Beyond tail median and conditional tail expectation: extreme risk estimation using tail Lp –optimisation

Laurent Gardes, Stephane Girard, Gilles Stupfler

Beyond tail median and conditional tail expectation: extreme risk estimation using tail L^p−optimisation

Laurent Gardes[(1)], Stéphane Girard[(2)] & Gilles Stupfler[(3)]

[(1)] Université de Strasbourg & CNRS, IRMA, UMR 7501, France

[(2)] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France

[(3)] School of Mathematical Sciences, University of Nottingham, UK

Abstract. The Conditional Tail Expectation is an indicator of tail behaviour that takes into account both the frequency and magnitude of a tail event. However, the asymptotic normality of its empirical estimator requires that the underlying distribution possess a finite variance; this can be a strong restriction in actuarial and financial applications. A valuable alternative is the Median Shortfall, although it only gives information about the frequency of a tail event. We construct a class of tail L^p−medians encompassing the Median Shortfall and Conditional Tail Expectation. For p in $(1, 2)$, a tail L^p−median depends on both the frequency and magnitude of tail events, and its empirical estimator is, within the range of the data, asymptotically normal under a condition weaker than a finite variance. We extrapolate this estimator and another technique to extreme levels using the heavy-tailed framework. The estimators are showcased on a simulation study and on real fire insurance data.

Keywords: asymptotic normality, conditional tail expectation, extreme value statistics, heavy-tailed distribution, L^p−optimisation, median shortfall, semiparametric extrapolation.

1 Introduction

A precise assessment of extreme risk is a crucial question in a number of fields of statistical applications. In actuarial science and finance, especially, a major question is to get a good understanding of potential extreme claims and losses that would constitute a threat to the survival of a company.

This has historically been done by simply using a quantile, also called Value-at-Risk (VaR) in the actuarial and financial world. The quantile of a real-valued random variable X at level $\alpha \in (0, 1)$ is, by definition, the lowest value $q(\alpha)$ which is exceeded by X with probability not larger than $1 - \alpha$. As a consequence, quantiles only provide an information about the frequency of a tail event; in particular, the quantile $q(\alpha)$ gives no indication about how heavy the right tail of X is or, perhaps more concretely, as to what a typical value of X above $q(\alpha)$ would be. From a risk assessment perspective, this is clearly an issue, as quantiles allow to quantify what a risky situation is (*i.e.* the quantile level), but not what the consequences of a risky situation would be (*i.e.* the behaviour of X beyond the quantile level).

The notion of Conditional Tail Expectation (CTE) precisely addresses this point. This risk measure, defined as $\text{CTE}(\alpha) := \mathbb{E}[X \mid X > q(\alpha)]$, is exactly the average value of X given that $X > q(\alpha)$. When X is continuous, $\text{CTE}(\alpha)$ coincides with the so-called Expected Shortfall $\text{ES}(\alpha)$, also known as Tail Value-at-Risk or Conditional Value-at-Risk and discussed in Acerbi & Tasche (2002), Rockafellar & Uryasev (2002), Tasche (2002), which is the average value of the quantile function $\tau \mapsto q(\tau)$ over the interval $[\alpha, 1)$. The potential of CTE for use in actuarial and financial risk management has been considered by a number of studies, such as Brazauskas *et al.* (2008), Tasche (2008), Wüthrich & Merz (2013), Emmer *et al.* (2015). Outside academic contexts, the CTE risk measure is used in capital requirement calculations by the Canadian financial and actuarial sectors (IMF, 2014), as well as for guaranteeing the sustainability of life insurance annuities in the USA (OECD, 2016). European regulators, *via* the Basel Committee on Banking Supervision, also recently recommended to use CTE rather than VaR in internal market risk models (BCBS, 2013). Whether the CTE or VaR should be used in a given financial application is still very much up for debate: we point to, among others, Artzner *et al.* (1999), Embrechts *et al.* (2009), Gneiting (2011), Daníelsson *et al.* (2013), Kou & Peng (2016) who discuss the various intrinsic axiomatic properties of CTE and/or VaR and their practical interpretations.

Aside from axiomatic considerations, the estimation of the CTE by the empirical conditional tail moment requires a finite (tail) second moment if this estimator is to be asymptotically normal. This should of course be expected, since a condition for the sample average to be an asymptotically normal estimator of the sample mean is precisely the finiteness of the variance. In heavy-tailed models, which are of interest in insurance/finance and that will be the focus of this paper, such moment restrictions can be essentially reformulated in terms of a condition on the tail index γ of

X. The variable X is heavy-tailed with index $\gamma > 0$ if its distribution behaves approximately like a power law distribution with exponent $1/\gamma$:

$$\mathbb{P}(X > x) = x^{-1/\gamma}\ell(x) \text{ for } x \text{ large enough},\qquad(1)$$

where ℓ is a slowly varying function at infinity, namely $\ell(tx)/\ell(t) \to 1$ as $t \to \infty$ for any $x > 0$ (Beirlant *et al.*, 2004, p.57). In such a model, the finite second moment condition is violated when the tail index γ, controlling the right tail heaviness of X, is such that $\gamma > 1/2$. At the finite-sample level, it is often desirable to have a finite fourth moment; this stronger condition is violated whenever $\gamma > 1/4$. Table 1 in the simulation study of El Methni & Stupfler (2017) shows a quite strong deterioration in the finite-sample performance of the empirical conditional tail moment as γ increases to $1/4$. A number of actuarial and financial data sets have been found to violate such integrability assumptions: we refer to, for instance, the Danish fire insurance data set considered in Example 4.2 of Resnick (2007), emerging market stock returns data (Hill, 2013, Ling, 2005) and exchange rates data (Hill, 2015). The deterioration in finite-sample performance as the number of finite moments decreases is due in no small part to increased variability in the right tail of X as γ increases, see Section 3.1 in El Methni & Stupfler (2018). To counteract this variability, one could move away from the CTE and use the Median Shortfall MS (Kou *et al.*, 2013, Kou & Peng, 2016), just as one can use the median instead of the mean for added robustness. However, $\text{MS}(\alpha)$ is nothing but $q([1 + \alpha]/2)$ and as such, similarly to VaR, does not contain any information on the behaviour of X beyond its value.

Our goal here is to design a class of tail indicators that realise a compromise between the sensitivity of the Conditional Tail Expectation and the robustness of the Median Shortfall. We will start by showing that these two quantities can be obtained in a new unified framework, which we call the class of *tail L^p-medians*. A tail L^p-median at level $\alpha \in (0, 1)$ is obtained, for $p \geq 1$, by minimising an L^p-moment criterion which only considers the probabilistic behaviour of X above a quantile $q(\alpha)$. As such, it satisfies a number of interesting properties we will investigate. Most importantly, the class of tail L^p-medians is linked to the notion of L^p-quantiles introduced by Chen (1996) and recently studied in heavy-tailed frameworks by Daouia *et al.* (2019), a tail L^p-median being the L^p-quantile of order $1/2$ (or L^p-median) of the variable $X \,|\, X > q(\alpha)$. In particular, the classical median of X is the L^1-median and the mean of X is the L^2-median (or expectile of order $1/2$, in the terminology of Newey & Powell, 1987), and we will show that similarly, the Median Shortfall is the tail L^1-median, while the Conditional Tail Expectation is the tail

L^2-median. Like L^p-quantiles, tail L^p-medians with $p > 1$ depend on both the frequency of the event $\{X > q(\alpha)\}$ and the actual behaviour of X beyond $q(\alpha)$. At the technical level, a condition for a tail L^p-median to exist is that $\gamma < 1/(p-1)$, and it can be empirically estimated at high levels by an asymptotically Gaussian estimator if $\gamma < 1/[2(p-1)]$. When $p \in (1,2)$, the tail L^p-median and the L^p-quantile therefore simultaneously exist and are accurately estimable in a wider class of models than the CTE is.

However, there are a number of differences between tail L^p-medians and L^p-quantiles. For example, the theoretical analysis of the asymptotic behaviour of tail L^p-medians, as $\alpha \to 1^-$ (where throughout the paper, "$\to 1^-$" denotes taking a left limit at 1), is technically more complex than that of L^p-quantiles. The asymptotic results that arise show that the tail L^p-median at level α is asymptotically proportional to the quantile $q(\alpha)$, as $\alpha \to 1^-$, through a non-explicit but very accurately approximable constant, and the remainder term in the asymptotic relationship is exclusively controlled by extreme value parameters of X. This stands in contrast with L^p-quantiles, which also are asymptotically proportional to the quantile $q(\alpha)$ through a simpler constant, although the remainder term crucially features the expectation and left-tail behaviour of X. The remainder term plays an important role in the estimation, as it determines the bias term in our eventual estimators, and we will then argue that the extreme value behaviour of tail L^p-medians is easier to understand and more natural than that of L^p-quantiles. We will also explain why, for heavy-tailed models, extreme tail L^p-medians are able to interpolate monotonically between extreme MS and extreme CTE, as p varies in $(1,2)$ and for $\gamma < 1$. By contrast, L^p-quantiles are known not to interpolate monotonically between quantiles and expectiles (see Figure 1 in Daouia *et al.*, 2019). The interpolation property also makes it possible to interpret an extreme tail L^p-median as a weighted average of extreme MS and extreme CTE. This is likely to be helpful as far as the practical applicability of tail L^p-medians is concerned, to the extent that it allows for a simple choice of p reflecting a pre-specified compromise between the extreme MS and CTE.

We shall then examine how to estimate an extreme tail L^p-median. This amounts to estimating a tail L^p-median at a level $\alpha = \alpha_n \to 1^-$ as the size n of the sample of data tends to infinity. We start by suggesting two estimation methods in the so-called intermediate case $n(1 - \alpha_n) \to \infty$. Although the design stage of our tail L^p-median estimators has similarities with that of the L^p-quantile estimators in Daouia *et al.* (2019), the investigation of the asymptotic properties of the tail L^p-median estimators is more challenging and technically involved. These methods will

provide us with basic estimators that we will extrapolate at a proper extreme level α_n, which satisfies $n(1 - \alpha_n) \to c \in [0, \infty)$, using the heavy-tailed assumption. We will, as a theoretical byproduct, demonstrate how our results make it possible to recover or improve upon known results in the literature on Conditional Tail Expectation. Our final step will be to assess the finite-sample behaviour of the suggested estimators. Our focus will not be to consider heavy-tailed models with a small value of γ: for these models and more generally in models with finite fourth moment, it is unlikely that any improvement will be brought on the CTE, whether in terms of quality of estimation or interpretability. Our view is rather that the use of tail L^p-medians with $p \in (1, 2)$ will be beneficial for very heavy-tailed models, in which γ is higher than the finite fourth moment threshold $\gamma = 1/4$, and possibly higher than the finite variance threshold $\gamma = 1/2$. For such values of γ and with an extreme level set to be $\alpha_n = 1 - 1/n$ (a typical choice in applications, used recently in Cai *et al.*, 2015, Gong *et al.*, 2015), we shall then evaluate the finite-sample performance of our estimators on simulated data sets, as well as on a real set of fire insurance data featuring an estimated value of γ larger than $1/2$.

The paper is organised as follows. We first give a rigorous definition of our concept of tail L^p-median and state some of its elementary properties in Section 2. Section 3 then focuses on the analysis of asymptotic properties of the population tail L^p-median, as $\alpha \to 1^-$, in heavy-tailed models. Estimators of an extreme tail L^p-median, obtained by first constructing two distinct estimators at intermediate levels which are then extrapolated to extreme levels, will be discussed in Section 4. A simulation study of the finite-sample performance of our estimators is presented in Section 5, and an application to real fire insurance data is discussed in Section 6. Proofs and auxiliary results are deferred to an online Supplementary Material document.

2   Definition and first properties

Let X be a real-valued random variable with distribution function F and quantile function q given by $q(\alpha) := \inf\{x \in \mathbb{R} \,|\, F(x) \geq \alpha\}$. **It is assumed throughout the paper that F is continuous**, so that $F(q(\alpha)) = \alpha$ for any $\alpha \in (0, 1)$. Our construction is motivated by the following two observations. Firstly, the median and mean of X can respectively be obtained by

minimising an expected absolute deviation and expected squared deviation:

$$
\begin{aligned}
q(1/2) &= \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m|-|X|\right) \\
\text{and } \mathbb{E}(X) &= \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m|^2-|X|^2\right) \ \ (\text{provided } \mathbb{E}|X|<\infty).
\end{aligned}
$$

The first identity is shown in *e.g.* Koenker & Bassett (1978) and Koenker (2005); the minimiser on the right-hand side thereof may actually not be unique (although it is if F is strictly increasing). Our convention throughout this paper will be that in such a situation, the minimiser is taken as the smallest possible minimiser, making the identity valid in any case. Note also that subtracting $|X|$ (resp. $|X|^2$) within the expectation in the cost function for $q(1/2)$ (resp. $\mathbb{E}(X)$) makes it possible to define this cost function, and therefore its minimiser, without assuming any integrability condition on X (resp. by assuming only $\mathbb{E}|X|<\infty$), as a consequence of the triangle inequality (resp. the identity $|X-m|^2-|X|^2 = m(m-2X)$). These optimisation problems extend their arguably better-known formulations

$$
q(1/2) = \arg\min_{m\in\mathbb{R}} \mathbb{E}|X-m| \text{ and } \mathbb{E}(X) = \arg\min_{m\in\mathbb{R}} \mathbb{E}(X-m)^2
$$

which are only well-defined when $\mathbb{E}|X|<\infty$ and $\mathbb{E}(X^2)<\infty$ respectively.

