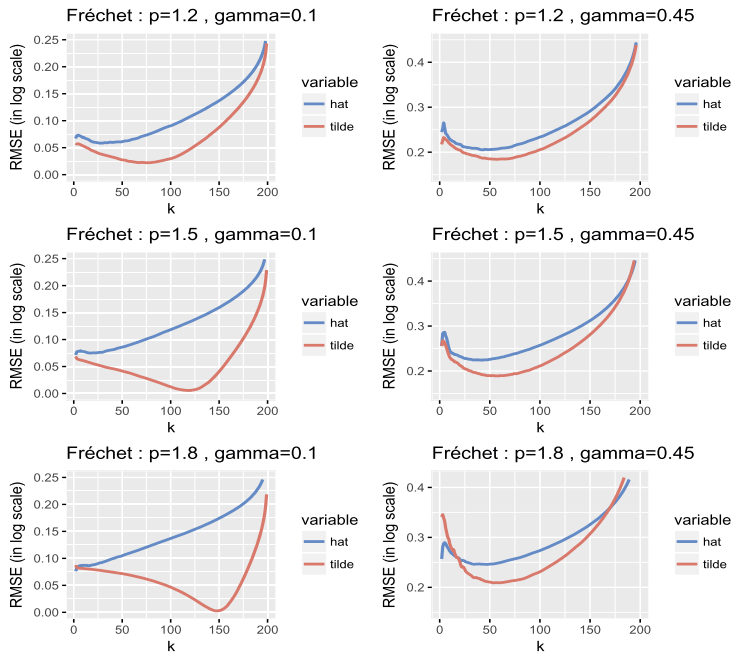
Figure 8. Fréchet distribution–RMSE (in log scale) of $\widehat{q}_{\tau_n'}^W(p)$ (blue) and $\widetilde{q}_{\tau_n'}^W(p)$ (red) as function of $k \in \{2, \ldots, n-1\}$. From left to right, $\gamma = 0.1, 0.45$. From top to bottom, $p = 1.2, 1.5, 1.8$.

In the case of the Student distribution, it may be seen from Figures 6 and 10 that $\widetilde{q}_{\tau_n'}^W(p)$ remains still competitive, but $\widehat{q}_{\tau_n'}^W(p)$ becomes more reliable for large values of p, say, $p \geq 1.9$. In particular, $\widehat{q}_{\tau_n'}^W(p)$ is clearly the winner for $p = 2$ as already demonstrated in Daouia *et al.* [18] via other scenarios. We repeated this kind of exercise with different values of γ, and arrived at the same tentative conclusions.

Interestingly, for p close to 1 and for the positive distributions, both estimators seem to perform comparably. The important gap in performance which sometimes occurs as p increases is most certainly due to the sensitivity of the least asymmetrically weighted estimator to the top extreme values in the sample. The estimator $\widetilde{q}_{\tau_n'}^W(p)$ does of course benefit from more robustness since it is computed using a single sample quantile.

6.3. Selection of the sample fraction k

The computation of the different presented extreme-value estimators requires the determination of the optimal value of the sample fraction k involved in the intermediate level $\tau_n = 1 - k/n$. A commonly used heuristic approach is to plot each estimator versus k and then pick out a suitable k corresponding to the first stable part of the plot (see, *e.g.*, Section 3 in de Haan and Ferreira [23]). A vexing defect with this heuristic approach from a forecasting perspective is that it re-
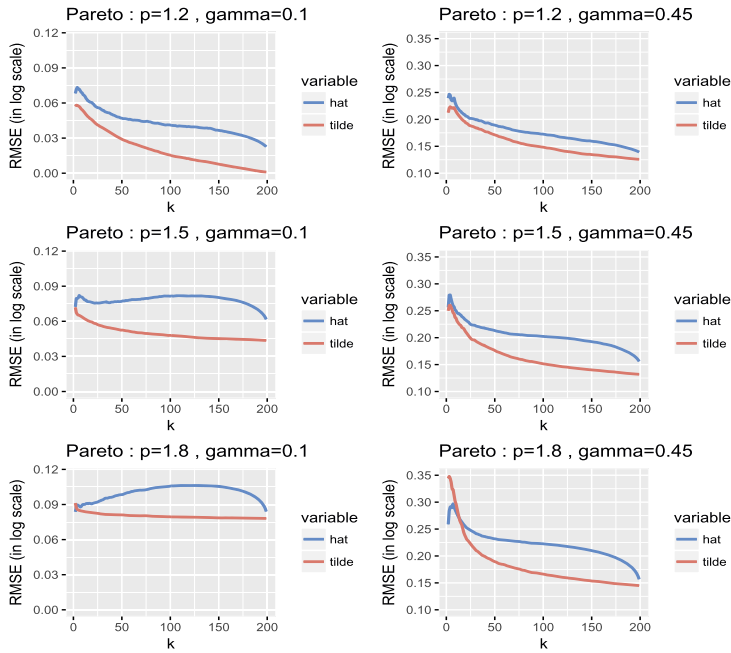
Figure 9. Pareto distribution–RMSE (in log scale) of $\widehat{q}_{\tau_n'}^{W}(p)$ (blue) and $\widetilde{q}_{\tau_n'}^{W}(p)$ (red) as function of $k \in \{2, \ldots, n-1\}$. From left to right, $\gamma = 0.1, 0.45$. From top to bottom, $p = 1.2, 1.5, 1.8$.

quires looking at the plot of the estimator at each forecast case. Instead of such a semi-automatic procedure, a fully automatic data-driven device can be performed to recover a suitable $\hat{k}$ in each forecast case. The basic idea is to evaluate first the estimator of interest [*e.g.*, $\widehat{\gamma}_n$, $\widehat{q}_{\tau_n'}^{W}(p)$, $\widetilde{q}_{\tau_n'}^{W}(p)$, $\widehat{\tau}_n'(p, \alpha_n; 1)$ or $\widecheck{q}_{\alpha_n}^{p}(2)$] over the range of values of k, and then to select the k where the variation of the results is the smallest. We achieve this by computing the standard deviations of the estimator over a "window" of successive values of k. The value of k where the standard deviation is minimal defines the desired sample fraction $\hat{k}$. This idea was already implemented recently by Daouia *et al.* [16,17], Stupfler [67], Goegebeur *et al.* [39] and Gardes and Stupfler [36], among others. Here, we apply the improved algorithm developed by El Methni and Stupfler [32], pages 919–920. The calculations all employ the same window of approximately 10 successive values of k.

