Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions

Jonathan El Methni, Gilles Stupfler

Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions

Jonathan El Methni(1) & Gilles Stupfler(2,*)

(1) Université Paris Descartes, Sorbonne Paris Cité, Laboratoire MAP5, UMR CNRS 8145, 75006 Paris, France

(2) Aix Marseille Université, CNRS, EHESS, Centrale Marseille, GREQAM UMR 7316, 13002 Marseille, France

Abstract. Among the many possible ways to study the right tail of a real-valued random variable, a particularly general one is given by considering the family of its Wang distortion risk measures. This class of risk measures encompasses various interesting indicators, such as the widely used Value-at-Risk and Tail Value-at-Risk, which are especially popular in actuarial science, for instance. In this paper, we first build simple extreme analogues of Wang distortion risk measures and we show how this makes it possible to consider many standard measures of extreme risk, including the usual extreme Value-at-Risk or Tail-Value-at-Risk, as well as the recently introduced extreme Conditional Tail Moment, in a unified framework. We then introduce adapted estimators when the random variable of interest has a heavy-tailed distribution and we prove their asymptotic normality. The finite sample performance of our estimators is assessed on a simulation study and we showcase our techniques on two sets of real data.

AMS Subject Classifications: 62G05, 62G30, 62G30, 62G32.

Keywords: asymptotic normality, conditional tail moment, distortion risk measure, extreme-value statistics, heavy-tailed distribution.

1 Introduction

Understanding the extremes of a random phenomenon is a major question in various areas of statistical application. The first motivating problem for extreme value theory is arguably to determine how high the dykes surrounding the areas below sea level in the Netherlands should be so as to protect these zones from flood risk in case of extreme storms affecting Northern Europe, see de Haan and Ferreira (2006). Further climate-related examples are the estimation of extreme rainfall at a given location (Koutsoyiannis, 2004), the estimation of extreme daily wind speeds (Beirlant *et al.*, 1996) or the modeling of large forest fires (Alvarado *et al.*, 1998). Another stimulating topic comes from the fact that extreme phenomena may have strong adverse effects on financial institutions or insurance companies, and the investigation of those effects on financial returns makes up a large part of the recent extreme value literature. Examples of such studies include the analysis of extreme log-returns of financial time series (Drees, 2003) or the study of extreme risks related to large losses for an insurance company (Rootzén and Tajvidi, 1997). A further application in actuarial science is, for insurance companies operating in Europe, the computation of their own solvency capital so as to fulfill the European Union Solvency II directive requirement that an insurance company should be able to survive the upcoming calendar year with a probability not less than 0.995.

A commonly encountered problem when analyzing the extremes of a random variable is that the straightforward empirical estimator of the quantile function is not consistent at extreme levels, that is, when the true quantile at the chosen level exceeds the range covered by the available data, and this makes beyond-the-sample estimation impossible. In many of the aforementioned applications, this issue can actually be bypassed because the problem can be accurately modeled using univariate heavy-tailed distributions. Roughly speaking, a distribution is said to be heavy-tailed if and only if its related survival function decays like a power function with negative exponent at infinity; its so-called tail index is then the parameter which controls its rate of convergence to 0 at infinity. A heuristic consequence of this is, if q denotes the underlying quantile function:

$$q(\delta) \approx \left(\frac{1-\beta}{1-\delta} \right)^{\gamma} q(\beta)$$

when β, δ are close to 1 and γ is the tail index of the distribution. The quantile function at an arbitrarily high extreme level can then be consistently deduced from its value at a typically much smaller level provided γ can be consistently estimated; the estimation of the tail index γ, an excellent overview of which is given in the recent monograph by de Haan and Ferreira (2006), is therefore a crucial step to gain understanding of the extremes of a random variable whose distribution is heavy-tailed. In concrete terms, on an $n-$sample of data, after a consistent estimator $\widehat{\gamma}$ of γ has been computed, the quantile function is first consistently estimated by the empirical quantile function at an intermediate level β_n, *i.e.* such that $n(1-\beta_n) \to \infty$, and then estimated at an arbitrarily high extreme level δ_n by plugging the two aforementioned estimators in the right-hand side of the above relationship warranted by the heavy-tailed framework. This procedure, suggested by Weissman

(1978), is arguably the simplest and most popular device as far as extreme quantile estimation is concerned.