Secondly, since F is continuous, the Median Shortfall $\mathrm{MS}(\alpha) = q([1+\alpha]/2)$ is the median of X given $X>q(\alpha)$, see Example 3 in Kou & Peng (2016). Since $\mathrm{CTE}(\alpha)$ is the expectation of X given $X>q(\alpha)$, we find that

$$
\begin{aligned}
\mathrm{MS}(\alpha) &= \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m|-|X|\,|\,X>q(\alpha)\right) \\
\text{and } \mathrm{CTE}(\alpha) &= \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m|^2-|X|^2\,|\,X>q(\alpha)\right) \ \ (\text{provided } \mathbb{E}(|X|\,|\,X>q(\alpha))<\infty).
\end{aligned}
$$

Our construction now encompasses these two quantities by replacing the absolute or squared deviations by power deviations.

Definition 1. The *tail L^p-median* of X, of order $\alpha \in (0,1)$, is (when it exists)

$$
m_p(\alpha) = \arg\min_{m\in\mathbb{R}} \mathbb{E}\left(|X-m|^p-|X|^p\,|\,X>q(\alpha)\right).
$$

Let us highlight the following important connection between the tail L^p-median and the notion of L^p-quantiles: recall, from Chen (1996) and Daouia *et al.* (2019), that an L^p-quantile of order $\tau \in (0,1)$ of a univariate random variable Y, with $\mathbb{E}|Y|^{p-1}<\infty$, is defined as

$$
q_\tau(p) = \arg\min_{q\in\mathbb{R}} \mathbb{E}\left(|\tau - \mathbb{1}_{\{Y\leq q\}}|\cdot|Y-q|^p - |\tau - \mathbb{1}_{\{Y\leq 0\}}|\cdot|Y|^p\right).
$$

Consequently, the L^p-median of Y, obtained for $\tau = 1/2$, is

$$q_{1/2}(p) = \arg\min_{q \in \mathbb{R}} \mathbb{E}\left(|Y - q|^p - |Y|^p\right).$$

For an arbitrary $p \geq 1$, the tail L^p-median $m_p(\alpha)$ of X is then exactly an L^p-median of X given that $X > q(\alpha)$. This construction of $m_p(\alpha)$ as an L^p-median given the tail event $\{X > q(\alpha)\}$ is what motivated the name "tail L^p-median" for $m_p(\alpha)$.

We underline again that subtracting $|X|^p$ inside the expectation in Definition 1 above makes the cost function well-defined whenever $\mathbb{E}(|X|^{p-1} \mid X > q(\alpha))$ is finite (or equivalently $\mathbb{E}(X_+^{p-1}) < \infty$, where $X_+ := \max(X, 0)$). This is a straightforward consequence of the triangle inequality when $p = 1$; when $p > 1$, this is a consequence of the fact that the function $x \mapsto |x|^p$ is continuously differentiable with derivative $x \mapsto p|x|^{p-1}\,\mathrm{sign}(x)$, together with the mean value theorem. If X is moreover assumed to satisfy $\mathbb{E}(X_+^p) < \infty$, then the definition of $m_p(\alpha)$ is equivalently and perhaps more intuitively

$$m_p(\alpha) = \arg\min_{m \in \mathbb{R}} \mathbb{E}\left(|X - m|^p \mid X > q(\alpha)\right).$$

The following result shows in particular that if $\mathbb{E}(X_+^{p-1}) < \infty$ for some $p \geq 1$, then the tail L^p-median always exists and is characterised by a simple equation. Especially, for $p \in (1, 2)$, the tail L^p-median $m_p(\alpha)$ exists and is unique under a weaker integrability condition than the assumption of a finite first tail moment which is necessary for the existence of $\mathrm{CTE}(\alpha)$.

Proposition 1. *Let $p \geq 1$. Pick $\alpha \in (0, 1)$ and assume that $\mathbb{E}(X_+^{p-1}) < \infty$. Then:*

(i) *The tail L^p-median $m_p(\alpha)$ exists and is such that $m_p(\alpha) > q(\alpha)$.*

(ii) *The tail L^p-median is equivariant with respect to increasing affine transformations: if $m_p^X(\alpha)$ is the tail L^p-median of X and $Y = aX + b$ with $a > 0$ and $b \in \mathbb{R}$, then the tail L^p-median $m_p^Y(\alpha)$ of Y is $m_p^Y(\alpha) = a\, m_p^X(\alpha) + b$.*

(iii) *When $p > 1$, the tail L^p-median $m_p(\alpha)$ is the unique $m \in \mathbb{R}$ solution of the equation*

$$\mathbb{E}[(m - X)^{p-1}\mathbb{1}_{\{q(\alpha) < X < m\}}] = \mathbb{E}[(X - m)^{p-1}\mathbb{1}_{\{X > m\}}].$$

(iv) *The tail L^p-median defines a monotonic functional with respect to first-order stochastic dominance: if Y is another random variable such that $\mathbb{E}(Y_+^{p-1}) < \infty$ then*

$$(\forall t \in \mathbb{R}, \ \mathbb{P}(X > t) \leq \mathbb{P}(Y > t)) \Rightarrow m_p^X(\alpha) \leq m_p^Y(\alpha).$$

(v) When $\mathbb{E}|X|^{p-1} < \infty$, the function $\alpha \mapsto m_p(\alpha)$ is nondecreasing on $(0,1)$.

Let us highlight that, in addition to these properties, we have most importantly

$$\mathrm{MS}(\alpha) = m_1(\alpha) \text{ and } \mathrm{CTE}(\alpha) = m_2(\alpha). \tag{2}$$

In other words, the Median Shortfall is the tail L^1-median and the Conditional Tail Expectation is the tail L^2-median: the class of tail L^p-medians encompasses the notions of MS and CTE. We conclude this section by noting that, like MS, the tail L^p-median is not subadditive (for $1 < p < 2$). Our objective here is not, however, to construct an alternative class of risk measures which have perfect axiomatic properties. There have been several instances when non-subadditive measures were found to be of practical value: besides the non-subadditive VaR and MS, L^p-quantiles define, for $1 < p < 2$, non-subadditive risk measures that can be fruitfully used for, among others, the backtesting of extreme quantile estimates (see Daouia *et al.*, 2019). This work is, rather, primarily intended to provide an interpretable middleway between the risk measures $\mathrm{MS}(\alpha)$ and $\mathrm{CTE}(\alpha)$, for a level α close to 1. This will be useful when the number of finite moments of X is low (typically, less than 4) because the estimation of $\mathrm{CTE}(\alpha)$, for high α, is then a difficult task in practice.

3 Asymptotic properties of an extreme tail L^p-median

It has been found in the literature that heavy-tailed distributions generally constitute appropriate models for the extreme value behaviour of actuarial and financial data, and particularly for extremely high insurance claims and atypical financial log-returns (Embrechts *et al.*, 1997, Resnick, 2007). Our first focus is therefore to analyse whether $m_p(\alpha)$, for $p \in (1,2)$, does indeed provide a middle ground between $\mathrm{MS}(\alpha)$ and $\mathrm{CTE}(\alpha)$, as $\alpha \to 1^-$, in heavy-tailed models. This is done by introducing the following regular variation assumption on the survival function $\overline{F} := 1 - F$ of X.

$\mathcal{C}_1(\gamma)$ There exists $\gamma > 0$ such that for all $x > 0$, $\lim_{t \to \infty} \overline{F}(tx)/\overline{F}(t) = x^{-1/\gamma}$.

In other words, condition $\mathcal{C}_1(\gamma)$ means that the function $\overline{F}$ is regularly varying with index $-1/\gamma$ in a neighbourhood of $+\infty$ (Bingham *et al.*, 1987); for this reason, condition $\mathcal{C}_1(\gamma)$ is equivalent to assuming (1). Note that, most importantly for our purposes, assuming this condition is equivalent to supposing that the tail quantile function $t \mapsto U(t) := q(1 - t^{-1})$ of X is regularly varying with tail index γ. We also remark that when $\mathcal{C}_1(\gamma)$ is satisfied then X does not have any finite conditional tail moments of order larger than $1/\gamma$; similarly, all conditional tail moments of X of

order smaller than $1/\gamma$ are finite (a rigorous statement is Exercise 1.16 in de Haan & Ferreira, 2006). Since the existence of a tail L^p−median requires finite conditional tail moments of order $p - 1$, this means that our minimal working condition on the pair (p, γ) should be $\gamma < 1/(p - 1)$, a condition that will indeed appear in many of our asymptotic results.

We now provide some insight into what asymptotic result on $m_p(\alpha)$ we should aim for under condition $\mathcal{C}_1(\gamma)$. A consequence of this condition is that above a high threshold u, the variable X/u is approximately Pareto distributed with tail index γ, or, in other words:

$$\forall x > 1, \ \mathbb{P}\left(\frac{X}{q(\alpha)} > x \ \middle| \ \frac{X}{q(\alpha)} > 1\right) \to x^{-1/\gamma} \ \text{ as } \ \alpha \to 1^-.$$

This is exactly the first statement of Theorem 1.2.1 in de Haan & Ferreira (2006). The above conditional Pareto approximation then suggests that when α is close enough to 1, the optimisation criterion in Definition 1 can be approximately written as follows:

$$\frac{m_p(\alpha)}{q(\alpha)} \approx \underset{M \in \mathbb{R}}{\arg\min} \ \mathbb{E}\left[|Z_\gamma - M|^p - |Z_\gamma|^p\right]$$

where Z_γ has a Pareto distribution with tail index γ. For the variable Z_γ, we can differentiate the cost function and use change-of-variables formulae to get that the minimiser M of the right-hand side should satisfy the equation $g_{p,\gamma}(1/M) = B(p, \gamma^{-1} - p + 1)$, where $g_{p,\gamma}(t) := \int_t^1 (1 - u)^{p-1} u^{-1/\gamma-1} du$ for $t \in (0, 1)$ and $B(x, y) = \int_0^1 v^{x-1}(1 - v)^{y-1} dv$ denotes the Beta function. In other words, and for a Pareto random variable, we have

$$\frac{m_p(\alpha)}{q(\alpha)} = \frac{1}{\kappa(p, \gamma)} \ \text{ where } \ \kappa(p, \gamma) := g_{p,\gamma}^{-1}(B(p, \gamma^{-1} - p + 1))$$

and $g_{p,\gamma}^{-1}$ denotes the inverse of the decreasing function $g_{p,\gamma}$ on $(0, 1)$. Our first asymptotic result states that this proportionality relationship is still valid asymptotically under condition $\mathcal{C}_1(\gamma)$.

Proposition 2. *Suppose that $p \geq 1$ and $\mathcal{C}_1(\gamma)$ holds with $\gamma < 1/(p - 1)$. Then:*

$$\lim_{\alpha \to 1^-} \frac{m_p(\alpha)}{q(\alpha)} = \frac{1}{\kappa(p, \gamma)}.$$

It follows from Proposition 2 that a tail L^p−median above a high exceedance level is approximately a multiple of this exceedance level. This first-order result is similar in spirit to other asymptotic proportionality relationships linking extreme risk measures to extreme quantiles: we refer to Daouia *et al.* (2018) for a result on extreme expectiles and Daouia *et al.* (2019) on the general class of L^p−quantiles, as well as to Zhu & Li (2012), Yang (2015), El Methni & Stupfler (2017) for a similar

analysis of extreme Wang distortion risk measures. A consequence of this result is that, similarly to extreme L^p-quantiles, an extreme tail L^p-median contains both the information contained in the quantile $q(\alpha)$ plus the information on tail heaviness provided by the tail index γ. This point shall be further used and discussed in Sections 4.2 and 5. Let us also highlight that Proposition 2 does not hold true for $\gamma = 1/(p-1)$, since $\kappa(p, \gamma)$ is then not well-defined.

The asymptotic proportionality constant $\kappa(p, \gamma) \in (0,1)$ does not have a simple closed form in general, due to the complicated expression of the function $g_{p,\gamma}$. It does however have a nice explicit expression in the two particular cases $p = 1$ and $p = 2$. For $p = 1$, we have $\kappa(1, \gamma) = 2^{-\gamma}$, see Lemma 2(ii) in the Supplementary Material document. This clearly yields the same equivalent as the one obtained using the regular variation of the tail quantile function U with index γ, i.e. in virtue of (2):

$$\frac{m_1(\alpha)}{q(\alpha)} = \frac{q((1+\alpha)/2)}{q(\alpha)} = \frac{U(2(1-\alpha)^{-1})}{U((1-\alpha)^{-1})} \to 2^\gamma \text{ as } \alpha \to 1^-.$$

When $p = 2$ and $\gamma \in (0,1)$, $\kappa(2, \gamma) = 1 - \gamma$, see Lemma 3(ii) in the Supplementary Material document. Since in this case, $m_2(\alpha)$ is nothing but $\mathrm{CTE}(\alpha)$ by (2), Proposition 2 agrees here with the asymptotic equivalent of $\mathrm{CTE}(\alpha)$ in terms of the exceedance level $q(\alpha)$, see e.g. Hua & Joe (2011).