The main difficulty when employing this automatic selection method is that the estimator of interest may be so unstable as a function of k that reasonable values of k (which would correspond to the true quantity we want to estimate) may be hidden in the plot. Consequently, the final estimates obtained from the selected $\hat{k}$ may exhibit considerable volatility as a function of the power p. Typical realizations are shown in Figure 11 when computing the optimal estimates $\widehat{q}_{\tau_n'}^{W}(p)$ and $\widetilde{q}_{\tau_n'}^{W}(p)$ of the extreme L^p-quantile $q_{\tau_n'}(p)$, with $\tau_n' = 1 - 1/n$ and $p \in (1, 2]$. Based on simulated samples from Fréchet, Pareto and Student distributions with $\gamma \in \{0.10, 0.45\}$, the resulting graphs of $p \mapsto \widehat{q}_{\tau_n'}^{W}(p)$ and $p \mapsto \widetilde{q}_{\tau_n'}^{W}(p)$ are plotted in red and blue, respectively, along
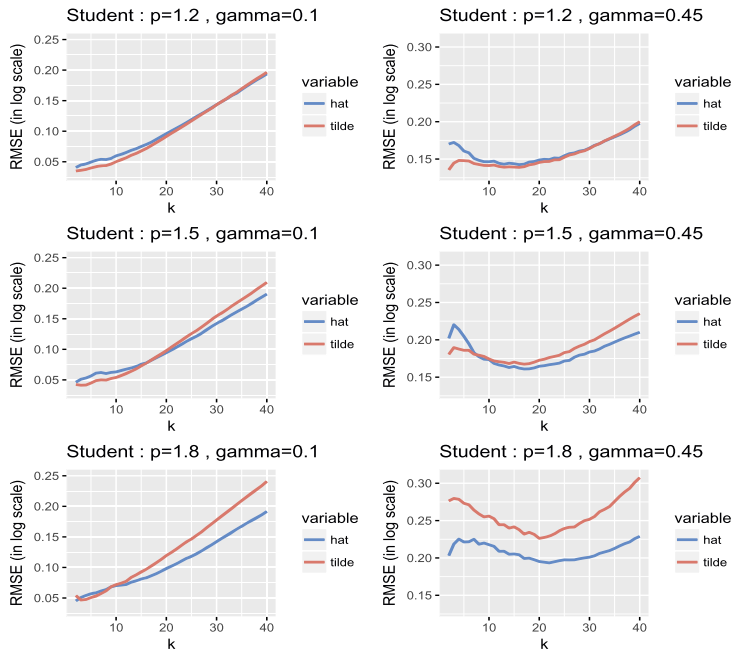
Figure 10. Student distribution–RMSE (in log scale) of $\widehat{q}_{\tau_n'}^W(p)$ (blue) and $\widetilde{q}_{\tau_n'}^W(p)$ (red) as function of k. From left to right, $\gamma = 0.1, 0.45$. From top to bottom, $p = 1.2, 1.5, 1.8$.

with the true L^p-quantile function $p \mapsto q_{\tau_n'}(p)$ in green. It may be seen that the selection data-driven method affords reasonable estimates regarding the very small sample size $n = 200$, but very good results with stable plots may require a large sample size of the order of several thousands.

To evaluate the performance of the automatic data-driven method, we have undertaken some Monte Carlo experiments using the same sample size $n = 200$. As a benchmark method for selecting the optimal k, we have used the value of k which minimizes the relative MSE of each estimator. The final RMSE and bias estimates of the two estimators $\widehat{q}_{\tau_n'}^W(p)$ and $\widetilde{q}_{\tau_n'}^W(p)$, computed over 3000 replications, are graphed in Figures 12 and 13 as functions of the power p. The solid curves in red and blue indicate the respective Monte Carlo estimates for $\widehat{q}_{\tau_n'}^W(p)$ and $\widetilde{q}_{\tau_n'}^W(p)$ obtained via the data-driven method, while the dashed versions give the benchmark optimal results obtained via the RMSE minimization. These results give a good overall impression of the precision of the two estimators $\widehat{q}_{\tau_n'}^W(p)$ and $\widetilde{q}_{\tau_n'}^W(p)$ as well as the adopted data-driven method. In particular, it may be seen that the evolution of the Monte Carlo estimates obtained via the data-driven method (solid lines) is generally coherent with the evolution of those obtained via the RMSE minimization (dashed lines).
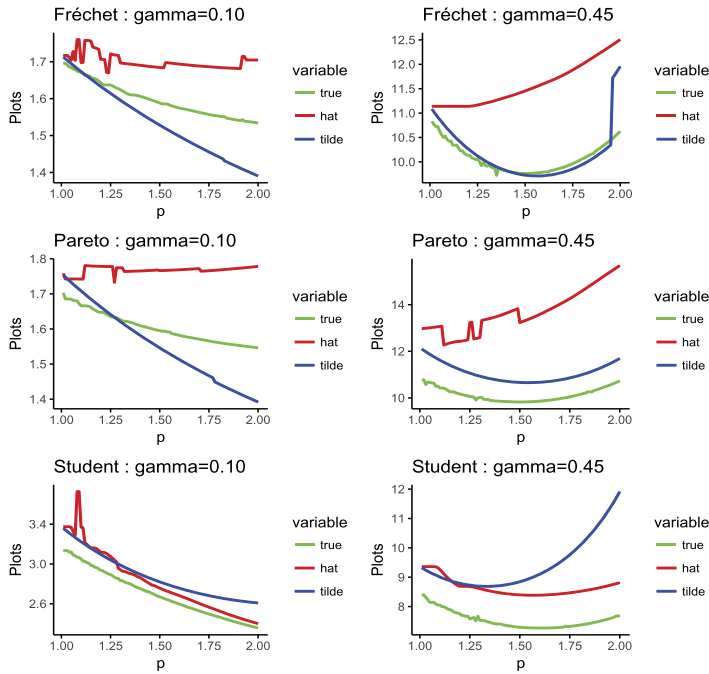
Figure 11. The plots of the estimators $\widehat{q}_{\tau_n'}^W(p)$ and $\widetilde{q}_{\tau_n'}^W(p)$ against p, respectively, in red and blue, with $\tau_n' = 1 - 1/n$ and k selected by the data-driven method. The true L^p-quantile function $p \mapsto q_{\tau_n'}(p)$ in green.

7. Validation and comparison on historical data

An important step beyond estimation of extreme L^p-quantiles $q_{\tau_n'}(p)$ from historical data is to be able to validate and compare the presented estimation procedures. We already know that, in the case of non-negative loss distributions, it is more efficient to use the plug-in Weissman estimator $\widetilde{q}_{\tau_n'}^W(p)$ than the least asymmetrically weighted L^p estimator $\widehat{q}_{\tau_n'}^W(p)$, as indicated by the Monte Carlo evidence above. In contrast, there is no clear winner in terms of the MSE in the case of real-valued profit-loss random variables. Here, we focus on the latter case when the ultimate interest is in an estimate of the loss return amount (negative log-return) that will be fallen below (on average) only once in N cases, with N being typically larger than or equal to the sample size n [see, *e.g.*, Acharya *et al.* [2], Chavez-Demoulin *et al.* [14], Gong *et al.* [40] and Cai *et al.* [12] for similar recent studies]. More specifically, we wish to use $\widetilde{q}_{\tau_n'}^W(p)$ and $\widehat{q}_{\tau_n'}^W(p)$ as estimators of the $(1/n)$th L^1-quantile $q_{1/n}(1) \equiv q_{\tau_n'}(p)$, for which verification and comparison is possible thanks to its elicitability property (see, *e.g.*, Gneiting [38] and Ziegel [72]). Following the ideas of Gneiting [38] and Ziegel [72], we consider in this section the evaluation and comparison of the two competing estimators $\widetilde{q}_{\tau_n'}^W(p)$ and $\widehat{q}_{\tau_n'}^W(p)$ with the standard left tail Weissman quantile estimator $\widehat{q}_{1/n}^W(1)$ from a forecasting perspective, trying to give the best possible point estimate
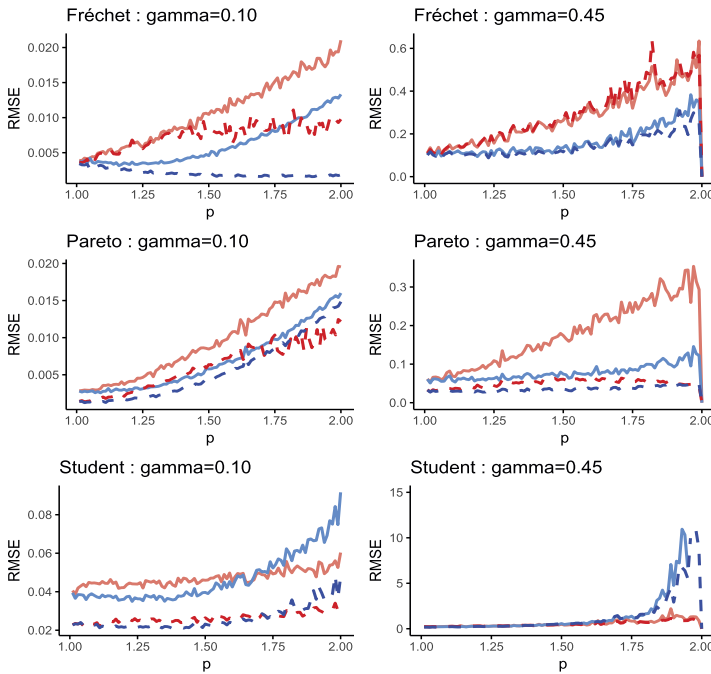
Figure 12. Relative MSE estimates of $\widehat{q}^{W}_{\tau'_n}(p)$ in red and $\widetilde{q}^{W}_{\tau'_n}(p)$ in blue, as functions of p, with $\tau'_n = 1 - 1/n$. In solid lines the estimates obtained via the data-driven method, in dashed lines the estimates obtained via the RMSE minimization.