Of course, the estimation of a single extreme quantile, or Value-at-Risk (VaR) as it is known in the actuarial and financial literature, only gives incomplete information on the extremes of a random variable. To put it differently, it may well be the case that a light-tailed distribution (*e.g.* a Gaussian distribution) and a heavy-tailed distribution share a quantile at some common level, although they clearly do not have the same behavior in their extremes. Besides, the VaR is not a coherent risk measure in the sense of Artzner *et al.* (1999), which is an undesirable feature from the financial point of view. This is why other quantities, which take into account the whole right tail of the random variable of interest, were developed and studied. Examples of such indicators include the Tail Value-at-Risk (TVaR), also called Expected Shortfall, and the Stop-loss Premium for reinsurance problems; we refer to Embrechts *et al.* (1997) and McNeil *et al.* (2005) for the study of such risk measures in an actuarial or financial context. When the related survival function is continuous, these measures can be obtained by combining the VaR and a Conditional Tail Moment (CTM) as introduced by El Methni *et al.* (2014), which is a general notion of moment of a random variable in its right tail.

One may then wonder if such risk measures may be encompassed in a single, unified class. An answer, in our opinion, lies in considering Wang distortion risk measures (DRMs), introduced by Wang (1996). The aforementioned VaR, TVaR and CTM actually are particular cases of Wang DRMs, and so are many other interesting risk measures such as the Wang transform (Wang, 2000) which is very popular in finance, the tail standard deviation premium calculation principle (Furman and Landsman, 2006) and the newly introduced GlueVaR of Belles-Sampera *et al.* (2014). The flexibility of this class is a reason why it has received considerable attention recently, see *e.g.* Wirch and Hardy (1999, 2002), Cotter and Dowd (2006) who worked with the particular subclass of spectral risk measures and Sereda *et al.* (2010), among others. The focus of our paper is to show that Wang DRMs can be nicely extended to the study of extreme risk. To be specific, we show how a simple linear transformation allows one to construct an extreme analogue of a Wang DRM and we consider its estimation under classical conditions in extreme value theory; because our estimators are suitable linear functionals of the tail quantile process, our extreme versions of Wang DRMs can be and are estimated here using the Weissman argument outlined above. Our method, it appears, provides a unified framework for the study of many frequently used extreme risk metrics, and we shall underline in particular that the asymptotic properties of our estimators make it possible to recover several results that have been known for some time in the literature.

To the best of our knowledge, the only comparable study, albeit with different goals, is Vandewalle and Beirlant (2006), and we shall highlight the differences between our work and their approach.

The outline of our paper is as follows. We first recall the definition of a Wang DRM in Section 2. In Section 3, we present a simple way to build extreme analogues of Wang DRMs and we consider their estimation. Section 4 is devoted to the study of the finite-sample performance of our estimators, and we showcase our method on real data sets in Section 5. Section 6 concludes the paper with a discussion of our results. The proofs are deferred to the Appendix.

2 Wang risk measures

Let X be a positive random variable. Wang (1996) introduced a family of risk measures called distortion risk measures (DRMs) by the concept of a distortion function: a function $g : [0,1] \to [0,1]$ is a distortion function if it is nondecreasing with $g(0) = 0$ and $g(1) = 1$. For ease of exposition, distortion functions will also be assumed to be right-continuous, a very mild condition which holds in all usual examples. The Wang DRM of X with distortion function g is then defined by:

$$R_g(X) := \int_0^\infty g(1 - F(x))dx$$

where F is the cumulative distribution function (cdf) of X. An alternative, easily interpretable expression of $R_g(X)$ can actually be found. Denote by q the quantile function of X, namely $q(\alpha) = \inf\{x \geq 0 \,|\, F(x) \geq \alpha\}$ for all $\alpha \in (0,1)$. In other words, the function q is the left-continuous inverse of F. Let moreover $m = \inf\{\alpha \in [0,1] \,|\, g(\alpha) > 0\}$ and $M = \sup\{\alpha \in [0,1] \,|\, g(\alpha) < 1\}$. Assume for the moment that q is continuous on $U \cap (0,1)$ with U an open interval containing $[1 - M, 1 - m]$; an equivalent assumption is that F is strictly increasing on $V \cap (0, \infty)$ with V an open interval containing $[q(1 - M), q(1 - m)]$. Noticing that $F(x) = \inf\{\alpha \in (0,1) \,|\, q(\alpha) > x\}$ and thus F is the right-continuous inverse of q, a classical change-of-variables formula and an integration by parts then entail that $R_g(X)$, provided it is finite, can be written as a Lebesgue-Stieltjes integral:

$$R_g(X) = \int_0^1 g(\alpha)dq(1 - \alpha) = \int_0^1 q(1 - \alpha)dg(\alpha).$$

A Wang DRM can thus be understood as a weighted version of the expectation of the random variable X. Specific examples include:

- the quantile at level β or $\mathrm{VaR}(\beta)$, standing for the level exceeded on average in $100(1 - \beta)\%$ of cases, obtained by setting $g(x) = \mathbb{I}\{x \geq 1 - \beta\}$, with $\mathbb{I}\{\cdot\}$ denoting the indicator function;
- the Tail Value-at-Risk $\mathrm{TVaR}(\beta)$ in the worst $100(1 - \beta)\%$ of cases, namely the average of all quantiles exceeding $\mathrm{VaR}(\beta)$, is recovered by taking $g(x) = \min(x/(1 - \beta), 1)$.

In Table 1 we give further examples of classical DRMs and their distortion functions (see *e.g.* Wang, 1996, Wirch and Hardy, 1999, Wang, 2000, Cherny and Madan, 2009 and Guegan and Hassani, 2014). Broadly speaking, the class of Wang DRMs allows almost total flexibility as far as the weight function is considered: in particular, choosing a convex (resp. concave) continuously differentiable function g results in gradually putting more weight towards small (resp. high) quantiles of X. Besides, any spectral risk measure of X, namely

$$S_\psi(X) = \int_0^1 q(1-\alpha)\psi(\alpha)d\alpha$$

where ψ is a non-decreasing probability density function on $[0,1]$, is also a Wang DRM with the distortion function g being the antiderivative of ψ. An application of such risk measures is considered in Cotter and Dowd (2006).

Furthermore, we note that if $h : [0,\infty) \to [0,\infty)$ is a strictly increasing, continuously differentiable function then the Wang DRM of $h(X)$ with distortion function g is

$$R_g(h(X)) = \int_0^1 h \circ q(1-\alpha)dg(\alpha). \tag{1}$$

Of course, the choice $h(x) = x$ yields standard Wang DRMs of X, but we may recover other types of risk measures by changing the function h. For instance, the choices $g(x) = \min(x/(1-\beta), 1)$, $\beta \in (0,1)$ and $h(x) = x^a$, with a a positive real number, yield after integrating by parts:

$$R_g(X^a) = \mathrm{CTM}_a(\beta) := \mathbb{E}(X^a | X > q(\beta))$$

provided F is continuous. This is actually the Conditional Tail Moment (CTM) of order a of the random variable X as introduced in El Methni *et al.* (2014). Especially, when F is continuous, the TVaR coincides with the Conditional Tail Expectation of X. Table 2 gives several examples of risk measures, such as the Conditional Value-at-Risk, Conditional Tail Variance, or Stop-loss Premium (SP) which can then be obtained by combining a finite number of CTMs and the VaR; see Furman and Landsman (2006) and El Methni *et al.* (2014) for further details.