For other values of p, an accurate numerical computation of the constant $\kappa(p, \gamma)$ can be carried out instead. Results of such numerical computations on the domain $(p, \gamma) \in [1, 2] \times (0, 1)$ are included in Figure C.1 in the Supplementary Material document. One can observe from this Figure that the functions $p \mapsto \kappa(p, \gamma)$ and $\gamma \mapsto \kappa(p, \gamma)$ seem to be both decreasing. This and (2) entail in particular that, for all $p_1, p_2 \in (1, 2)$ such that $p_1 < p_2$, we have, for α close enough to 1:

$$\mathrm{MS}(\alpha) = m_1(\alpha) < m_{p_1}(\alpha) < m_{p_2}(\alpha) < m_2(\alpha) = \mathrm{CTE}(\alpha).$$

The tail L^p-median $m_p(\alpha)$ can therefore be seen as a risk measure interpolating monotonically between $\mathrm{MS}(\alpha)$ and $\mathrm{CTE}(\alpha)$, at a high enough level α. Actually, Proposition 2 yields, for all $p \in [1, 2]$ and $\gamma < 1/(p-1)$ that, as $\alpha \to 1^-$,

$$m_p(\alpha) \approx \lambda(p, \gamma)\mathrm{MS}(\alpha) + [1 - \lambda(p, \gamma)]\mathrm{CTE}(\alpha), \tag{3}$$

where the weighting constant $\lambda(p, \gamma) \in [0, 1]$ is defined by

$$\lambda(p, \gamma) := \lim_{\alpha \to 1^-} \frac{m_p(\alpha) - \mathrm{CTE}(\alpha)}{\mathrm{MS}(\alpha) - \mathrm{CTE}(\alpha)} = \frac{1 - (1-\gamma)/\kappa(p, \gamma)}{1 - 2^\gamma(1-\gamma)}. \tag{4}$$

Extreme tail L^p-medians of heavy-tailed models can then be interpreted, for $p \in (1,2)$, as weighted averages of extreme Median Shortfall and extreme Conditional Expectation at the same level. It should be noted that, by contrast, the monotonic interpolation property (and hence the weighted average interpretation) is demonstrably false in general for L^p-quantiles, as is most easily seen from Figure 1 in Daouia *et al.* (2019): this Figure suggests that high L^p-quantiles define, for γ close to $1/2$, a decreasing function of p when it is close to 1 and an increasing function of p when it is close to 2.

The fact that $\gamma \mapsto \kappa(p, \gamma)$ is decreasing, meanwhile, can be proven rigorously by noting that its partial derivative $\partial \kappa / \partial \gamma$ is negative (see Theorem 2 below). More intuitively, the monotonicity of $\gamma \mapsto \kappa(p, \gamma)$ can be seen as a consequence of the heavy-tailedness of the distribution function F. Indeed, we saw that heuristically, as $\alpha \to 1^-$,

$$\frac{m_p(\alpha)}{q(\alpha)} \approx \arg\min_{M \in \mathbb{R}} \mathbb{E}\left[|Z_\gamma - M|^p - |Z_\gamma|^p\right]$$

where Z_γ is a Pareto random variable with tail index γ. When γ increases, the random variable Z_γ tends to return higher values because its survival function $\mathbb{P}(Z_\gamma > z) = z^{-1/\gamma}$ (for $z > 1$) is an increasing function of γ. We can therefore expect that, as γ increases, a higher value of $M = m_p(\alpha)/q(\alpha)$ will be needed in order to minimise the above cost function.

Our next goal is to derive an asymptotic expansion of the tail L^p-median $m_p(\alpha)$, relatively to the high exceedance level $q(\alpha)$. This will be the key theoretical tool making it possible to analyse the asymptotic properties of estimators of an extreme tail L^p-median. For this, we need to quantify precisely the error term in the convergence given by Proposition 2, and this prompts us to introduce the following second-order regular variation condition:

$\mathcal{C}_2(\gamma, \rho, A)$ The function $\overline{F}$ is second-order regularly varying in a neighbourhood 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,

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

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

This standard condition on $\overline{F}$ controls the rate of convergence in $\mathcal{C}_1(\gamma)$: the larger $|\rho|$ is, the faster the function $|A|$ converges to 0 (since $|A|$ is regularly varying with index ρ, in virtue of Theorems 2.3.3 and 2.3.9 in de Haan & Ferreira, 2006) and the smaller the error in the approximation

of the right tail of X by a purely Pareto tail will be. Further interpretation of this assumption can be found in Beirlant *et al.* (2004) and de Haan & Ferreira (2006) along with numerous examples of commonly used continuous distributions satisfying it. Let us finally mention that it is a consequence of Theorem 2.3.9 in de Haan & Ferreira (2006) that $\mathcal{C}_2(\gamma, \rho, A)$ is actually equivalent to the following, perhaps more usual extremal assumption on the tail quantile function U:

$$\forall x > 0, \ \lim_{t \to \infty} \frac{1}{A(t)} \left[\frac{U(tx)}{U(t)} - x^\gamma \right] = x^\gamma \frac{x^\rho - 1}{\rho}. \tag{5}$$

In (5), the right-hand side should be read as $x^\gamma \log x$ when $\rho = 0$ (similarly, throughout this paper, quantities depending on ρ are extended by continuity using their left limit as $\rho \to 0^-$). The next result is the desired refinement of Proposition 2 giving the error term in the asymptotic proportionality relationship linking $m_p(\alpha)$ to $q(\alpha)$ when $\alpha \to 1^-$.

Proposition 3. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma, \rho, A)$ holds with $\gamma < 1/(p-1)$. Then, as $\alpha \to 1^-$,*

$$\frac{m_p(\alpha)}{q(\alpha)} = \frac{1}{\kappa(p, \gamma)} \left(1 + [R(p, \gamma, \rho) + o(1)] A \left((1-\alpha)^{-1} \right) \right)$$

with

$$R(p, \gamma, \rho) = \frac{[\kappa(p, \gamma)]^{-(\rho-1)/\gamma}}{(1 - \kappa(p, \gamma))^{p-1}} \times \frac{1-\rho}{\gamma \rho} \left[B(p, (1-\rho)\gamma^{-1} - p + 1) - g_{p, \gamma/(1-\rho)}(\kappa(p, \gamma)) \right].$$

This result is similar in spirit to second-order results that have been shown for other extreme risk measures: we refer again to Zhu & Li (2012) and Yang (2015), as well as to El Methni & Stupfler (2017), Daouia *et al.* (2018, 2019) for analogue results used as a basis to carry out extreme-value based inference on other types of indicators. It should, however, be underlined that the asymptotic expansion of an extreme tail L^p-median depends solely on the extreme parameters γ, ρ and A, along with the power p. By contrast, the asymptotic expansion of an extreme L^p-quantile depends on the expectation and left-tail behaviour of X, which are typically considered to be irrelevant to the understanding of the right tail of X. From an extreme value point of view, the asymptotic expansion of extreme tail L^p-medians is therefore easier to understand than that of L^p-quantiles. Statistically speaking, it also implies that there are less sources of potential bias in extreme tail L^p-median estimation than in extreme L^p-quantile estimation. Both of these statements can be explained by the fact that the tail L^p-median is constructed exclusively on the event $\{X > q(\alpha)\}$, while the equation defining an L^p-quantile $q_p(\alpha)$ is

$$(1 - \alpha)\mathbb{E}((q_p(\alpha) - X)^{p-1} \mathbb{1}_{\{X < q_p(\alpha)\}}) = \alpha \mathbb{E}((X - q_p(\alpha))^{p-1} \mathbb{1}_{\{X > q_p(\alpha)\}})$$

(Chen, 1996, Section 2). The left-hand side term ensures that the central and left-tail behaviour of X will necessarily have an influence on the value of any L^p−quantile, even at an extreme level. The construction of a tail L^p−median as an L^p−median in the right tail of X removes this issue.

Like the asymptotic proportionality constant $\kappa(p, \gamma)$ on which it depends, the remainder term $R(p, \gamma, \rho)$ does not have an explicit form in general. That being said, $R(1, \gamma, \rho)$ and $R(2, \gamma, \rho)$ have simple explicit values: for $p = 1$, $R(1, \gamma, \rho) = (2^\rho - 1)/\rho$ (with the convention $(x^\rho - 1)/\rho = \log x$ for $\rho = 0$), see Lemma 2(iii) in the Supplementary Material document. We then find back the result that comes as an immediate consequence of (2) and our second-order condition $\mathcal{C}_2(\gamma, \rho, A)$, via (5):

$$\frac{m_1(\alpha)}{q(\alpha)} = \frac{U(2(1-\alpha)^{-1})}{U((1-\alpha)^{-1})} = 2^\gamma \left(1 + A((1-\alpha)^{-1})\frac{2^\rho - 1}{\rho}(1 + \mathrm{o}(1))\right).$$

For $p = 2$ and $\gamma \in (0, 1)$, (2) and (5) suggest that:

$$\frac{m_2(\alpha)}{q(\alpha)} = \int_1^\infty \frac{U((1-\alpha)^{-1}x)}{U((1-\alpha)^{-1})}\frac{dx}{x^2} = \frac{1}{1-\gamma}\left(1 + \frac{1}{1-\gamma-\rho}A((1-\alpha)^{-1})(1 + \mathrm{o}(1))\right)$$

which coincides with Proposition 3, given that $R(2, \gamma, \rho) = 1/(1-\gamma-\rho)$ from Lemma 3(iii) in the Supplementary Material document.

We close this section by noting that all our results, and indeed the practical use of the tail L^p−median more generally, depend on the fixed value of the constant p. Just as when using L^p−quantiles, the choice of p in practice is a difficult but important question. Although Chen (1996) introduced L^p−quantiles in the context of testing for symmetry in non-parametric regression, it did not investigate the question of the choice of p. In Daouia *et al.* (2019), extreme L^p−quantiles were used as vehicles for the estimation of extreme quantiles and expectiles and for extreme quantile forecast validation; in connection with the latter, it is observed that neither $p = 1$ nor $p = 2$ provide the best performance in terms of forecast, but no definitive conclusion is reached as to which value of p should be chosen (see Section 7 therein). For extreme tail L^p−medians, which unlike extreme L^p−quantiles satisfy an interpolation property, we may suggest a potentially simpler and intuitive way to choose p. Recall the weighted average relationship (3):

$$m_p(\alpha) \approx \lambda(p, \gamma)\mathrm{MS}(\alpha) + [1 - \lambda(p, \gamma)]\mathrm{CTE}(\alpha), \text{ with } \lambda(p, \gamma) = \frac{1 - (1-\gamma)/\kappa(p, \gamma)}{1 - 2^\gamma(1-\gamma)}$$

for α close to 1. In practice, given the (estimated) value of γ, and a pre-specified weighting constant λ_0 indicating a compromise between robustness of MS and sensitivity of CTE, one can choose $p = p_0$ as the unique root of the equation $\lambda(p, \gamma) = \lambda_0$ with unknown p. Although $\lambda(p, \gamma)$

does not have a simple closed form, our experience shows that this equation can be solved very quickly and accurately with standard numerical solvers. This results in a tail L^p-median $m_{p_0}(\alpha)$ satisfying, for α close to 1,

$$m_{p_0}(\alpha) \approx \lambda_0 \text{MS}(\alpha) + (1 - \lambda_0)\text{CTE}(\alpha).$$

The interpretation of $m_{p_0}(\alpha)$ is easier than that of a generic tail L^p-median $m_p(\alpha)$ due to its explicit and fully-determined connection with the two well-understood quantities $\text{MS}(\alpha)$ and $\text{CTE}(\alpha)$. The question of which weighting constant λ_0 should be chosen is itself difficult and depends on the requirements of the situation at hand; our real data application in Section 6 provides an illustration with $\lambda_0 = 1/2$, corresponding to a simple average between extreme MS and CTE.

4   Estimation of an extreme tail L^p-median

Suppose that we observe a random sample $(X_1, \ldots, X_n)$ of independent copies of X, and denote by $X_{1,n} \leq \cdots \leq X_{n,n}$ the corresponding set of order statistics arranged in increasing order. Our goal in this section is to estimate an extreme tail L^p-median $m_p(\alpha_n)$, where $\alpha_n \to 1^-$ as $n \to \infty$. The final aim is to allow α_n to approach 1 at any rate, covering the cases of an intermediate tail L^p-median with $n(1 - \alpha_n) \to \infty$ and of a proper extreme tail L^p-median with $n(1 - \alpha_n) \to c$, where c is some finite positive constant.

4.1   Intermediate case: direct estimation by empirical L^p-optimisation

Recall that the tail L^p-median $m_p(\alpha_n)$ is, by Definition 1,

$$m_p(\alpha_n) = \underset{m \in \mathbb{R}}{\arg\min}\, \mathbb{E}\left(|X - m|^p - |X|^p \,|\, X > q(\alpha_n)\right).$$

Assume here that $n(1 - \alpha_n) \to \infty$, so that $\lfloor n(1 - \alpha_n) \rfloor > 0$ eventually. We can therefore define a direct empirical tail L^p-median estimator of $m_p(\alpha_n)$ by minimising the above empirical cost function:

$$\widehat{m}_p(\alpha_n) = \underset{m \in \mathbb{R}}{\arg\min}\, \frac{1}{\lfloor n(1 - \alpha_n) \rfloor} \sum_{i=1}^{\lfloor n(1-\alpha_n) \rfloor} \left(|X_{n-i+1,n} - m|^p - |X_{n-i+1,n}|^p\right). \tag{6}$$

We now pave the way for a theoretical study of this estimator. The key point is that since normalising constants and shifts are irrelevant in the definition of the empirical criterion, we

clearly have the equivalent definition

$$\widehat{m}_p(\alpha_n) = \arg\min_{m\in\mathbb{R}} \frac{1}{p[m_p(\alpha_n)]^p} \sum_{i=1}^{\lfloor n(1-\alpha_n)\rfloor} \left(|X_{n-i+1,n}-m|^p - |X_{n-i+1,n}-m_p(\alpha_n)|^p\right).$$

Consequently

$$\sqrt{n(1-\alpha_n)}\left(\frac{\widehat{m}_p(\alpha_n)}{m_p(\alpha_n)}-1\right) = \arg\min_{u\in\mathbb{R}} \psi_n(u;p)$$

with

$$\psi_n(u;p) = \frac{1}{p[m_p(\alpha_n)]^p} \sum_{i=1}^{\lfloor n(1-\alpha_n)\rfloor} \left(\left|X_{n-i+1,n}-m_p(\alpha_n)-\frac{um_p(\alpha_n)}{\sqrt{n(1-\alpha_n)}}\right|^p - |X_{n-i+1,n}-m_p(\alpha_n)|^p\right).$$
(7)

Note that the empirical criterion $\psi_n(u;p)$ is a continuous and convex function of u, so that the asymptotic properties of the minimiser follow directly from those of the criterion itself by the convexity lemmas of Geyer (1996) and Knight (1999). The empirical criterion, then, is analysed by using its continuous differentiability (for $p > 1$) in order to formulate an L^p-analogue of Knight's identity (Knight, 1998) and divide the work between, on the one hand, the study of a $\sqrt{n(1-\alpha_n)}-$consistent and asymptotically Gaussian term which is an affine function of u and, on the other hand, a bias term which converges to a nonrandom multiple of u^2. Further technical details are provided in the Supplementary Material document, see in particular Lemmas 6, 7, 9 and 11.