for tomorrow with our knowledge of today. The portfolio under consideration is represented by the S&P500 Index from 4 January 1994 to 30 September 2016, which corresponds to 5727 trading days. The corresponding logarithmic returns are reported in Figure 14.

Let the random variable X model the future observation of interest. If the (τ'_n)th L^p-quantile $q_{\tau'_n}(p)$ coincides with the $(1/n)$th L^1-quantile $q_{1/n}(1)$, then it equals the optimal point forecast for X given by the Bayes rule

$$q_{\tau'_n}(p) = q_{1/n}(1) = \arg\min_{q \in \mathbb{R}} \mathbb{E}\big[L_n(q, X)\big],$$

under the asymmetric piecewise linear scoring function

$$L_n : \mathbb{R}^2 \longrightarrow [0, \infty), \qquad (q, x) \mapsto \eta_{\frac{1}{n}}(x - q; 1),$$

where $L_n(q, x)$ represents the loss or penalty when the point forecast q is issued and the realization x of X materializes. Following Gneiting [38] and Ziegel [72], the point estimates $\widehat{q}^{W}_{\tau'_n}(p)$, $\widetilde{q}^{W}_{\tau'_n}(p)$ and $\widehat{q}^{W}_{1/n}(1)$ of $q_{\tau'_n}(p)$ can then be compared and assessed by means of the scoring function L_n. Suppose that, in T forecast cases, we have point forecasts $(q^{(m)}_1, \ldots, q^{(m)}_T)$ and realizing
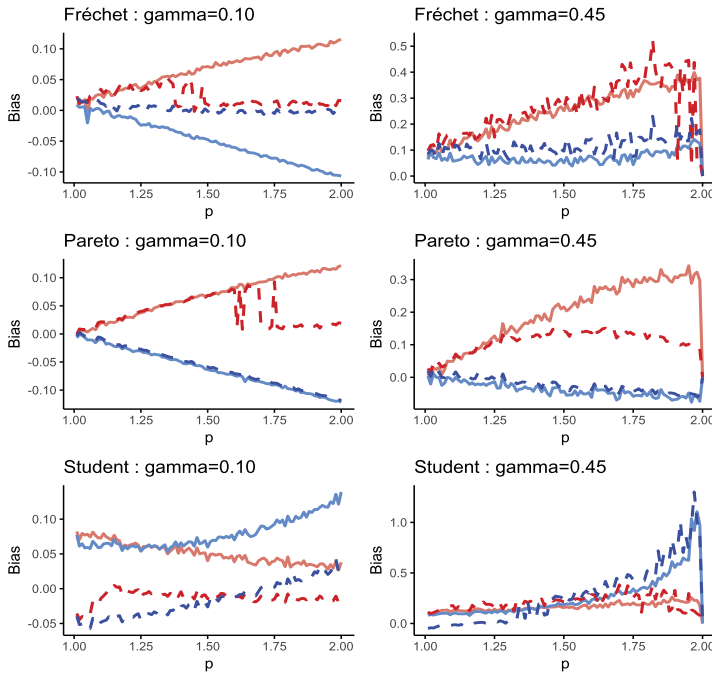
Figure 13. Bias estimates of $\widehat{q}^W_{\tau'_n}(p)$ in red and $\widetilde{q}^W_{\tau'_n}(p)$ in blue, as functions of p, with $\tau'_n = 1 - 1/n$. In solid lines the estimates obtained via the data-driven method, in dashed lines the estimates obtained via the RMSE minimization.

observations $(x_1, \dots, x_T)$, where the index m numbers the competing forecasters

$$q_t^{(1)} := \widehat{q}^W_{1/n}(1), \qquad q_t^{(2)} := \widehat{q}^W_{\tau'_n}(p) \quad \text{and} \quad q_t^{(3)} := \widetilde{q}^W_{\tau'_n}(p)$$

that are computed at each forecast case $t = 1, \dots, T$. These purely historical estimates can then be ranked in terms of their average scores (the lower the better):

$$\bar{L}_n^{(m)} = \frac{1}{T} \sum_{t=1}^{T} L_n\big(q_t^{(m)}, x_t\big), \qquad m = 1, 2, 3. \tag{18}$$

In our motivating application concerned with the logarithmic returns of the S&P500 Index, the three estimates were computed on rolling windows of length $n = 2510$, which corresponds to $T = 3217$ forecast cases. Based on the US market, there are on average 251 trading days in a year, and hence each rolling window of size $n = 10 \times 251$ trading days corresponds to a period of 10 years. Therefore, the tail quantity of interest $q_{\tau'_n}(p) \equiv q_{1/n}(1)$ represents the daily loss return (negative log-return) for a once-per-decade market crisis. Such a choice of once-in-a-decade extreme event is often used to evaluate systemic financial risk such as in, for instance, Brownlees and Engle [11] and Cai *et al.* [12] and the references therein. We also used the same
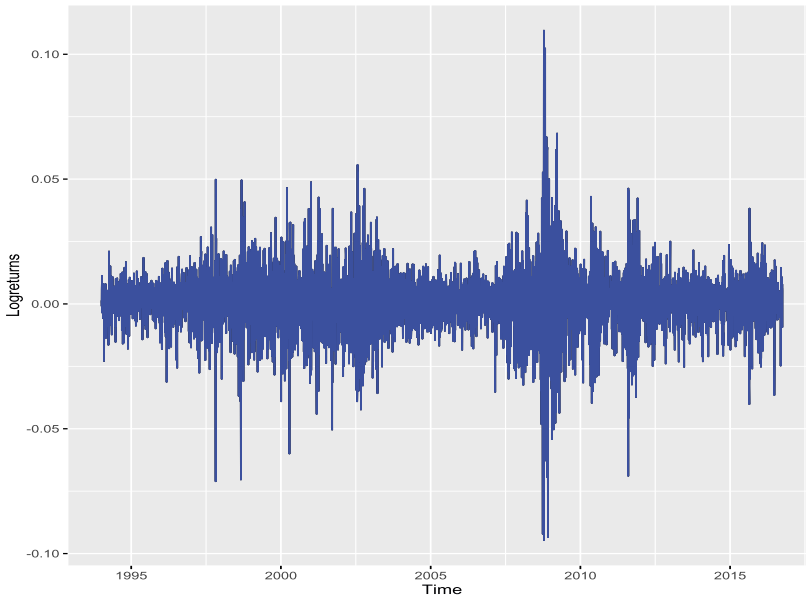
Figure 14. Log-returns of the S&P500 Index.