We close this section by mentioning that in an actuarial context, a DRM is a coherent risk measure (see Artzner *et al.*, 1999), that is, translation invariant, positive homogeneous, monotonic and subadditive, if and only if the distortion function g is concave, according to Wirch and Hardy (2002). Coherency of a risk measure is often thought of as a desirable feature from the actuarial point of view; in particular, it reflects on the diversification principle which asserts that aggregating two risks cannot be worse than handling them separately (Artzner *et al.*, 1999). A particular corollary of the result of Wirch and Hardy (2002) is that while the VaR is not a coherent risk measure, the TVaR is, for instance, and this has already been noted several times in the recent literature. It should be acknowledged nonetheless that the VaR is subadditive (and therefore coherent) in the right tail under certain conditions, see Daníelsson *et al.* (2013). More broadly, the result of Wirch and Hardy (2002) makes it easy to identify the subclass of coherent Wang DRMs such as the Dual Power or Proportional Hazard transform risk measure, whose respective practical interpretations can be found in the actuarial science literature (see *e.g.* Denuit *et al.*, 2005, pp. 94–95).

The discussion about the relative merits of VaR, distortion risk measures and TVaR is not limited to coherency: although the popular saying in insurance and finance is that TVaR is more conservative than VaR, Kou and Peng (2014) argue that TVaR should actually be compared to the median shortfall (see Kou *et al.*, 2013), just as a mean is usually compared to a median, and in this case the aforementioned conclusion is no longer necessarily true. Cont *et al.* (2010) show that VaR is more robust than TVaR against small departures from the model or from the data, although it might be less aggregation-robust, see Embrechts *et al.* (2014). Linton and Xiao (2013) argue that the inference procedure for the extreme VaR is easier than from the extreme TVaR because it does not depend on tail heaviness (at least theoretically, for heavy-tailed data). There are concerns related to the actual practical use of the VaR: for instance, in the Basel II and III accords (see Basel Committee on Banking Supervision, 2006, 2011) the VaR-based risk measure used to compute capital requirements for trading books, whose relationship to the 99.9% VaR is studied in Gordy (2003), has been criticized for being procyclical (see Adrian and Brunnermeier, 2008), or, as Kou and Peng (2014) point out, for being low in booms and high in crises, which is of course a problem as far as regulation is concerned. Keppo *et al.* (2010) even show that the Basel accords capital requirements may sometimes increase the default probability of a bank, contrary to the regulators' original aim.

Let us finally mention that yet another property of risk measures, namely elicitability (Gneiting, 2011; Ziegel, 2015), has gained prominence in recent years since it has been argued to allow for correct forecast performance comparisons. A related concept is consistency, introduced by Davis (2013). While the VaR is an elicitable (and consistent) risk measure, the TVaR is not; more generally, it has been shown recently by Kou and Peng (2014) and Wang and Ziegel (2015) that Wang DRMs different from either the VaR or the simple expectation do not satisfy such a property. An example of a risk measure that is both coherent and elicitable is the expectile (Newey and Powell, 1987; in a financial context, Kuan *et al.*, 2009) when it is larger than the expectation. Studying the estimation of extreme expectiles, which to the best of our knowledge cannot be written as a simple combination of extreme Wang DRMs of X, is beyond the scope of this paper.

3 Framework

3.1 Extreme versions of Wang risk measures and their estimation

Extreme versions of Wang risk measures may be obtained as follows. Let g be a distortion function and for every $\beta \in (0,1)$, consider the function g_β which is defined by:

$$\forall y \in [0,1],\ g_\beta(y) := g\left(\min\left[1, \frac{y}{1-\beta}\right]\right) = \begin{cases} g\left(\dfrac{y}{1-\beta}\right) & \text{if} \quad y \leq 1-\beta \\ 1 & \text{otherwise.} \end{cases}$$