This programme of work is broadly similar to that of Daouia *et al.* (2019) for the convergence of the direct intermediate L^p-quantile estimator. The difficulty in this particular case, however, is twofold: first, the affine function of u is a generalised $L-$statistic (in the sense of for instance Borisov & Baklanov, 2001) whose analysis requires delicate arguments relying on the asymptotic behaviour of the tail quantile process *via* Theorem 5.1.4 p.161 in de Haan & Ferreira (2006). For L^p-quantiles, this is not necessary because the affine term is actually a sum of independent, identically distributed and centred variables. Second, the bias term is essentially a doubly integrated oscillation of a power function with generally noninteger exponent. The examination of its convergence requires certain precise real analysis arguments which do not follow from those developed in Daouia *et al.* (2019) for the asymptotic analysis of intermediate L^p-quantiles.

With this in mind, the asymptotic normality result for the direct intermediate tail L^p-median estimator $\widehat{m}_p(\alpha_n)$ is the following.

Theorem 1. *Suppose that $p \geq 1$ and $C_2(\gamma, \rho, A)$ holds with $\gamma < 1/[2(p-1)]$. Assume further that $\alpha_n \to 1^-$ is such that $n(1 - \alpha_n) \to \infty$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) = O(1)$. Then we have, as $n \to \infty$:*

$$\sqrt{n(1-\alpha_n)}\left(\frac{\widehat{m}_p(\alpha_n)}{m_p(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, V(p, \gamma)\right).$$

Here $V(p, \gamma) = V_1(p, \gamma)/V_2(p, \gamma)$ with

$$V_1(p, \gamma) = \frac{[\kappa(p, \gamma)]^{1/\gamma}}{\gamma}\left(B(2p - 1, \gamma^{-1} - 2(p-1)) + g_{2p-1,\gamma}(\kappa(p, \gamma))\right) + [1 - \kappa(p, \gamma)]^{2(p-1)}$$

while $V_2(p, \gamma)$ is defined by: $V_2(1, \gamma) = 1/\gamma^2$ and, for $p > 1$,

$$V_2(p, \gamma) = \left(\frac{p-1}{\gamma}[\kappa(p, \gamma)]^{1/\gamma}\left[B(p - 1, \gamma^{-1} - p + 2) + g_{p-1,\gamma}(\kappa(p, \gamma))\right]\right)^2.$$

Moreover, the functions $V_2(\cdot, \gamma)$ and $V(\cdot, \gamma)$ defined this way are right-continuous at 1.

The asymptotic variance $V(p, \gamma)$ has a rather involved expression. Figures C.2 and C.3 in the Supplementary Material document provide graphical representations of this variance term.

It can be seen on Figure C.2 that the function $\gamma \mapsto V(p, \gamma)$ appears to be increasing. This reflects the increasing tendency of the underlying distribution to generate extremely high observations when the tail index increases (El Methni & Stupfler, 2018, Section 3.1), thus increasing the variability of the empirical criterion $\psi_n(\cdot; p)$ and consequently that of its minimiser. It is not, however, clear from this figure that the function $p \mapsto V(p, \gamma)$ is monotonic, like the proportionality constant κ was. It turns out that, somewhat surprisingly, the function $p \mapsto V(p, \gamma)$ is not in general a monotonic function of p, and an illustration is provided for $\gamma = 0.3$ in Figure C.3. A numerical study, which is not reported here, actually shows that we have $V(p, \gamma) < V(1, \gamma)$ for any $(p, \gamma) \in (1, 1.2] \times [0.25, 0.5]$. This suggests that for all heavy-tailed distributions having only a second moment (an already difficult case as far as estimation in heavy-tailed models is concerned), a direct L^p-tail median estimator with $p \in (1, 1.2]$ will have a smaller asymptotic variance than the empirical L^1-tail median estimator, or, in other words, the empirical Median Shortfall.

We conclude this section by noting that, like the constants appearing in our previous asymptotic results, the variance term $V(p, \gamma)$ has a simple expression when $p = 1$ or $p = 2$. In the case $p = 1$, we have $V(1, \gamma) = 2\gamma^2$ by Lemma 2(iv) in the Supplementary Material document. Statement (2) suggests that this should be identical to the asymptotic variance of the high quantile estimator $\widehat{MS}(\alpha_n) = X_{n-\lfloor n(1-\alpha_n)/2\rfloor,n}$, and indeed

$$\sqrt{n(1-\alpha_n)}\left(\frac{\widehat{MS}(\alpha_n)}{MS(\alpha_n)} - 1\right) = \sqrt{2}\sqrt{\frac{n(1-\alpha_n)}{2}}\left(\frac{X_{n-\lfloor n(1-\alpha_n)/2\rfloor,n}}{q(1 - (1-\alpha_n)/2)} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, 2\gamma^2\right)$$

16

by Theorem 2.4.8 p.52 in de Haan & Ferreira (2006). For $p = 2$, we have $V(2, \gamma) = 2\gamma^2(1-\gamma)/(1-2\gamma)$ by Lemma 3(iv) in the Supplementary Material document. By (2), we should expect this constant to coincide with the asymptotic variance of the empirical counterpart of the Conditional Tail Expectation, namely

$$\widehat{\text{CTE}}(\alpha_n) = \frac{1}{\lfloor n(1-\alpha_n) \rfloor} \sum_{i=1}^{\lfloor n(1-\alpha_n) \rfloor} X_{n-i+1,n}.$$

This is indeed the case as Corollary 1 in El Methni *et al.* (2014) shows; see also Theorem 2 in El Methni *et al.* (2018). In particular, the function $\gamma \mapsto V(2, \gamma)$ tends to infinity as $\gamma \uparrow 1/2$, reflecting the increasing difficulty of estimating a high Conditional Tail Expectation by its direct empirical counterpart as the right tail of X gets heavier.

4.2 Intermediate case: indirect quantile-based estimation

We can also design an estimator of $m_p(\alpha_n)$ based on the asymptotic equivalence between $m_p(\alpha_n)$ and $q(\alpha_n)$ that is provided by Proposition 2. Indeed, since this result suggests that $m_p(\alpha)/q(\alpha) \sim 1/\kappa(p, \gamma)$ when $\alpha \to 1^-$ (with $\sim$ denoting asymptotic equivalence throughout), it makes sense to build a plug-in estimator of $m_p(\alpha_n)$ by setting $\widetilde{m}_p(\alpha_n) = \widehat{q}(\alpha_n)/\kappa(p, \widehat{\gamma}_n)$, where $\widehat{q}(\alpha_n)$ and $\widehat{\gamma}_n$ are respectively two consistent estimators of the high quantile $q(\alpha_n)$ and of the tail index γ. Since we work here in the intermediate case $n(1-\alpha_n) \to \infty$, we know that the sample counterpart $X_{\lceil n\alpha_n \rceil, n}$ of $q(\alpha_n)$ is a relatively consistent estimator of $q(\alpha_n)$, see Theorem 2.4.1 in de Haan & Ferreira (2006). This suggests to use the estimator

$$\widetilde{m}_p(\alpha_n) := \frac{X_{\lceil n\alpha_n \rceil, n}}{\kappa(p, \widehat{\gamma}_n)}. \tag{8}$$

Our next result analyses the asymptotic distribution of this estimator, assuming that the pair $(\widehat{\gamma}_n, X_{\lceil n\alpha_n \rceil, n})$ is jointly $\sqrt{n(1-\alpha_n)}-$consistent.

Theorem 2. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma, \rho, A)$ holds with $\gamma < 1/(p-1)$. Assume further that $\alpha_n \to 1^-$ is such that $n(1-\alpha_n) \to \infty$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) \to \lambda \in \mathbb{R}$, and that*

$$\sqrt{n(1-\alpha_n)} \left(\widehat{\gamma}_n - \gamma, \frac{X_{\lceil n\alpha_n \rceil, n}}{q(\alpha_n)} - 1 \right) \xrightarrow{d} (\xi_1, \xi_2)$$

where (ξ_1, ξ_2) is a pair of nondegenerate random variables. Then we have, as $n \to \infty$:

$$\sqrt{n(1-\alpha_n)} \left(\frac{\widetilde{m}_p(\alpha_n)}{m_p(\alpha_n)} - 1 \right) \xrightarrow{d} \sigma(p, \gamma)\xi_1 + \xi_2 - \lambda R(p, \gamma, \rho),$$

with the positive constant $\sigma(p, \gamma)$ being $\sigma(p, \gamma) = -\dfrac{1}{\kappa(p, \gamma)}\dfrac{\partial \kappa}{\partial \gamma}(p, \gamma)$. In other words,

$$\sigma(p, \gamma) = \frac{B(p, \gamma^{-1} - p + 1)[\Psi(\gamma^{-1} + 1) - \Psi(\gamma^{-1} - p + 1)] - \int_{\kappa(p,\gamma)}^{1}(1 - u)^{p-1}u^{-1/\gamma - 1}\log(u)\, du}{\gamma^2[1 - \kappa(p, \gamma)]^{p-1}[\kappa(p, \gamma)]^{-1/\gamma}}$$

where $\Psi(x) = \Gamma'(x)/\Gamma(x)$ is Euler's digamma function.

Again, the constant $\sigma(p, \gamma)$ does not generally have a simple explicit form, but we can compute it when $p = 1$ or $p = 2$. Lemma 2(v) shows that $\sigma(1, \gamma) = \log 2$, while Lemma 3(v) entails $\sigma(2, \gamma) = 1/(1 - \gamma)$ (see the Supplementary Material document). Contrary to our previous analyses, it is more difficult to relate these constants to pre-existing results in high quantile or high CTE estimation because Theorem 2 is a general result that applies to a wide range of estimators $\widehat{\gamma}_n$. To the best of our knowledge, there is no general analogue of this result in the literature for the case $p = 1$. In the case $p = 2$, we find back the asymptotic distribution result in Theorem 1 of El Methni & Stupfler (2017):

$$\sqrt{n(1 - \alpha_n)}\left(\frac{\widetilde{m}_2(\alpha_n)}{\mathrm{CTE}(\alpha_n)} - 1\right) \xrightarrow{d} \frac{1}{1 - \gamma}\xi_1 + \xi_2 - \frac{\lambda}{1 - \gamma - \rho}.$$

It should be highlighted that, for $p = 2$, Theorem 2 in the present paper is a stronger result than Theorem 1 of El Methni & Stupfler (2017), since the condition on γ is less stringent than that of Theorem 1 therein. As the above convergence is valid for any $\gamma < 1$, one may therefore think that the estimator $\widetilde{m}_2(\alpha_n)$ is a widely applicable estimator of the CTE at high levels. Since it is also more robust than the direct empirical CTE estimator due to its reliance on the sample quantile $X_{\lceil n\alpha_n \rceil, n}$, this would defeat the point of looking for a middle ground solution between the sensitivity of CTE to high values and the robustness of VaR-type measures in very heavy-tailed models. The simulation study in El Methni & Stupfler (2017) shows however that in general, the estimator $\widetilde{m}_2$ fares worse than the direct CTE estimator $\widehat{m}_2$, and increasingly so as γ increases within the range $(0, 1/4]$. We will also confirm this in our simulation study by considering several cases with higher values of γ and showing that the estimator $\widetilde{m}_2$ should in general not be preferred to $\widehat{m}_2$. The benefit of using the indirect estimator $\widetilde{m}_p$ will rather be found for values of p away from 2, when a genuine compromise between sensitivity and robustness is sought.

Theorem 2 applies whenever $\widehat{\gamma}_n$ is a consistent estimator of γ that satisfies a joint convergence condition together with the intermediate order statistic $X_{\lceil n\alpha_n \rceil, n}$. This is not a restrictive requirement.

For instance, if $\widehat{\gamma}_n = \widehat{\gamma}_n^H$ is the widely used Hill estimator (Hill, 1975):

$$\widehat{\gamma}_n^H := \frac{1}{\lfloor n(1-\alpha_n)\rfloor} \sum_{i=1}^{\lfloor n(1-\alpha_n)\rfloor} \log\left(X_{n-i+1,n}\right) - \log\left(X_{n-\lfloor n(1-\alpha_n)\rfloor,n}\right), \tag{9}$$

then, under the bias condition $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) \to \lambda$, we have the joint convergence

$$\sqrt{n(1-\alpha_n)}\left(\widehat{\gamma}_n^H - \gamma, \frac{X_{\lceil n\alpha_n\rceil,n}}{q(\alpha_n)} - 1\right) \xrightarrow{d} \left(\gamma N_1 + \frac{\lambda}{1-\rho}, \gamma N_2\right)$$

where (N_1, N_2) is a pair of independent standard Gaussian random variables (for a proof, combine Theorem 2.4.1, Lemma 3.2.3 and Theorem 3.2.5 in de Haan & Ferreira, 2006). As a corollary of Theorem 2, we then get the following asymptotic result on $\widetilde{m}_p(\alpha_n)$ when the estimator $\widehat{\gamma}_n$ is the Hill estimator $\widehat{\gamma}_n^H$.