considerations as before for the choice of the intermediate level τ_n and the estimator $\hat{\gamma}_n$ of the tail index γ, with the sign convention for values of $Y = -X$ as the negative of returns. With this sign convention, the quantile of interest, $q_{1/n}(1)$, can be written as $-Q_{1-1/n}(1)$, where $Q_\tau(p)$ stands for the τth L^p-quantile of Y. The extreme level τ'_n such that $Q_{\tau'_n}(p) = Q_{1-1/n}(1)$ has the closed form expression $\tau'_n(p, \alpha_n; 1)$ described in (9), with $\alpha_n = 1 - 1/n$, and can be estimated by $\widehat{\tau}'_n(p, 1 - 1/n; 1)$ in (10). Alternatively, without recourse to this sign convention as the negative of returns, it is not hard to check that the level τ'_n such that $q_{\tau'_n}(p) = q_{1/n}(1)$ can directly be estimated by

$$\widehat{\tau}'_n(p, 1/n) := \frac{1}{n} \frac{1}{\widehat{\gamma}_n} B\left(p, \frac{1}{\widehat{\gamma}_n} - p + 1\right).$$

This might suggest the following strategy at each forecast case $t = 1, \ldots, T$:

(a) Calculate the first competing estimate $q_t^{(1)} = \widehat{q}_{1/n}^W(1)$;
(b) For a given value of $p \in (1, 2]$, calculate the τ'_n estimate $\hat{\tau}'_n(t) = \widehat{\tau}'_n(p, 1/n)$;
(c) Calculate the other competing estimates $q_t^{(2)} = \widehat{q}_{\tau'_n}^W(p)$ and $q_t^{(3)} = \widetilde{q}_{\tau'_n}^W(p)$ by substituting the estimated value $\hat{\tau}'_n(t)$ in place of τ'_n.

As a matter of fact, we use in step (c) the two-stage estimators $q_t^{(2)} = \widehat{q}_{\hat{\tau}'_n(t)}^W(p)$ and $q_t^{(3)} = \widetilde{q}_{\hat{\tau}'_n(t)}^W(p)$: first, we estimate τ'_n by $\hat{\tau}'_n(t)$ in step (b) and, second, we use the estimators $\widehat{q}_{\tau'_n}^W(p)$ and $\widetilde{q}_{\tau'_n}^W(p)$, as if τ'_n were known, by substituting $\hat{\tau}'_n(t)$ in place of τ'_n.

Of course, the computation of the different point estimates in steps (a), (b) and (c) requires the determination of the optimal values of the sample fraction k involved in the intermediate levels τ_n of these estimates. Here, we apply the data-driven method described in Section 6.3. For instance, the plot of the estimator $q_t^{(1)}(k) = \hat{q}_{1/n}^W(1)$, that is obtained at the first forecast case $t = 1$, can be visualized in Figure 15(a) as a rainbow curve. The effect of the Hill estimator $\hat{\gamma}_n(k)$ on $q_t^{(1)}(k)$ is highlighted by a colour-scheme, ranking from dark red (low $\hat{\gamma}_n$) to dark violet (high $\hat{\gamma}_n$). The resulting optimal estimate $q_t^{(1)}(\hat{k}_1)$ is indicated by the horizontal yellow dashed line, which affords a less pessimistic forecast than the worst observed loss return, $X_{1,n}$, indicated by the horizontal pink dashed line.

When proceeding to step (b) in the first forecast case, with the choice of $p = 2$, we obtain the plot of the estimator $\hat{\tau}_n'(t) = \hat{\tau}_n'(p, k)$ graphed in Figure 15(b) as a rainbow curve, along with its optimal value indicated by the horizontal yellow dashed line. We can see that the resulting optimal expectile level, $\hat{\tau}_n'(t) = 0.000137$, is much more extreme than the chosen quantile level (relative frequency $1/n = 0.000398$) indicated by the horizontal pink dashed line. Finally, by proceeding to step (c), we get the estimators $q_t^{(2)}(k) = \hat{q}_{\tau_n'}^W(p)$ and $q_t^{(3)}(k) = \tilde{q}_{\tau_n'}^W(p)$ displayed in Figure 15(c), along with their optimal values $q_t^{(2)}(\hat{k}_2)$ and $q_t^{(3)}(\hat{k}_3)$.

It may be seen in this first forecast case that both expectile estimators $\hat{q}_{\tau_n'}^W(2)$ and $\tilde{q}_{\tau_n'}^W(2)$ point towards similar forecasts as the rival quantile estimator $\hat{q}_{1/n}^W(1)$. This can be visualized more clearly in Figure 16, where the plots of the three estimators are superimposed. Interestingly, $\hat{q}_{1/n}^W(1)$ (in blue) remains very close to $\tilde{q}_{\tau_n'}^W(2)$ (in black) before being extrapolated beyond the minimum log-return $X_{1,n}$ (in pink). Then it becomes very close to the other expectile estimator $\hat{q}_{\tau_n'}^W(2)$ (in orange). In order to decide on the global accuracy of the three competing methods, we shall need to rank the values of their realized losses $\bar{L}_n^{(m)}$ by making use of the T forecasts and realizing observations, as described in (18).

The plots of the realized loss versus k are graphed in Figure 17(a) for $\hat{q}_{1/n}^W(1)$ and $\hat{q}_{\tau_n'}^W(p)$, with various values of $p \in \{1.1, 1.2, \ldots, 1.9, 2\}$, and in Figure 17(b) for the $\hat{q}_{1/n}^W(1)$ benchmark and $\tilde{q}_{\tau_n'}^W(p)$ with the same values of p. We can already see that $p = 2$ is a worse choice for both $\hat{q}_{\tau_n'}^W(p)$ and $\tilde{q}_{\tau_n'}^W(p)$ estimators.

The optimal values of the realized loss for the three methods (the lower the better), displayed in Table 1, indicate that the popular Weissman quantile estimator $\hat{q}_{1/n}^W(1)$ does not ensure the best accurate forecasts of the classical risk measure $q_{1/n}(1)$.