Such a function, which is deduced from g by a simple piecewise linear transform of its argument, is thus constant equal to 1 on $[1-\beta, 1]$. Moreover, if g is concave then so is g_β: in other words, if g gives rise to a coherent Wang DRM, so does g_β. We now consider the Wang DRM of X with distortion function g_β:

$$R_{g,\beta}(X) := \int_0^\infty g_\beta(1-F(x))dx.$$

Because the inequality $F(x) \geq \beta$ is equivalent to $x \geq q(\beta)$, it is actually straightforward to obtain from the definition of g_β that

$$R_{g,\beta}(X) = \int_0^\infty g(1-F_\beta(x))dx \quad \text{with} \quad F_\beta(x) := \max\left[0, \frac{F(x)-\beta}{1-\beta}\right]. \tag{2}$$

In the case when q is continuous and strictly increasing in a neighborhood of β, then $\beta = F(q(\beta))$ and

$$F_\beta(x) = \max\left[0, \frac{F(x) - F(q(\beta))}{1-q(\beta)}\right] = \mathbb{P}(X \leq x | X > q(\beta))$$

which makes the interpretation of the risk measure $R_{g,\beta}(X)$ clear: it is the Wang DRM of X given that it lies above the level $q(\beta)$. In other words, we have shown the following result:

Proposition 1. *Assume that for some $t > 0$, the function q is continuous and strictly increasing on $[t, 1)$. Then for all $\beta > t$ and any strictly increasing and continuously differentiable function h on $(0, \infty)$, it holds that:*

$$R_{g,\beta}(h(X)) = R_g(h(X_\beta)) \quad \textit{with} \quad \mathbb{P}(X_\beta \leq x) = \mathbb{P}(X \leq x | X > q(\beta)).$$

When $\beta \uparrow 1$, we may then think of this construction as a way to consider Wang DRMs of the extremes of X.

Choosing $h(x) = x$ makes it possible to recover some simple and widely used extreme risk measures:

- the usual extreme VaR is obtained by setting $g(x) = \mathbb{I}\{x = 1\}$,
- an extreme version of the TVaR is obtained by taking $g(x) = x$,

and the same idea yields extreme analogues of the various risk measures shown in Table 1. Furthermore, as highlighted in Section 1, choosing $g(x) = x$ and $h(x) = x^a$, $a > 0$, yields an extreme version of a CTM of X, and therefore extreme versions of quantities such as those introduced in Table 2 can be studied.

It is worth noting at this point that the construction presented in this paper is different from that of Vandewalle and Beirlant (2006). In the latter paper, the authors look at the asymptotic behavior of the quantity

$$\int_R^\infty g(1 - F(x))dx = \int_0^\infty g(1 - F(x + R))dx$$

as $R \to \infty$. This amounts to considering the Wang DRM R_g of $(X - R)\mathbb{I}\{X > R\} = \max(X - R, 0)$ for large R. Their construction is thus adapted to the examination of excess-of-loss reinsurance policies for extreme losses, a prominent example of risk premium being then obtained by the Stop-loss Premium; their work is, by the way, restricted to the case of a concave function g satisfying a regular variation condition in a neighborhood of 0. It therefore excludes the simple VaR risk measure, for instance, as well as the Conditional-Value-at-Risk (CVaR) and the GlueVaR of Belles-Sampera *et al.* (2014). Our idea is rather to consider a conditional construction in the sense that we look at the Wang DRMs of X given that it lies above a high level, with conditions as weak as possible on the function g, in an effort to be able to examine the extremes of X in as unified a way as possible.

3.2 Estimation using an asymptotic equivalent of a Wang DRM

We now give a first idea to estimate this type of extreme risk measure. Let $(X_1, \ldots, X_n)$ be a sample of independent and identically distributed copies of a random variable X having cdf F, and let (β_n) be a nondecreasing sequence of real numbers belonging to $(0, 1)$, which converges to 1. Assume for the time being that X is Pareto distributed, that is

$$\forall x > 1, \ \mathbb{P}(X \le x) = 1 - x^{-1/\gamma}$$

where $\gamma > 0$ is the so-called tail index of X. This is, as we shall recall in a short while, the simplest example of a heavy-tailed distribution. In this case, the quantile function of X is $q(\alpha) = (1 - \alpha)^{-\gamma}$ for all $\alpha \in (0, 1)$. In particular, q is continuous and strictly increasing on $(0, 1)$ so that using (1) in