Corollary 1. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma, \rho, A)$ holds with $\gamma < 1/(p-1)$. Assume further that $\alpha_n \to 1^-$ is such that $n(1-\alpha_n) \to \infty$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) \to \lambda \in \mathbb{R}$. Then, as $n \to \infty$:*

$$\sqrt{n(1-\alpha_n)}\left(\frac{\widetilde{m}_p(\alpha_n)}{m_p(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}\left(\lambda\left(\frac{\sigma(p,\gamma)}{1-\rho} - R(p,\gamma,\rho)\right), \mathfrak{v}(p,\gamma)\right),$$

where $\mathfrak{v}(p,\gamma) = \gamma^2\left([\sigma(p,\gamma)]^2 + 1\right)$.

The asymptotic variance $\mathfrak{v}(p,\gamma)$ is plotted, for several values of p, on Figure C.4 in the Supplementary Material document, and a comparison of the asymptotic variance $V(p,\gamma)$ of the direct estimator with $\mathfrak{v}(p,\gamma)$ is depicted on Figure C.5 of the Supplementary Material document, for $\gamma \in [1/4, 1/2]$.

It can be seen from these figures that the indirect estimator has a lower variance than the direct one. The difference between the two variances becomes sizeable when the quantity $2\gamma(p-1)$ gets closer to 1, as should be expected since the asymptotic variance of the direct estimator asymptotically increases to infinity (see Theorem 1), while the asymptotic variance of the indirect estimator is kept under control (see Corollary 1). This seems to indicate that the indirect estimator should be preferred to the direct estimator in terms of variability. However, the indirect estimator is asymptotically biased (due to the use of the approximation $m_p(\alpha)/q(\alpha) \sim 1/\kappa(p,\gamma)$ in its construction), while the direct estimator is not. We will see that this can make one prefer the direct estimator in terms of mean squared error on finite samples, even for large values of γ and p.

We conclude this section by mentioning that similar joint convergence results on the pair $(\widehat{\gamma}_n, X_{\lceil n\alpha_n\rceil,n})$, and therefore analogues of Corollary 1, can be found for a wide range of other estimators of γ.

We mention for instance the maximum likelihood estimator in an approximate Generalised Pareto model for exceedances and probability-weighted moment estimators; we refer to *e.g.* Sections 3 and 4 of de Haan & Ferreira (2006) for an asymptotic analysis of these alternatives. This can be used to construct several other indirect tail L^p-median estimators.

4.3   Extreme case: an extrapolation device

Both the direct and indirect estimators constructed so far are only consistent for intermediate sequences α_n such that $n(1 - \alpha_n) \to \infty$. Our purpose is now to extrapolate these intermediate tail L^p-median estimators to proper extreme levels $\beta_n \to 1^-$ with $n(1 - \beta_n) \to c < \infty$ as $n \to \infty$. The extrapolation argument is based on the fact that under the regular variation condition $\mathcal{C}_1(\gamma)$, the function $t \mapsto U(t) = q(1 - t^{-1})$ is regularly varying with index γ. In particular, we have:

$$\frac{q(\beta_n)}{q(\alpha_n)} = \frac{U((1 - \beta_n)^{-1})}{U((1 - \alpha_n)^{-1})} \approx \left(\frac{1 - \beta_n}{1 - \alpha_n} \right)^{-\gamma}$$

when $\alpha_n, \beta_n \to 1^-$. This approximation is at the heart of the construction of the classical Weissman extreme quantile estimator $\widehat{q}^W(\beta_n)$, introduced in Weissman (1978):

$$\widehat{q}^W(\beta_n) := \left(\frac{1 - \beta_n}{1 - \alpha_n} \right)^{-\widehat{\gamma}_n} X_{\lceil n\alpha_n \rceil, n}.$$

The key point is then that, when $\gamma < 1/(p-1)$, the quantity $m_p(\alpha)$ is asymptotically proportional to $q(\alpha)$, by Proposition 2. Combining this with the above approximation on ratios of high quantiles suggests the following extrapolation formula:

$$m_p(\beta_n) \approx \left(\frac{1 - \beta_n}{1 - \alpha_n} \right)^{-\gamma} m_p(\alpha_n).$$

An estimator of the extreme tail L^p-median $m_p(\beta_n)$ is obtained from this approximation by plugging in a consistent estimator $\widehat{\gamma}_n$ of γ and a consistent estimator of $m_p(\alpha_n)$. In our context, the latter can be the direct, empirical L^p-estimator $\widehat{m}_p(\alpha_n)$, or the indirect, intermediate quantile-based estimator $\widetilde{m}_p(\alpha_n)$, yielding the extrapolated estimators

$$\widehat{m}_p^W(\beta_n) := \left(\frac{1 - \beta_n}{1 - \alpha_n} \right)^{-\widehat{\gamma}_n} \widehat{m}_p(\alpha_n) \;\; \text{and} \;\; \widetilde{m}_p^W(\beta_n) := \left(\frac{1 - \beta_n}{1 - \alpha_n} \right)^{-\widehat{\gamma}_n} \widetilde{m}_p(\alpha_n).$$

We note, moreover, that the latter estimator is precisely the estimator deduced by plugging the Weissman extreme quantile estimator $\widehat{q}^W(\beta_n)$ in the asymptotic proportionality relationship $m_p(\beta_n)/q(\beta_n) \sim 1/\kappa(p, \gamma)$, since

$$\widetilde{m}_p^W(\beta_n) = \left(\frac{1 - \beta_n}{1 - \alpha_n} \right)^{-\widehat{\gamma}_n} \times \left\{ \frac{X_{\lceil n\alpha_n \rceil, n}}{\kappa(p, \widehat{\gamma}_n)} \right\} = \frac{\widehat{q}^W(\beta_n)}{\kappa(p, \widehat{\gamma}_n)}.$$

The asymptotic behaviour of the two extrapolated estimators $\widehat{m}_p^W(\beta_n)$ or $\widetilde{m}_p^W(\beta_n)$ is analysed in our next main result.

Theorem 3. *Suppose that $p \geq 1$ and $\mathcal{C}_2(\gamma, \rho, A)$ holds with $\rho < 0$. Assume also that $\alpha_n, \beta_n \to 1^-$ are such that $n(1-\alpha_n) \to \infty$ and $n(1-\beta_n) \to c < \infty$, with $\sqrt{n(1-\alpha_n)}/\log([1-\alpha_n]/[1-\beta_n]) \to \infty$. Assume finally that $\sqrt{n(1-\alpha_n)}(\widehat{\gamma}_n - \gamma) \xrightarrow{d} \xi$, where ξ is a nondegenerate limiting random variable.*

(i) *If $\gamma < 1/[2(p-1)]$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) = O(1)$ then, as $n \to \infty$:*

$$\frac{\sqrt{n(1-\alpha_n)}}{\log[(1-\alpha_n)/(1-\beta_n)]}\left(\frac{\widehat{m}_p^W(\beta_n)}{m_p(\beta_n)} - 1\right) \xrightarrow{d} \xi.$$

(ii) *If $\gamma < 1/(p-1)$ and $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) \to \lambda \in \mathbb{R}$ then, as $n \to \infty$:*

$$\frac{\sqrt{n(1-\alpha_n)}}{\log[(1-\alpha_n)/(1-\beta_n)]}\left(\frac{\widetilde{m}_p^W(\beta_n)}{m_p(\beta_n)} - 1\right) \xrightarrow{d} \xi.$$

This result shows that both of the estimators $\widehat{m}_p^W(\beta_n)$ and $\widetilde{m}_p^W(\beta_n)$ have their asymptotic properties governed by those of the tail index estimator $\widehat{\gamma}_n$. This is not an unusual phenomenon for extrapolated estimators: actually, the very fact that these two estimators are built on an intermediate tail L^p−median estimator and a tail index estimator $\widehat{\gamma}_n$ sharing the same rate of convergence guarantees that the asymptotic behaviour of $\widehat{\gamma}_n$ will dominate. A brief, theoretical justification for this is that while the intermediate tail L^p−median estimator is $\sqrt{n(1-\alpha_n)}$−relatively consistent, the (estimated) extrapolation factor $([1-\beta_n]/[1-\alpha_n])^{-\widehat{\gamma}_n}$, whose asymptotic behaviour only depends on that of $\widehat{\gamma}_n$, converges relatively to $([1-\beta_n]/[1-\alpha_n])^{-\gamma}$ with the slower rate of convergence $\sqrt{n(1-\alpha_n)}/\log[(1-\alpha_n)/(1-\beta_n)]$. This is explained in detail in the proof of Theorem 3, and we also refer to Theorem 4.3.8 of de Haan & Ferreira (2006) and its proof for a detailed exposition regarding the Weissman quantile estimator. In particular, if $\widehat{\gamma}_n$ is the Hill estimator (9), then the common asymptotic distribution of our extrapolated estimators will be Gaussian with mean $\lambda/(1-\rho)$ and variance γ^2, provided $\sqrt{n(1-\alpha_n)}A((1-\alpha_n)^{-1}) \to \lambda \in \mathbb{R}$, see Theorem 3.2.5 in de Haan & Ferreira (2006).

Let us highlight though that while the asymptotic behaviour of $\widehat{\gamma}_n$ is crucial, we should anticipate that in finite-sample situations, an accurate estimation of the intermediate tail L^p−median $m_p(\alpha_n)$ is also important. A mathematical reason for this is that in the typical situation when $1 - \beta_n = 1/n$ (considered recently by for instance Cai *et al.*, 2015, Gong *et al.*, 2015), the logarithmic

term $\log[(1-\alpha_n)/(1-\beta_n)]$ has order at most $\log(n)$, and thus for a moderately high sample size n, the quantity $\sqrt{n(1-\alpha_n)}/\log[(1-\alpha_n)/(1-\beta_n)]$ representing the rate of convergence of the extrapolation factor may only be slightly lower than the quantity $\sqrt{n(1-\alpha_n)}$ representing the rate of convergence of the estimator at the intermediate step. Hence the idea that, while for n very large the difference in finite-sample behaviour between any two estimators of the tail L^p-median at the basic intermediate level α_n will be eventually wiped out by the performance of the estimator $\widehat{\gamma}_n$, there may still be a significant impact of the quality of the intermediate tail L^p-median estimator used on the overall accuracy of the extrapolated estimator when n is moderately large. This is illustrated in the simulation study below.

5   Simulation study

Our goal in the present section is to assess the finite-sample performance of our direct and indirect estimators of an extreme tail L^p-median, for $p \in [1,2]$. In addition, we shall do so in a way that provides guidance as to how an extreme tail L^p-median, and its estimates, can be used and interpreted in practical setups. Let us recall that our focus is not to consider cases with low γ, as in such cases the easily interpretable CTE risk measure can be used and estimated with good accuracy, including at extreme levels. We shall rather consider cases with $\gamma > 1/4$, where the fact that the tail extreme tail L^p-median realises a compromise between the robust MS and the sensitive CTE should be expected to result in estimators with an improved finite-sample performance compared to that of the classical empirical CTE estimator. It was actually highlighted in (3) and (4) that, for $p \in [1,2]$ and $\gamma < 1/(p-1)$, an extreme tail L^p-median $m_p(\alpha)$ can be understood asymptotically as a weighted average of MS(α) and CTE(α). In other words, defining the interpolating risk measure

$$R_\lambda(\alpha) := \lambda \text{MS}(\alpha) + (1-\lambda)\text{CTE}(\alpha),$$

we have $m_p(\alpha) \approx R_\lambda(\alpha)$ as $\alpha \to 1^-$ with $\lambda = \lambda(p,\gamma)$ as in (4). It then turns out that at the population level, we have two distinct (but asymptotically equivalent) possibilities to interpolate, and thus create a compromise, between extreme Median Shortfall and extreme Conditional Tail Expectation:

- Consider the family of measures $m_p(\alpha)$, $p \in [1,2]$;

- Consider the family of measures $R_\lambda(\alpha)$, $\lambda \in [0,1]$.

In the rest of this section, based on a sample of data $(X_1, \ldots, X_n)$ of size n, we consider the estimation of the tail L^p-median $m_p(\alpha_n)$ for both intermediate and extreme levels α_n, and how this estimation compares with direct estimation of the interpolating measure $R_\lambda(\alpha_n)$.

5.1 Intermediate case

We first investigate the estimation of $m_p(\alpha_n)$, for $p \in [1,2]$, and of $R_\lambda(\alpha_n)$, for $\lambda \in [0,1]$, in the intermediate case when $\alpha_n \to 1^-$ and $n(1-\alpha_n) \to \infty$. As far as the estimation of $m_p(\alpha_n)$ is concerned, we compare the direct estimator $\widehat{m}_p(\alpha_n)$ defined in (6) and the indirect estimator in (8). In the latter, $\widehat{\gamma}_n$ is taken as the Hill estimator defined in (9): in other words,

$$\widetilde{m}_p(\alpha_n) = \frac{X_{n-\lfloor n(1-\alpha_n)\rfloor,n}}{\kappa(p, \widehat{\gamma}_n^H(\lfloor n(1-\alpha_n)\rfloor))}$$

$$\text{with } \widehat{\gamma}_n^H(k) = \frac{1}{k}\sum_{i=1}^{k}\log(X_{n-i+1,n}) - \log(X_{n-k,n}).$$

To further compare the performance of these two estimators, and therefore the practical applicability of the measure $m_p(\alpha_n)$ for interpolating between extreme MS and extreme CTE, the finite sample behaviour of these estimators are compared to that of the estimator of $R_\lambda(\alpha_n)$ given by

$$\widehat{R}_\lambda(\alpha_n) := \lambda X_{n-\lfloor n(1-\alpha_n)/2\rfloor,n} + (1-\lambda)\widehat{\text{CTE}}(\alpha_n)$$

$$\text{with } \widehat{\text{CTE}}(\alpha_n) = \widehat{m}_2(\alpha_n) = \frac{1}{\lfloor n(1-\alpha_n)\rfloor}\sum_{i=1}^{\lfloor n(1-\alpha_n)\rfloor} X_{n-i+1,n}.$$

In order to be able to compare this estimator of $R_\lambda(\alpha_n)$ to our estimators of the tail L^p-median $m_p(\alpha_n)$, we take $\lambda = \lambda(p,\gamma)$ as in (4). This results in $m_p(\alpha_n) \approx R_{\lambda(p,\gamma)}(\alpha_n)$ as $n \to \infty$, and we can then compare the estimators $\widehat{m}_p(\alpha_n)$, $\widetilde{m}_p(\alpha_n)$ and $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$ on the range $p \in [1,2]$.