The top forecaster is $\hat{q}_{\tau_n'}^W(p)$ for $p = 1.3, 1.6, 1.5$ in this order, followed by $\tilde{q}_{\tau_n'}^W(p)$ for $p = 1.7, 1.6, 1.5, 1.4, 1.3, 1.2, 1.1$, and then $\hat{q}_{1/n}^W(1)$. Their optimal values obtained in the first and last forecast cases are shown in Table 2. In the forecast case $t = 1$, based on the loss returns observed during the first decade from 1994-01-05 to 2003-12-19, all forecasts of the Value at Risk $q_{1/n}(1)$ do not succeed in falling below the worst recorded loss return $X_{1,n}$. Yet, all of the generalized L^p-quantiles $\hat{q}_{\tau_n'}^W(p)$ and $\tilde{q}_{\tau_n'}^W(p)$ appear to be smaller and hence more conservative than the usual L^1-quantile $\hat{q}_{1/n}^W(1)$. Here, the tail index estimate is found to be $\hat{\gamma}_n(\hat{k}) = 0.256$. In the forecast case $t = T$, based on the last decade from 2006-10-11 to 2016-09-29, all forecasts of the Value at Risk were capable of extrapolating outside the minimal loss return $X_{1,n}$. This is due to the turbulent episodes that have been experienced by financial markets during 2007–2008,
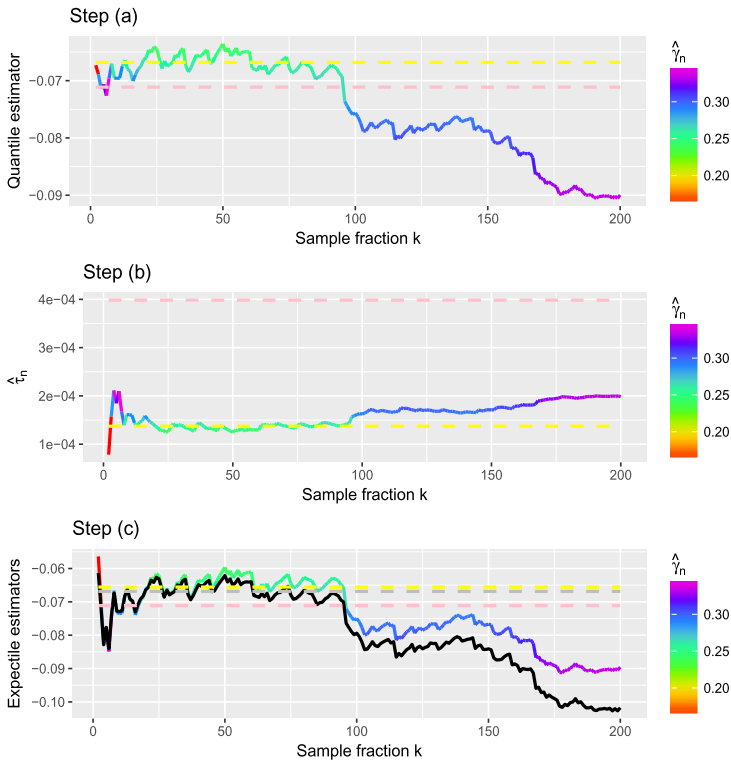
Figure 15. Step (a) The plot of the estimator $q_t^{(1)}(k) = \widehat{q}_{1/n}^W(1)$ versus k as a rainbow curve. The selected optimal estimate $q_t^{(1)}(\hat{k}_1)$ indicated by the horizontal yellow dashed line. The sample minimum log-return $X_{1,n}$ indicated by the horizontal pink dashed line. Step (b) The plot of the estimator $\hat{\tau}_n'(p, k)$, with $p = 2$, versus k as a rainbow curve. The selected optimal estimate $\hat{\tau}_n'(p, k)$ indicated by the horizontal yellow dashed line. The relative frequency $1/n$ indicated by the horizontal pink dashed line. Step (c) The plots of the estimators $q_t^{(2)}(k) = \widehat{q}_{\tau_n'}^W(p)$ and $q_t^{(3)}(k) = \widetilde{q}_{\tau_n'}^W(p)$ versus k, respectively, as rainbow and black curves, with $p = 2$. The selected optimal estimates $q_t^{(2)}(\hat{k}_2)$ and $q_t^{(3)}(\hat{k}_3)$ indicated, respectively, by the horizontal yellow and grey dashed lines. The minimum log-return $X_{1,n}$ indicated by the horizontal pink dashed line. The effect of $\hat{\gamma}_n(k)$ on $q_t^{(2)}(k)$ is highlighted by the colour-scheme.

as visualized in Figure 14. In particular, the tail index estimate becomes $\widehat{\gamma}_n(\hat{k}) = 0.359$. Yet, the top forecasters $\widehat{q}_{\tau_n'}^W(p)$ and $\widetilde{q}_{\tau_n'}^W(p)$ appear to be larger and hence less pessimistic than the L^1-quantile $\widehat{q}_{1/n}^W(1)$. In both forecast cases, that eternal maxim of the pessimists, "expect the worst, and you won't be disappointed" seems to be transformed into a more realistic calculus via tail L^p-quantiles than classical quantiles.

Finally, we would like to comment on the evolution of the extreme L^p-quantile level $\hat{\tau}_n'(t)$ with t. The optimal estimates $t \mapsto \hat{\tau}_n'(t)$, obtained for the different values of p, are graphed in Figure 18. It can be seen that $\hat{\tau}_n'(t)$ decreases, uniformly in t, as p increases. Also, it may be seen
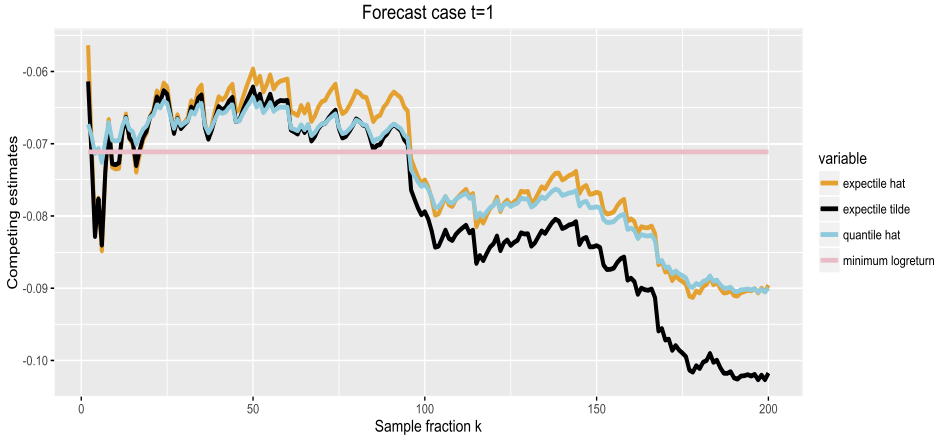
Figure 16. The plots of the three estimators $q_t^{(1)}(k) = \widehat{q}_{1/n}^{W}(1)$ in blue, $q_t^{(2)}(k) = \widehat{q}_{\tau_n'}^{W}(2)$ in orange, and $q_t^{(3)}(k) = \widetilde{q}_{\tau_n'}^{W}(2)$ in black. The sample minimum $X_{1,n}$ in pink.

Table 1. Optimal values $\bar{L}_n^{(1)}$, $\bar{L}_n^{(2)}$ and $\bar{L}_n^{(3)}$ of the realized loss for the three forecasters $\widehat{q}_{1/n}^{W}(1)$, $\widehat{q}_{\tau_n'}^{W}(p)$ and $\widetilde{q}_{\tau_n'}^{W}(p)$, respectively. Results based on daily loss returns

	$\bar{L}_n^{(1)} = 5.758\ 10^{-05}$				
	$p = 1.1$	$p = 1.2$	$p = 1.3$	$p = 1.4$	$p = 1.5$
$\bar{L}_n^{(2)}$	$5.856\ 10^{-05}$	$5.842\ 10^{-05}$	$5.632\ 10^{-05}$	$5.861\ 10^{-05}$	$5.727\ 10^{-05}$
$\bar{L}_n^{(3)}$	$5.755\ 10^{-05}$	$5.752\ 10^{-05}$	$5.748\ 10^{-05}$	$5.745\ 10^{-05}$	$5.742\ 10^{-05}$
	$p = 1.6$	$p = 1.7$	$p = 1.8$	$p = 1.9$	$p = 2$
$\bar{L}_n^{(2)}$	$5.712\ 10^{-05}$	$5.847\ 10^{-05}$	$5.918\ 10^{-05}$	$6.024\ 10^{-05}$	$6.118\ 10^{-05}$
$\bar{L}_n^{(3)}$	$5.739\ 10^{-05}$	$5.735\ 10^{-05}$	$5.919\ 10^{-05}$	$6.030\ 10^{-05}$	$6.025\ 10^{-05}$

Table 2. Optimal values of the top forecasters, obtained in the first and last forecast cases, along with the sample minimum $X_{1,n}$. Results based on daily loss returns

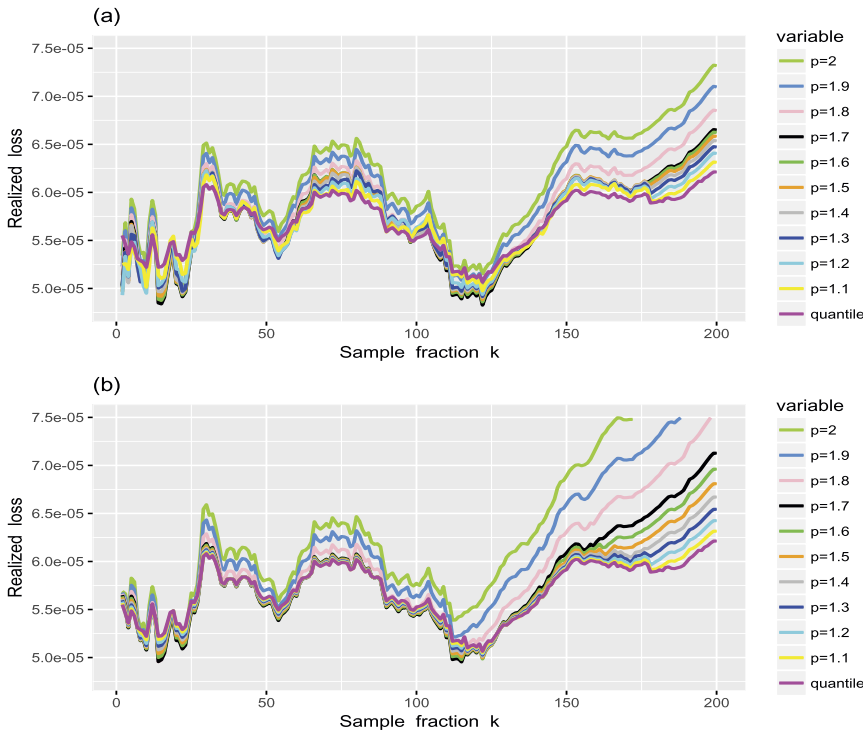| Forecaster	$\widehat{q}_{\tau_n'}^{W}(1.3)$	$\widehat{q}_{\tau_n'}^{W}(1.6)$	$\widehat{q}_{\tau_n'}^{W}(1.5)$	$\widetilde{q}_{\tau_n'}^{W}(1.7)$	$\widetilde{q}_{\tau_n'}^{W}(1.6)$	$\widetilde{q}_{\tau_n'}^{W}(1.5)$
$t = 1$	-0.067901	-0.067169	-0.067288	-0.066846	-0.066841	-0.066837
$t = T$	-0.105717	-0.103084	-0.104613	-0.101254	-0.102211	-0.103131
Forecaster	$\widetilde{q}_{\tau_n'}^{W}(1.4)$	$\widetilde{q}_{\tau_n'}^{W}(1.3)$	$\widetilde{q}_{\tau_n'}^{W}(1.2)$	$\widetilde{q}_{\tau_n'}^{W}(1.1)$	$\widehat{q}_{1/n}^{W}(1)$	$X_{1,n}$
$t = 1$	-0.066832	-0.066828	-0.066824	-0.066820	-0.066816	-0.071127
$t = T$	-0.104016	-0.104870	-0.105695	-0.106493	-0.107266	-0.094695

Figure 17. (a) Plots of the realized loss $k \mapsto \bar{L}_n^{(m)}(k)$ for $\widehat{q}_{1/n}^W(1)$ in magenta and $\widehat{q}_{\tau_n'}^W(p)$ with different values of p. (b) Results with $\widetilde{q}_{\tau_n'}^W(p)$ in place of $\widehat{q}_{\tau_n'}^W(p)$. Results based on daily loss returns.

that the curve corresponding to the best choice $p = 1.3$ (dark blue) exhibits two different trends before and after the severe losses of 2007–2008. Both trends appear to be much more extreme than the quantile level $1/n$.

Let us now consider lower frequency data to reduce the potential serial dependence in this application. The theory for the extreme L^p-quantile estimators is derived for dependent random variables $X_1, \ldots, X_n$ under mixing conditions. Our theorems also work under independence with reduced asymptotic variances. Here, similarly to Cai *et al.* [12], we reduce substantially the potential serial dependence by choosing weekly (Wednesday to Wednesday) returns in the same sample period. This results in a sample of size 1176. We compute the three estimates $\widehat{q}_{1/n}^W(1)$, $\widehat{q}_{\tau_n'}^W(p)$ and $\widetilde{q}_{\tau_n'}^W(p)$ of $q_{1/n}(1)$ on rolling windows of length $n = 520$, which corresponds to $T = 656$ forecast cases. Given that there are 52 weeks in a year, $q_{1/n}(1)$ can be viewed as the weekly loss return for a once-per-decade financial crisis. The plots of the realized loss $k \mapsto \bar{L}_n^{(m)}(k)$ are graphed in Figure 19(a) for $\widehat{q}_{1/n}^W(1)$ and $\widehat{q}_{\tau_n'}^W(p)$, and in Figure 19(b) with $\widetilde{q}_{\tau_n'}^W(p)$ in place of $\widehat{q}_{\tau_n'}^W(p)$.
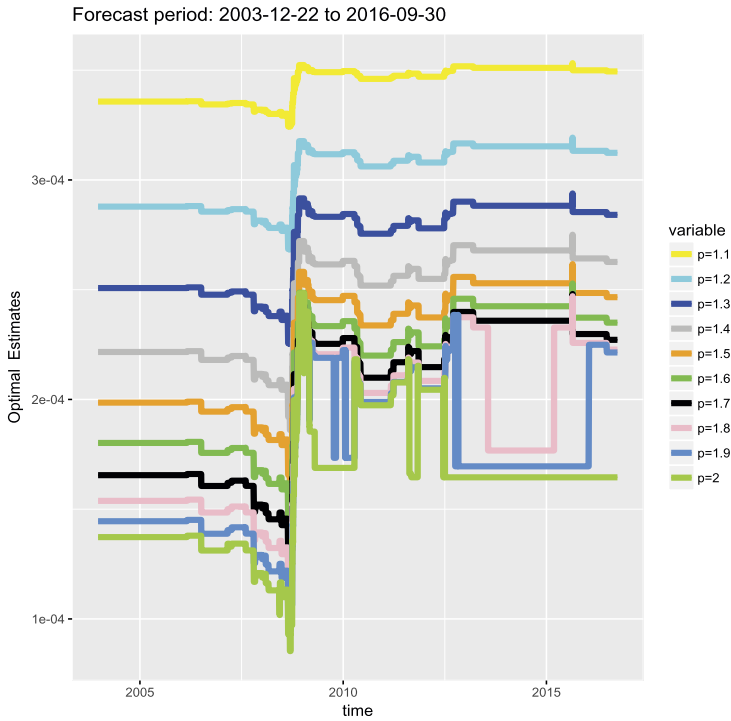
Forecast period: 2003-12-22 to 2016-09-30

Figure 18. The final estimates $t \mapsto \hat{\tau}'_n(t)$, obtained for $p \in \{1.1, 1.2, \ldots, 1.9, 2\}$.