We do so on $N = 500$ simulated random samples of size $n = 500$, with $\alpha_n = 1 - 75/n = 0.85$. Two distributions satisfying condition $\mathcal{C}_2(\gamma, \rho, A)$ are considered:

- The Burr distribution having distribution function $F(x) = 1 - (1 + x^{3/(2\gamma)})^{-2/3}$ on $(0, \infty)$ (here $\gamma > 0$). This distribution has tail index γ and second-order parameter $\rho = -3/2$.

- The Student distribution with $1/\gamma$ degrees of freedom, where $\gamma > 0$. This distribution has tail index γ and second-order parameter $\rho = -2\gamma$.

For each of these two distributions, we examine the cases $\gamma \in \{1/4, 1/2, 3/4\}$, corresponding to, respectively, a borderline case for finite fourth moment, a borderline case for finite variance, and a case where there is no finite variance. The accuracy of each of the three estimators $\widehat{m}_p(\alpha_n)$, $\widetilde{m}_p(\alpha_n)$ and $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$ is measured by their respective empirical Mean Log-Squared Error (MLSE) on the N simulated samples. This is defined by

$$\text{MLSE}(\widehat{m}_p(\alpha_n)) = \frac{1}{N} \sum_{j=1}^{N} \log^2 \left(\frac{\widehat{m}_p^{(j)}(\alpha_n)}{m_p(\alpha_n)} \right)$$

where $\widehat{m}_p^{(j)}(\alpha_n)$ is the estimate of $m_p(\alpha_n)$ calculated on the jth sample, and similarly for $\widetilde{m}_p(\alpha_n)$ and $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$ (in the latter case, we replace $m_p(\alpha_n)$ by $R_{\lambda(p,\gamma)}(\alpha_n)$). The rationale for reporting MLSEs instead of the straightforward Relative Mean-Squared Errors is that, in the case $\gamma \geq 1/2$, the direct CTE estimator $\widehat{m}_2(\alpha_n)$ occasionally produces a very large error, due to its variance being infinite (a similar point is made in Section 4.2 of El Methni & Stupfler, 2018). Computing a logarithmic error therefore helps to assess the relative performance of all the considered estimators across our cases. MLSEs of our estimators are then represented, as functions of $p \in [1, 2]$, in the top panels of Figures 1–3.

It can be seen from these graphs that, on the Burr distribution, the indirect estimator $\widetilde{m}_p(\alpha_n)$ has a lower MLSE than the direct estimator, except for γ close to 1 and p close to 2. This instability is very likely due to the fact that, for p close to 2, we have $1/\kappa(p,\gamma) \approx 1/(1-\gamma)$, and thus if γ is also close to 1, the quantity $1/\kappa(p, \widehat{\gamma}_n^H(\lfloor n(1-\alpha_n) \rfloor))$ appearing in the estimator $\widetilde{m}_p(\alpha_n)$ will be a highly unstable estimator of $1/\kappa(p,\gamma)$. By contrast, the direct estimator is generally more accurate than the indirect one when the underlying distribution is a Student distribution, especially for $p \geq 1.6$. Furthermore, it should be noted that the direct estimator $\widehat{m}_p(\alpha_n)$ performs overall noticeably better than the estimator $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$. This confirms our theoretical expectation that estimation of $m_p(\alpha_n)$ should be easier than estimation of $R_{\lambda(p,\gamma)}(\alpha_n)$, since $\widehat{m}_p(\alpha_n)$ is relatively $\sqrt{n(1-\tau_n)}-$consistent as soon as $\gamma < 1/[2(p-1)]$, which is a weaker condition than the assumption $\gamma < 1/2$ needed to ensure the relative $\sqrt{n(1-\tau_n)}-$consistency of $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$ (due to its reliance on the estimator $\widehat{\text{CTE}}(\alpha_n)$). In other words, on finite samples and at intermediate levels, it is preferable to interpolate between extreme MS and extreme CTE *via* tail L^p-medians rather than using direct linear interpolation, even though these two ideas are asymptotically equivalent at the population level.

5.2 Extreme case

We now focus on the estimation of $m_p(\beta_n)$ and $R_\lambda(\beta_n)$ for a proper extreme level $\beta_n \to 1^-$, such that $n(1-\beta_n) \to c \in [0,\infty)$. The estimators we consider are, for the estimation of the extreme tail L^p-median $m_p(\beta_n)$, the two extrapolated estimators defined in Section 4.3: first, the extrapolated direct estimator, given by

$$\widehat{m}_p^W(\beta_n, k) := \widehat{m}_p^W(\beta_n) = \left(\frac{n(1-\beta_n)}{k}\right)^{-\widehat{\gamma}_n^H(k)} \widehat{m}_p(1-k/n),$$

where $k \in \{1, \ldots, n-1\}$, and, second, the extrapolated indirect estimator

$$\widetilde{m}_p^W(\beta_n, k) := \widetilde{m}_p^W(\beta_n) = \left(\frac{n(1-\beta_n)}{k}\right)^{-\widehat{\gamma}_n^H(k)} \widetilde{m}_p(1-k/n).$$

Recall that here, $\widehat{\gamma}_n^H(k)$ denotes the Hill estimator introduced in (9). These two estimators are compared to the following extrapolated version of the estimator $\widehat{R}_\lambda$:

$$\widehat{R}_\lambda^W(\beta_n, k) := \left(\frac{n(1-\beta_n)}{k}\right)^{-\widehat{\gamma}_n^H(k)} \left[\lambda X_{n-\lfloor k/2 \rfloor, n} + (1-\lambda)\widehat{\mathrm{CTE}}(1-k/n)\right].$$

We take again $\lambda = \lambda(p, \gamma)$ so that the estimators $\widehat{m}_p^W(\beta_n, k)$, $\widetilde{m}_p^W(\beta_n, k)$ and $\widehat{R}_\lambda^W(\beta_n, k)$ can be compared on the range $p \in [1,2]$. In what follows, we also take $\beta_n = 1-1/n$ in all three estimators. These estimators depend on a tuning parameter k, which we take as

$$\hat{k}_{\mathrm{opt}} := \left\lfloor \frac{1}{J} \sum_{j=1}^J \mathop{\arg\min}_{k \in \{4, \ldots, \lfloor n/4 \rfloor\}} \int_{1-k/n}^{1-k/(4n)} \log^2\left(\frac{\widehat{m}_{p_j}^W(\alpha, k)}{\widehat{m}_{p_j}(\alpha)}\right) d\alpha \right\rfloor,$$

where $J \in \mathbb{N} \setminus \{0\}$ and $1 = p_1 < \cdots < p_J = 2$. The idea behind this criterion is that for an intermediate order α, the empirical and extrapolated estimators should both be able to estimate accurately $m_p(\alpha)$, across the range $p \in [1,2]$, and therefore the distance between these two estimators should be small at intermediate levels provided the parameter k is chosen properly. A related selection rule is used by Gardes & Stupfler (2015) for extreme quantile estimation. We take here $J = 4$ and $p_j = (j+2)/3$. Simulation settings (sample size, distributions, ...) are the same as those of the intermediate case; empirical MLSEs of the estimators $\widehat{m}_p^W(\beta_n, \hat{k}_{\mathrm{opt}})$, $\widetilde{m}_p^W(\beta_n, \hat{k}_{\mathrm{opt}})$ and $\widehat{R}_{\lambda(p,\gamma)}^W(\beta_n, \hat{k}_{\mathrm{opt}})$ are represented, as functions of $p \in [1,2]$, in the bottom panels of Figures 1–3.

These graphs show that the conclusions reached in the intermediate case remain true in the extreme case $\beta_n = 1-1/n$: on the Burr distribution, the extrapolated indirect estimator of $m_p(\beta_n)$ performs comparably or better than the extrapolated direct estimator, except for γ close to 1 and p close to 2.

The reverse conclusion holds true on the Student distribution. The extrapolated direct estimator $\widehat{m}_p^W(\beta_n, \hat{k}_{\text{opt}})$ also performs generally better than the extrapolated linear interpolation estimator $\widehat{R}_{\lambda(p,\gamma)}^W(\beta_n, \hat{k}_{\text{opt}})$. There is a noticeable improvement for $p \in [1.25, 1.75]$ in the case $\gamma = 1/2$ and even more so for $\gamma = 3/4$, which are the most relevant cases for our purpose. The accuracy of $\widehat{m}_p^W(\beta_n, \hat{k}_{\text{opt}})$ is also comparable overall to that of $\widehat{R}_{\lambda(p,\gamma)}^W(\beta_n, \hat{k}_{\text{opt}})$ for $\gamma = 1/4$, although this is not the case we originally constructed the tail L^p-median for.

As a conclusion, this simulation study indicates that, whether at intermediate or proper extreme levels, our tail L^p-median estimators provide a way of interpolating between extreme MS and extreme CTE that is more accurate in practice than a simple linear interpolation. This appears to be true on a wide range of values of γ, allowing for a flexible use of the class of tail L^p-medians in practice, although the improvement is clearer for $\gamma \geq 1/2$, where it is known that CTE estimation is a difficult problem. Finally, the fact that the extrapolated direct estimator performs better than its indirect competitor for certain distributions, and for high γ generally, is evidence that an extreme tail L^p-median actually contains more information than a simple combination of a quantile with the tail index γ. This was also observed by El Methni & Stupfler (2017), Daouia *et al.* (2018, 2019) in the context of, respectively, the estimation of extreme Wang distortion risk measures, expectiles and L^p-quantiles.

6   Real data analysis

The data set we consider is made of $n = 1098$ commercial fire losses recorded between 1st January 1995 and 31st December 1996 by the FFSA (an acronym for the *Fédération Française des Sociétés d'Assurance*). This data set is available from the R package `CASdatasets` as `data(frecomfire)`. The data is converted into euros from French francs, and denoted by $(X_1, \ldots, X_n)$. Insurance and financial companies have a strong interest in the analysis of this type of data set. For example, extremely high losses have to be taken into account in order to estimate, at company level, the capital requirements that have to be put in place so as to survive the upcoming calendar year with a probability not less than 0.995, as part of compliance with the Solvency II directive. At the same time, it is in the interest of companies to carry out a balanced assessment of risk: an underestimation of risk could threaten the company's survival, but an overestimation may lead the insurer to ask for higher premiums and deductibles, reducing the attractiveness of its products

to the consumer, thus negatively affecting the company's competitiveness on the market. A single quantile, as in the above Solvency II compliance example, cannot account for a detailed picture of risk, which is why we investigate here the use of the alternative class of tail L^p-medians.

Our first step in the analysis of the extreme losses in this data set is to estimate the tail index γ. To this end, the procedure of choice of $\hat{k}_{\mathrm{opt}}$ outlined in Section 5.2 is used: the value of k returned by this procedure is $\hat{k}_{\mathrm{opt}} = 64$, and the corresponding estimate of γ provided by the Hill estimator is $\widehat{\gamma}^H = 0.67$. This estimated value of γ is in line with the findings of El Methni & Stupfler (2018), and it suggests that there is evidence for an infinite second moment of the underlying distribution. We know, in this context, that the estimation of the extreme CTE is going to be a difficult problem, although it would give a better understanding of risk in this data set than a single quantile such as the VaR or MS would do. It therefore makes sense, on this data set, to use the class of tail L^p-medians to find a middle ground between MS and CTE estimation by interpolation.

Estimates of the tail L^p-median $m_p(1 - 1/n)$, obtained through our extrapolated direct and extrapolated indirect estimators, are depicted on Figure 4. These estimates are fairly close overall on the range $p \in [1, 2]$. There is a difference for p close to 2, where the increased theoretical sensitivity of the direct estimator to the highest values in the sample makes it exceed the more robust indirect estimator; note that here γ is estimated to be 0.67, which is sufficiently far from 1 to ensure that the instability of the indirect estimator for very high γ is not an issue, and the estimate returned by the indirect extrapolated method can be considered to be a reasonable one.