The optimal values of the realized loss for the three methods, displayed in Table 3, indicate that the best forecaster is $\widetilde{q}^W_{\tau'_n}(p)$ for $p = 2, 1.9, 1.8$ in this order, followed by $\widehat{q}^W_{\tau'_n}(p)$ for $p = 1.2$, $\widetilde{q}^W_{\tau'_n}(p)$ for $p = 1.7, 1.4, 1.5, 1.6$, and then $\widehat{q}^W_{1/n}(1)$. All in all, the final results based on weekly

Table 3. Optimal values $\bar{L}^{(1)}_n$, $\bar{L}^{(2)}_n$ and $\bar{L}^{(3)}_n$ of the realized loss for the three forecasters $\widehat{q}^W_{1/n}(1)$, $\widehat{q}^W_{\tau'_n}(p)$ and $\widetilde{q}^W_{\tau'_n}(p)$, respectively. Results based on weekly loss returns

	$\bar{L}^{(1)}_n = 0.0001791$				
	$p = 1.1$	$p = 1.2$	$p = 1.3$	$p = 1.4$	$p = 1.5$
$\bar{L}^{(2)}_n$	0.0001834	0.0001771	0.0001845	0.0001911	0.0001821
$\bar{L}^{(3)}_n$	0.0001792	0.0001796	0.0001801	0.0001790	0.0001790
	$p = 1.6$	$p = 1.7$	$p = 1.8$	$p = 1.9$	$p = 2$
$\bar{L}^{(2)}_n$	0.0001829	0.0001830	0.0001814	0.0001811	0.0001815
$\bar{L}^{(3)}_n$	0.0001790	0.0001788	0.0001768	0.0001763	0.0001762

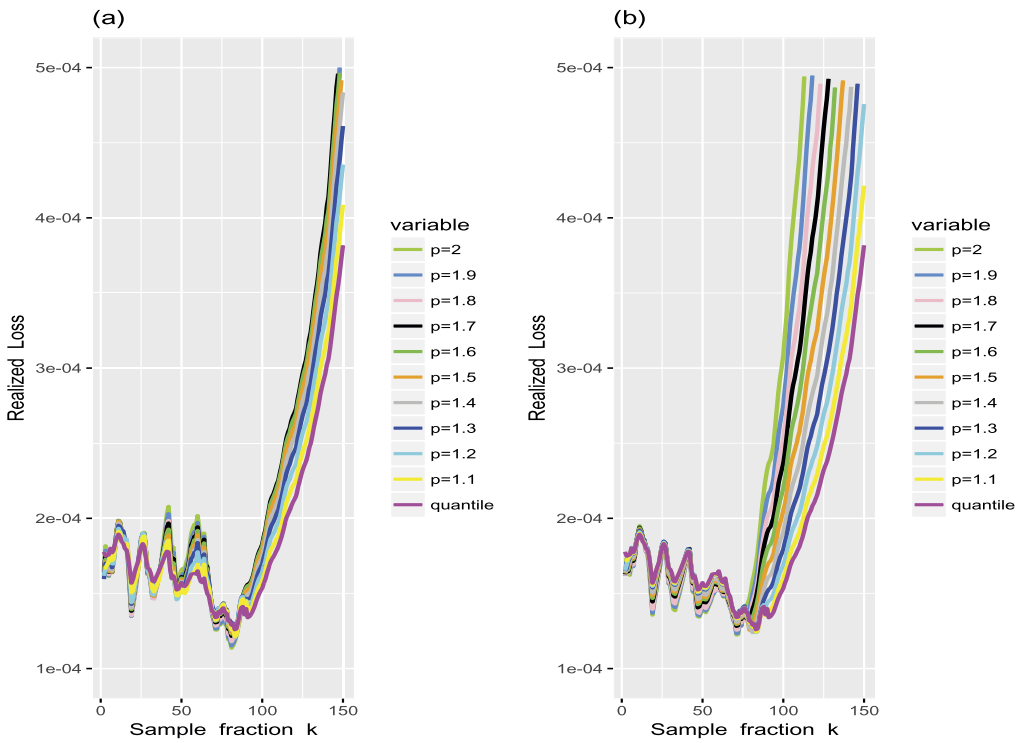**Figure 19.** (a) Plots of the realized loss $k \mapsto \bar{L}_n^{(m)}(k)$ for $\widehat{q}_{1/n}^W(1)$ in magenta and $\widehat{q}_{\tau_n'}^W(p)$ with different values of p. (b) Results with $\widetilde{q}_{\tau_n'}^W(p)$ in place of $\widehat{q}_{\tau_n'}^W(p)$. Results based on weekly loss returns.

loss returns seem to indicate that $\widetilde{q}_{\tau_n'}^W(p)$ is the winner, while the results based on daily loss returns tend to favor the use of $\widehat{q}_{\tau_n'}^W(p)$.

Acknowledgements

Supplementary Material

Supplement to "Extreme M-quantiles as risk measures: From L^1 to L^p optimization."
(DOI: [10.3150/17-BEJ987SUPP](); .pdf). The supplement to this article contains additional simu-

lations, a second application to medical insurance data, technical lemmas and the proofs of all theoretical results of the main article.

References

[1] Abdous, B. and Rémillard, B. (1995). Relating quantiles and expectiles under weighted-symmetry. *Ann. Inst. Statist. Math*. **47** 371–384. MR1345429

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

[3] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Math. Finance* **9** 203–228. MR1850791

[4] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes. Theory and Applications. Wiley Series in Probability and Statistics*. Chichester: Wiley. With contributions from Daniel De Waal and Chris Ferro. MR2108013

[5] Bellini, F. and Di Bernardino, E. (2017). Risk management with expectiles. *Eur. J. Finance* **23** 487–506. DOI:10.1080/1351847X.2015.1052150.

[6] Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance Math. Econom*. **54** 41–48. MR3145849

[7] Boente, G. and Fraiman, R. (1995). Asymptotic distribution of smoothers based on local means and local medians under dependence. *J. Multivariate Anal*. **57** 77–90.

[8] Bradley, R.C. (1988). A central limit theorem for stationary ρ-mixing sequences with infinite variance. *Ann. Probab*. **16** 313–332.

[9] Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv*. **2** 107–144. Update of, and a supplement to, the 1986 original. MR2178042

[10] Breckling, J. and Chambers, R. (1988). *M*-quantiles. *Biometrika* **75** 761–771. MR0995118

[11] Brownlees, C.T. and Engle, R. (2012). Volatility, correlation and tails for systemic risk measurement. Preprint.