To get a further idea of the proximity between the two estimators, we calculate asymptotic confidence intervals, on the basis of the convergence

$$\frac{\sigma_n^{-1}}{\gamma} \left(\frac{m_p(1 - 1/n)}{\widehat{m}_p^W(1 - 1/n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, 1), \text{ where } \sigma_n = \frac{\log[n(1 - \alpha_n)]}{\sqrt{n(1 - \alpha_n)}}.$$

This results from a straightforward combination of Theorem 3 and the delta-method, recalling that γ is estimated by the Hill estimator $\widehat{\gamma}^H$ which is asymptotically Gaussian with variance γ^2. Letting u_ξ be the $\xi-$quantile of the standard Gaussian distribution, and $\alpha_n = 1 - \hat{k}_{\mathrm{opt}}/n$, a $(1 - \xi)-$asymptotic confidence interval for $m_p(1 - 1/n)$ is then

$$\left[\widehat{m}_p^W(1 - 1/n, \hat{k}_{\mathrm{opt}}) \left(1 + \widehat{\gamma}^H \hat{\sigma} u_{\xi/2} \right), \ \widehat{m}_p^W(1 - 1/n, \hat{k}_{\mathrm{opt}}) \left(1 + \widehat{\gamma}^H \hat{\sigma} u_{1-\xi/2} \right) \right], \text{ with } \hat{\sigma} = \frac{\log \hat{k}_{\mathrm{opt}}}{\sqrt{\hat{k}_{\mathrm{opt}}}}.$$

An analogue confidence interval is available using $\widetilde{m}_p^W(1 - 1/n, \hat{k}_{\mathrm{opt}})$ rather than $\widehat{m}_p^W(1 - 1/n, \hat{k}_{\mathrm{opt}})$. On Figure 5, we report asymptotic 90% confidence intervals, corresponding to $\xi = 0.1$, constructed

using both estimators. The confidence intervals largely overlap, which was to be expected since the two estimators $\widehat{m}_p^W(\beta_n, \hat{k}_{\mathrm{opt}})$ and $\widetilde{m}_p^W(\beta_n, \hat{k}_{\mathrm{opt}})$ were seen to be fairly close on Figure 4. The confidence intervals are also quite wide, which was again to be expected because the effective sample size is the relatively low $\hat{k}_{\mathrm{opt}} = 64$.

With these estimates at hand comes the question of how to choose p. This is of course a difficult question that depends on the objective of the analysis from the perspective of risk assessment: should the analysis be conservative (*i.e.* return a higher, more prudent estimation) or not? We do not wish to enter into such a debate here, which would be better held within the financial and actuarial communities. Let us, however, illustrate a simple way to choose p based on our discussion at the very end of Section 3. Recall from (3) and (4) that a tail L^p-median has an asymptotic interpretation as a weighted average of MS and CTE:

$$
\begin{aligned}
m_p(1 - 1/n) &\approx \lambda(p, \gamma)\mathrm{MS}(1 - 1/n) + (1 - \lambda(p, \gamma))\mathrm{CTE}(1 - 1/n) \\
\text{with } \lambda(p, \gamma) &= \frac{1 - (1 - \gamma)/\kappa(p, \gamma)}{1 - 2\gamma(1 - \gamma)}.
\end{aligned}
$$

We also know from our simulation study that it is generally more accurate to estimate $m_p(1 - 1/n)$ rather than the corresponding linear combination of $\mathrm{MS}(1 - 1/n)$ and $\mathrm{CTE}(1 - 1/n)$. With $\lambda_0 = 1/2$, representing the simple average between $\mathrm{MS}(1 - 1/n)$ and $\mathrm{CTE}(1 - 1/n)$, choosing $p = \widehat{p}$ as the unique root of the equation $\lambda(\widehat{p}, \widehat{\gamma}) = \lambda_0$ yields $\widehat{p} = 1.711$. The corresponding estimates, in million euros, of the linear combination

$$
\lambda_0\mathrm{MS}(1 - 1/n) + (1 - \lambda_0)\mathrm{CTE}(1 - 1/n)
$$

are $\widehat{m}_{\widehat{p}}^W(1 - 1/n) = 160.8$ and $\widetilde{m}_{\widehat{p}}^W(1 - 1/n) = 155.4$. It is interesting to note that, although these quantities estimate the average of $\mathrm{MS}(1 - 1/n) = m_1(1 - 1/n)$ and $\mathrm{CTE}(1 - 1/n) = m_2(1 - 1/n)$, we also have $\widehat{m}_1^W(1 - 1/n) = 106.3$ and $\widehat{m}_2^W(1 - 1/n) = 225.2$ so that

$$
\widetilde{m}_{\widehat{p}}^W(1 - 1/n) = 155.4 < \widehat{m}_{\widehat{p}}^W(1 - 1/n) = 160.8 < \frac{1}{2}\left[\widehat{m}_1^W(1 - 1/n) + \widehat{m}_2^W(1 - 1/n)\right] = 165.8.
$$

The estimate $\widehat{m}_{\widehat{p}}^W(1 - 1/n)$ of the simple average between $\mathrm{MS}(1 - 1/n)$ and $\mathrm{CTE}(1 - 1/n)$ obtained through extrapolating the direct tail L^p-median estimator is therefore itself a middleway between the indirect estimator $\widetilde{m}_{\widehat{p}}^W(1 - 1/n)$, which relies on the VaR estimator $\widehat{q}(1 - 1/n)$, and the direct estimator of this average, which depends on the highly variable estimator $\widehat{\mathrm{CTE}}(1 - 1/n)$.

Acknowledgements

Supplementary Material

A Supplementary Material document, available online, contains all necessary proofs as well as additional figures referred to in the present article.

References

Acerbi, C. & Tasche, D. (2002). On the coherence of expected shortfall. *J. Banking & Finance* **26**, 1487–1503.

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

Basel Committee on Banking Supervision (2013). *Fundamental review of the trading book: A revised market risk framework*.

Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of extremes: theory and applications*. John Wiley & Sons, Chichester.

Bingham, N.H., Goldie, C.M. & Teugels, J.L. (1987). *Regular variation*. Cambridge University Press.

Borisov, I.S. & Baklanov, E.A. (2001). Probability inequalities for generalized $L-$statistics. *Sib. Math. J.* **42**, 217–231.

Brazauskas, V., Jones, B.L., Puri, M.L. & Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view. *J. Statist. Plann. Inference* **138**, 3590–3604.

Cai, J.-J., Einmahl, J.H.J., de Haan, L. & Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *J. Roy. Statist. Soc. Ser. B* **77**, 417–442.

Chen, Z. (1996). Conditional L_p-quantiles and their application to testing of symmetry in non-parametric regression. *Statist. Probab. Lett.* **29**, 107–115.

Daouia, A., Girard, S. & Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *J. Roy. Statist. Soc. Ser. B* **80**, 263–292.

Daouia, A., Girard, S. & Stupfler, G. (2019). Extreme M-quantiles as risk measures: From L^1 to L^p optimization. *Bernoulli* **25**, 264–309.

Daníelsson, J., Jorgensen, B.N., Samorodnitsky, G., Sarma, M. & de Vries, C.G. (2013). Fat tails, VaR and subadditivity. *J. Econometrics* **172**, 283–291.

de Haan, L. & Ferreira, A. (2006). *Extreme value theory: an introduction*, Springer.

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

El Methni, J., Gardes, L. & Girard, S. (2018). Kernel estimation of extreme regression risk measures. *Electron. J. Stat.* **12**, 359–398.

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

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

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*, Springer.

Embrechts, P., Lambrigger, D.P. & Wüthrich, M.V. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes* **12**, 107–127.

Emmer, S., Kratz, M. & Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *J. Risk* **18**, 31–60.

Gardes, L. & Stupfler, G. (2015). Estimating extreme quantiles under random truncation. *TEST* **24**, 207–227.

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

Gneiting, T. (2011). Making and evaluating point forecasts. *J. Am. Statist. Assoc.* **106**, 746–762.

Gong, J., Li, Y., Peng, L. & Yao, Q. (2015). Estimation of extreme quantiles for functions of dependent random variables. *J. Roy. Statist. Soc. Ser. B* **77**, 1001–1024.

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

Hill, J.B. (2013). Least tail-trimmed squares for infinite variance autoregressions. *J. Time Series Anal.* **34**, 168–186.

Hill, J.B. (2015). Expected shortfall estimation and Gaussian inference for infinite variance time series. *J. Financ. Econom.* **13**, 1–44.

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

International Monetary Fund (2014). *Canada: financial sector assessment program–insurance core principles–detailed assessment of observance*. Washington, D.C.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755–770.

Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. Technical report, University of Toronto. Available at `http://www.utstat.toronto.edu/keith/EPFL/equi.ps`.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press.

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

Kou, S., Peng, X. & Heyde, C.C. (2013). External risk measures and Basel accords. *Math. Oper. Res.* **38**, 393–417.

Kou, S. & Peng, X. (2016). On the measurement of economic tail risk. *Oper. Res.* **64**, 1056–1072.

Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J. Roy. Statist. Soc. Ser. B* **67**, 381–393.

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

Organisation for Economic Co-operation and Development (2016). *Life annuity products and their guarantees.* OECD Publishing, Paris.

Resnick, S. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer, New York.

Rockafellar, R.T. & Uryasev, S. (2002). Conditional Value-at-Risk for general loss distributions. *J. Banking & Finance* **26**, 1443–1471.

Tasche, D. (2002). Expected shortfall and beyond. *J. Banking & Finance* **26**, 1519–1533.

Tasche, D. (2008). Capital allocation to business units and sub-portfolios: the Euler principle. In *Pillar II in the new Basel accord: the challenge of economic capital* (ed A. Resti), Risk Books, 423–453.

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

Wüthrich, M. & Merz, M. (2013). *Financial modeling, actuarial valuation and solvency in insurance.* Springer-Verlag, Berlin Heidelberg.

Yang, F. (2015). First- and second-order asymptotics for the tail distortion risk measure of extreme risks. *Commun. Stat. Theory Methods* **44**, 520–532.

Zhu, L. & Li, H. (2012). Tail distortion risk and its asymptotic analysis. *Insurance Math. Econom.* **51**, 115–121.

Details of corresponding author:

Gilles Stupfler

School of Mathematical Sciences, University of Nottingham

University Park Campus, Nottingham NG7 2RD, United Kingdom
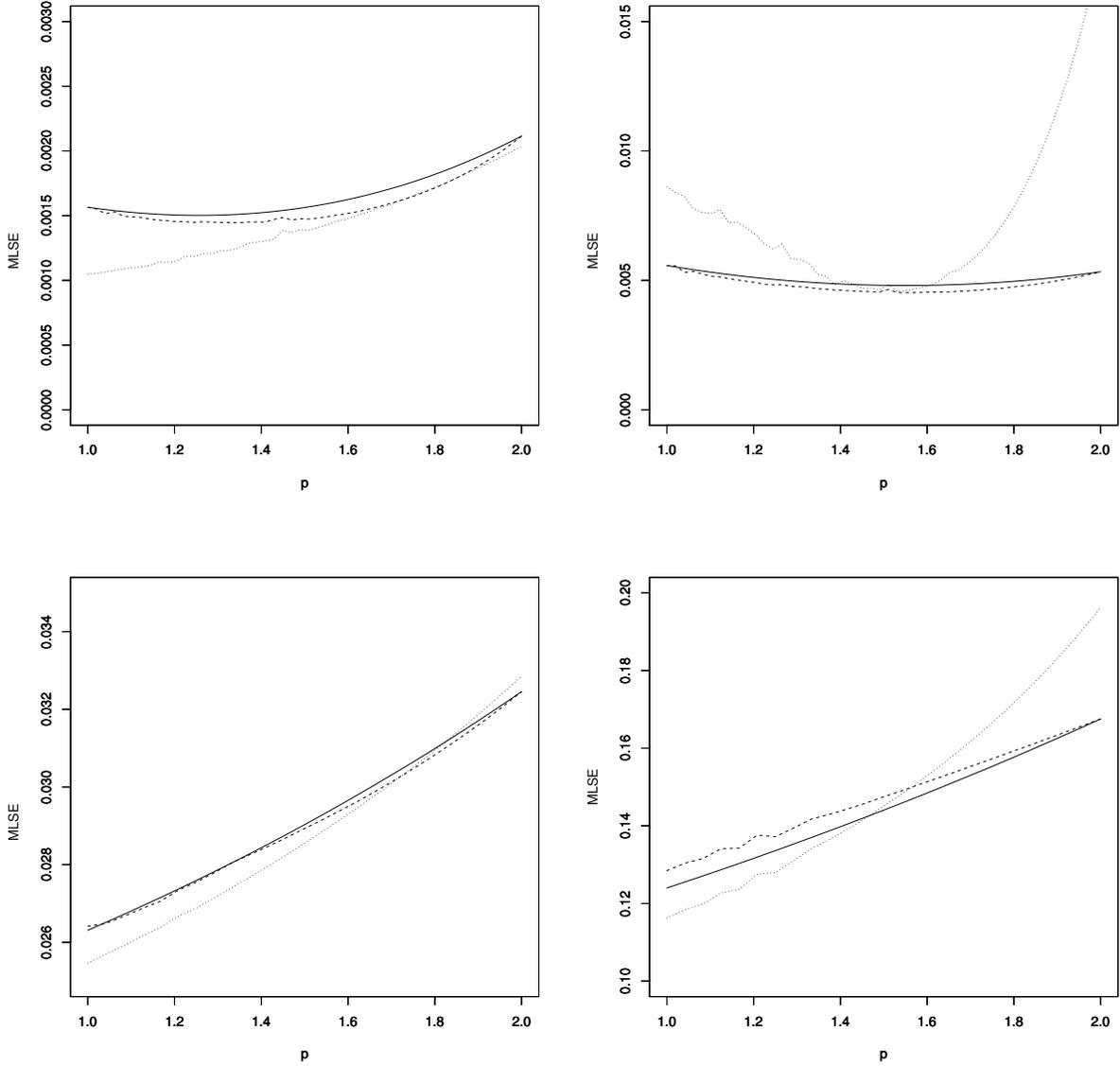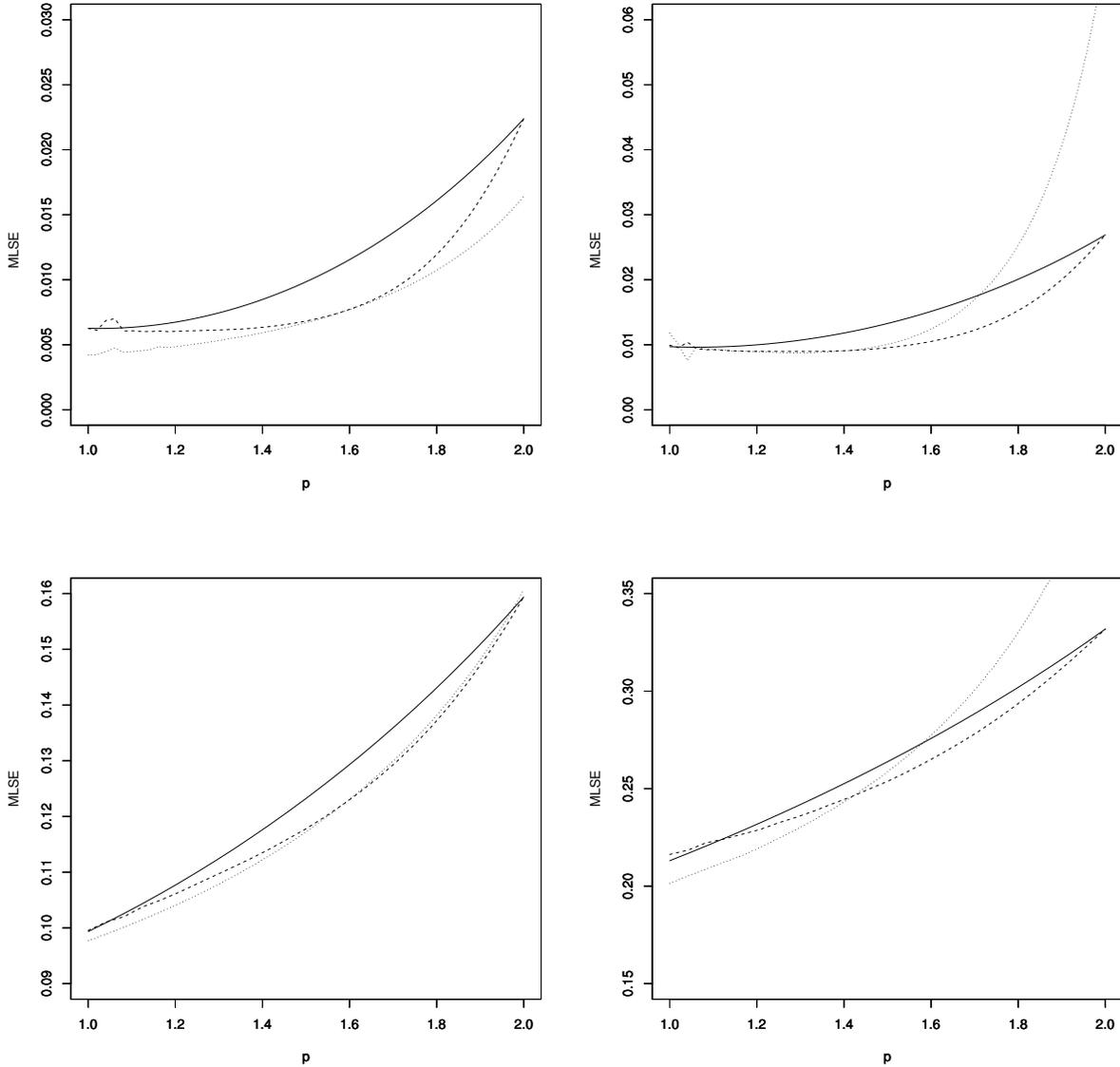
Email: gilles.stupfler@gmail.com

Figure 1: Simulation study, case $\gamma = 1/4$. **Top panels:** empirical MLSE of the estimator $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$ estimator (full line) and of the two tail L^p-median estimators $\widehat{m}_p(\alpha_n)$ (dashed line) and $\widetilde{m}_p(\alpha_n)$ (dotted line). **Bottom panels:** empirical MLSE of the extrapolated estimator $\widehat{R}^W_{\lambda(p,\gamma)}(\beta_n, \hat{k}_{\mathrm{opt}})$ (full line) and of the two extrapolated tail L^p-median estimators $\widehat{m}^W_p(\beta_n, \hat{k}_{\mathrm{opt}})$ (dashed line) and $\widetilde{m}^W_p(\beta_n, \hat{k}_{\mathrm{opt}})$ (dotted line). Left: Burr distribution, right: Student distribution.
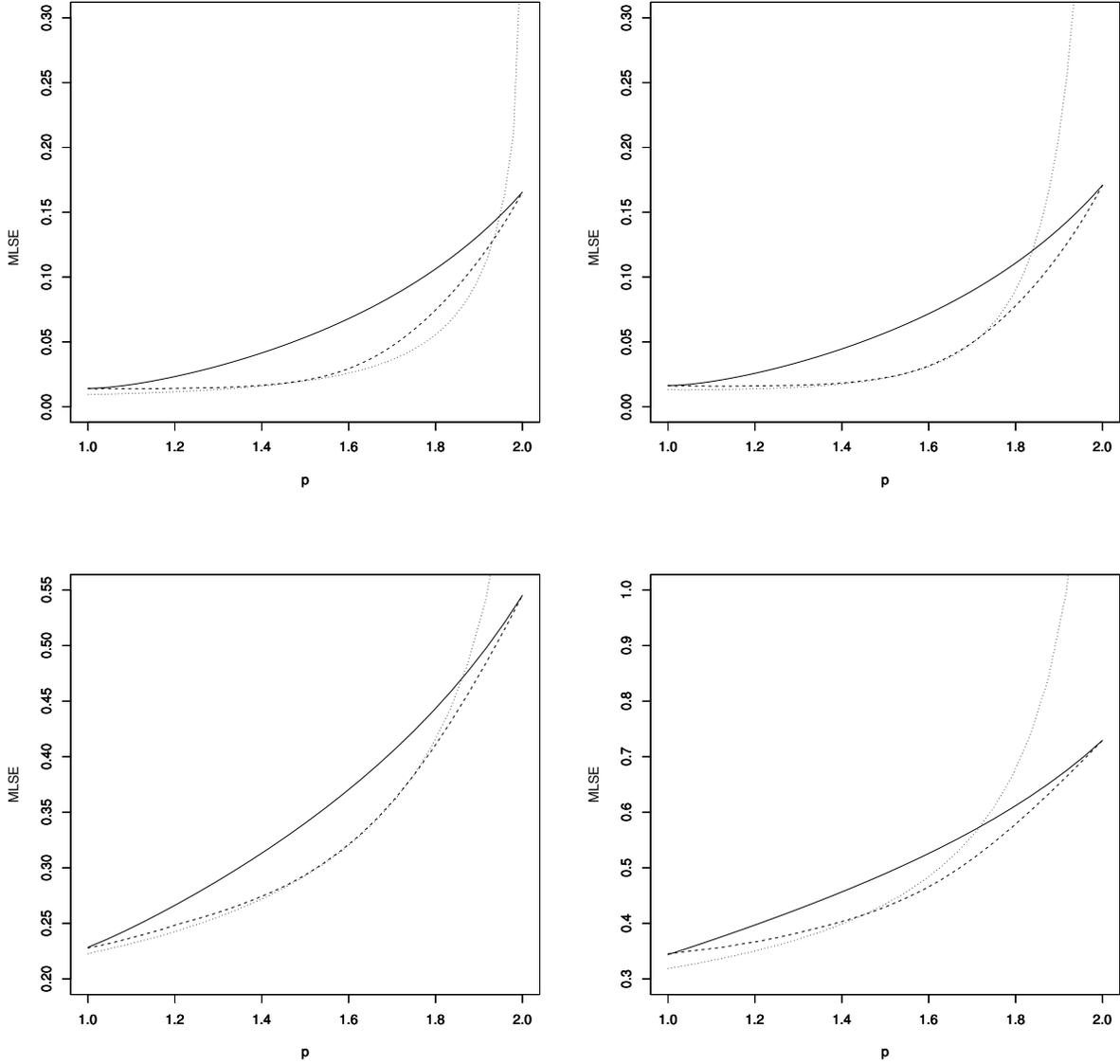
Figure 2: Simulation study, case $\gamma = 1/2$. **Top panels:** empirical MLSE of the estimator $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$ estimator (full line) and of the two tail L^p−median estimators $\widehat{m}_p(\alpha_n)$ (dashed line) and $\widetilde{m}_p(\alpha_n)$ (dotted line). **Bottom panels:** empirical MLSE of the extrapolated estimator $\widehat{R}^W_{\lambda(p,\gamma)}(\beta_n, \hat{k}_{\mathrm{opt}})$ (full line) and of the two extrapolated tail L^p−median estimators $\widehat{m}^W_p(\beta_n, \hat{k}_{\mathrm{opt}})$ (dashed line) and $\widetilde{m}^W_p(\beta_n, \hat{k}_{\mathrm{opt}})$ (dotted line). Left: Burr distribution, right: Student distribution.

Figure 3: Simulation study, case $\gamma = 3/4$. **Top panels:** empirical MLSE of the estimator $\widehat{R}_{\lambda(p,\gamma)}(\alpha_n)$ estimator (full line) and of the two tail L^p−median estimators $\widehat{m}_p(\alpha_n)$ (dashed line) and $\widetilde{m}_p(\alpha_n)$ (dotted line). **Bottom panels:** empirical MLSE of the extrapolated estimator $\widehat{R}^W_{\lambda(p,\gamma)}(\beta_n, \hat{k}_{\mathrm{opt}})$ (full line) and of the two extrapolated tail L^p−median estimators $\widehat{m}^W_p(\beta_n, \hat{k}_{\mathrm{opt}})$ (dashed line) and $\widetilde{m}^W_p(\beta_n, \hat{k}_{\mathrm{opt}})$ (dotted line). Left: Burr distribution, right: Student distribution.
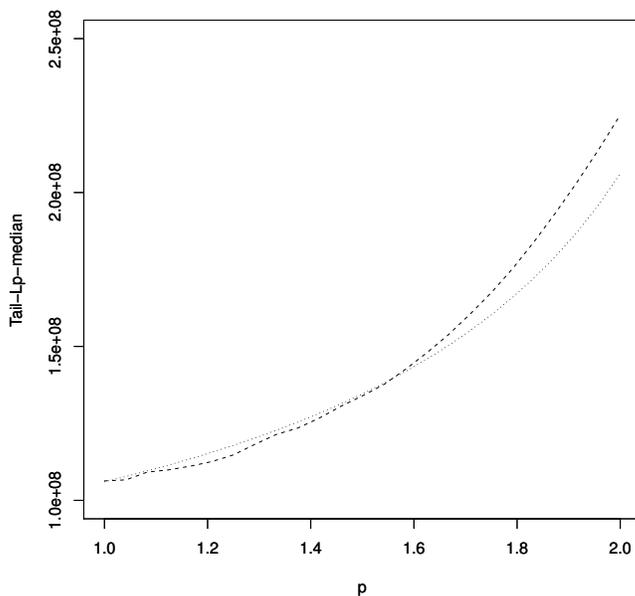
Figure 4: Real fire insurance data set. Extrapolated tail L^p−median estimators $\widehat{m}_p^W(1-1/n, \hat{k}_{\mathrm{opt}})$ (dashed line) and $\widetilde{m}_p^W(1-1/n, \hat{k}_{\mathrm{opt}})$ (dotted line) as functions of $p \in [1, 2]$.
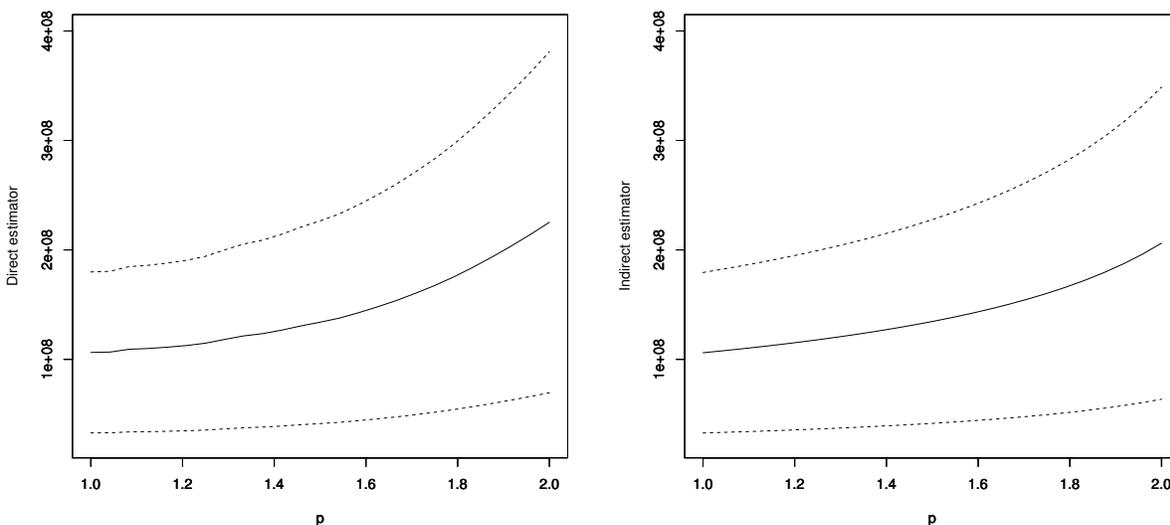

Figure 5: Real fire insurance data set. Extrapolated tail L^p−median estimators $\widehat{m}_p^W(1-1/n, \hat{k}_{\mathrm{opt}})$ (left panel) and $\widetilde{m}_p^W(1-1/n, \hat{k}_{\mathrm{opt}})$ (right panel) as functions of $p \in [1, 2]$. The lower and upper bounds of the asymptotic confidence intervals at the 90% level are represented by the dashed lines.