[12] Cai, J.-J., Einmahl, J.H.J., de Haan, L. and Zhou, C. (2015). Estimation of the marginal expected shortfall: The mean when a related variable is extreme. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **77** 417–442. MR3310533

[13] Cai, Z. (2003). Nonparametric estimation equations for time series data. *Statist. Probab. Lett*. **62** 379–390. MR1973313

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

[15] Chen, Z. (1996). Conditional L_p-quantiles and their application to the testing of symmetry in nonparametric regression. *Statist. Probab. Lett*. **29** 107–115. MR1411408

[16] Daouia, A., Florens, J.-P. and Simar, L. (2010). Frontier estimation and extreme value theory. *Bernoulli* **16** 1039–1063. MR2759168

[17] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli* **19** 2557–2589. MR3160564

[18] Daouia, A., Girard, S. and Stupfler, G. (2017). Estimation of tail risk based on extreme expectiles. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. To appear, DOI:10.1111/rssb.12254.

[19] Daouia, A., Girard, S. and Stupfler, G. (2017). Supplement to "Extreme M-quantiles as risk measures: From L^1 to L^p optimization." DOI:10.3150/17-BEJ987SUPP.

[20] Davis, R. and Mikosch, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli* **15** 977–1009.

[21] Davis, R., Mikosch, T. and Cribben, I. (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *J. Econometrics* **170** 142–152.

[22] Davis, R.A., Mikosch, T. and Zhao, Y. (2013). Measures of serial extremal dependence and their estimation. *Stochastic Process. Appl.* **123** 2575–2602. MR3054537

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

[24] de Haan, L., Mercadier, C. and Zhou, C. (2016). Adapting extreme value statistics to financial time series: Dealing with bias and serial dependence. *Finance Stoch.* **20** 321–354. MR3479324

[25] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** 1833–1855. MR1026315

[26] Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.* **10** 1274–1301. MR1810875

[27] Drees, H. (2002). Tail empirical processes under mixing conditions. In *Empirical Processes Techniques for Dependent Data* (H.G. Dehling, T. Mikosch and M. Sørensen, eds.). Boston: Birkhäuser.

[28] Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9** 617–657. MR1996273

[29] Drees, H. and Rootzén, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186.

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

[31] El Methni, J., Gardes, L. and Girard, S. (2014). Nonparametric estimation of extreme risks from conditional heavy-tailed distributions. *Scand. J. Stat.* **41** 988–1012.

[32] El Methni, J. and Stupfler, G. (2017). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. *Statist. Sinica* **27** 907–930. MR3675036

[33] El Methni, J. and Stupfler, G. (2017). Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions. *Econom. Stat.* To appear. DOI:10.1016/j.ecosta.2017.03.002.

[34] Embrechts, P. and Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Ann. Rev. Stat. Appl.* **1** 493–514.

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

[36] Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes* **17** 45–75. MR3179970

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

[38] Gneiting, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* **106** 746–762. MR2847988

[39] Goegebeur, Y., Guillou, A. and Osmann, M. (2014). A local moment type estimator for the extreme value index in regression with random covariates. *Canad. J. Statist.* **42** 487–507.

[40] Gong, J., Li, Y., Peng, L. and Yao, Q. (2015). Estimation of extreme quantiles for functions of dependent random variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 1001–1024.

[41] Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications*. *Lect. Notes Stat. Proc.* **198** 127–145. Springer, Heidelberg. MR3051266

[42] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204

[43] Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Ann. Inst. Statist. Math.* **52** 459–470.

[44] Hsing, T. (1991). On tail index estimation using dependent data. *Ann. Statist.* **19** 1547–1569.

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

[46] Ibragimov, I.A. (1975). A remark on the central limit theorem for dependent random variables. *Theor. Probab. Appl.* **20** 135–141. MR0362448

[47] Jones, M.C. (1994). Expectiles and *M*-quantiles are quantiles. *Statist*. *Probab*. *Lett*. **20** 149–153. MR1293293

[48] Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. Technical report, Univ. Toronto.

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

[50] Kuan, C-M., Yeh, J-H. and Hsu, Y-C. (2009). Assessing value at risk with CARE, the conditional autoregressive expectile models. *J. Econometrics* **2** 261–270.

[51] Mao, T., Ng, K.W. and Hu, T. (2015). Asymptotic expansions of generalized quantiles and expectiles for extreme risks. *Probab*. *Engrg*. *Inform*. *Sci*. **29** 309–327. MR3355606

[52] Mao, T. and Yang, F. (2015). Risk concentration based on expectiles for extreme risks under FGM copula. *Insurance Math*. *Econom*. **64** 429–439. MR3394666

[53] Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd ed. New York: Springer.

[54] Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55** 819–847. MR0906565

[55] Peligrad, M. (1987). On the central limit theorem for ρ-mixing sequences of random variables. *Ann*. *Probab*. **15** 1387–1394. MR0905338

[56] Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann*. *Statist*. **3** 119–131.

[57] Resnick, S. and Stărică, C. (1997). Asymptotic behavior of Hill's estimator for autoregressive data. *Commun*. *Stat*., *Stochastic Models* **13** 703–721. MR1482290

[58] Resnick, S. and Stărică, C. (1998). Tail index estimation for dependent data. *Ann*. *Appl*. *Probab*. **8** 1156–1183. MR1661160

[59] Resnick, S. and Stărică, C. (1995). Consistency of Hill's estimator for dependent data. *J*. *Appl*. *Probab*. **32** 139–167.

[60] Resnick, S.I. (2007). *Heavy-Tail Phenomena*. *Probabilistic and Statistical Modeling*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2271424

[61] Robert, C.Y. (2008). Estimating the multivariate extremal index function. *Bernoulli* **14** 1027–1064. MR2543585

[62] Robert, C.Y. (2009). Inference for the limiting cluster size distribution of extreme values. *Ann*. *Statist*. **37** 271–310.

[63] Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Process*. *Appl*. **119** 468–490.

[64] Schnabel, S. and Eilers, P. (2013). A location-scale model for non-crossing expectile curves. *Stat* **2** 171–183. DOI:10.1002/sta4.27.

[65] Schulze Waltrup, L., Sobotka, F., Kneib, T. and Kauermann, G. (2015). Expectile and quantile regression – David and Goliath? *Stat*. *Model*. **15** 433–456. MR3403125

[66] Sklar, M. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publ*. *Inst*. *Stat*. *Univ*. *Paris* **8** 229–231. MR0125600

[67] Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electron*. *J*. *Stat*. **7** 2298–2343. MR3108815

[68] Utev, S.A. (1990). On the central limit theorem for φ-mixing arrays of random variables. *Theory Probab*. *Appl*. **35** 131–139.

[69] Weissman, I. (1978). Estimation of parameters and large quantiles based on the *k* largest observations. *J*. *Amer*. *Statist*. *Assoc*. **73** 812–815. MR0521329

[70] Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach. *J*. *Nonparametr*. *Stat*. **6** 273–292. MR1383055

[71] Zhao, G.H., Teo, K.L. and Chan, K.S. (2005). Estimation of conditional quantiles by a new smoothing approximation of asymmetric loss functions. *Stat*. *Comput*. **15** 5–11.

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

[73] Zou, H. (2014). Generalizing Koenker's distribution. *J. Statist. Plann. Inference* **148** 123–127. MR3174153