Section 3.1 and a simple change of variables, we get:

$$\begin{aligned}
R_{g,\beta_n}(h(X)) = \int_0^1 h \circ q(1-\alpha) dg_{\beta_n}(\alpha) &= \int_0^1 h \circ q(1-(1-\beta_n)s) dg(s) \\
&= \int_0^1 h((1-\beta_n)^{-\gamma} s^{-\gamma}) dg(s) \\
&= \int_0^1 h(q(\beta_n) s^{-\gamma}) dg(s). \qquad (3)
\end{aligned}$$

In this case, an estimator of $R_{g,\beta_n}(h(X))$ would then be obtained by plugging estimators of $q(\beta_n)$ and γ in the right-hand side of (3).

Of course, in general, a strong relationship such as (3) cannot be expected to hold, but it shall stay true to some extent when X has a heavy-tailed distribution, the rigorous definition of which we recall now. A function f is said to be regularly varying at infinity with index $b \in \mathbb{R}$ if f is nonnegative and for any $x > 0$, $f(tx)/f(t) \to x^b$ as $t \to \infty$; the distribution of X is then said to be heavy-tailed when $1 - F$ is regularly varying with index $-1/\gamma < 0$, the parameter γ being the so-called tail index of the cdf F. This condition, which is a usual restriction in extreme value theory (see de Haan and Ferreira, 2006), essentially says that $1 - F(x)$ is in some sense close to $x^{-1/\gamma}$ when x is large. In the sequel, we therefore assume that X is heavy-tailed. We also suppose that the quantile function q of X is continuous and strictly increasing in a neighborhood of infinity, which shall make possible the use of (1) for n large enough.

Finally, we assume that the function h is a positive power of x: $h(x) = x^a$, where $a > 0$. This choice allows us to consider estimators of a large class of risk measures of X, including the aforementioned CTM. In this case (see Lemma 3 in the Appendix), it holds that

$$R_{g,\beta_n}(X^a) = [q(\beta_n)]^a \int_0^1 s^{-a\gamma} dg(s)(1 + \mathrm{o}(1)) \quad \text{as} \quad n \to \infty$$

provided $\int_0^1 s^{-a\gamma-\eta} dg(s) < \infty$ for some $\eta > 0$. This suggests that the above idea for the construction of the estimator can still be used provided n is large enough. Specifically, if $\widehat{F}_n$ denotes the empirical cdf related to this sample and $\widehat{q}_n$ denotes the empirical quantile function:

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\} \quad \text{and} \quad \widehat{q}_n(\alpha) = \inf\{t \in \mathbb{R} \,|\, \widehat{F}_n(t) \geq \alpha\} = X_{\lceil n\alpha \rceil, n}$$

in which $X_{1,n} \leq \cdots \leq X_{n,n}$ are the order statistics of the sample $(X_1, \ldots, X_n)$ and $\lceil \cdot \rceil$ is the ceiling function, we set

$$\widehat{R}^{AE}_{g,\beta_n}(X^a) := X^a_{\lceil n\beta_n \rceil, n} \int_0^1 s^{-a\widehat{\gamma}_n} dg(s) \qquad (4)$$

where $\widehat{\gamma}_n$ is any consistent estimator of γ. This estimator shall be called the AE estimator in what follows; notice that the integrability condition $\int_0^1 s^{-a\gamma-\eta} dg(s) < \infty$, which should be thought of as a condition that guarantees the existence of the considered Wang DRM, makes the estimator introduced here well-defined with probability arbitrarily large when n is large enough because of the consistency of $\widehat{\gamma}_n$. A related but different idea is used by Vandewalle and Beirlant (2006), although they work with the expression linking the Wang DRM with the survival function $1-F$, which is the reason why it is assumed there that g satisfies a regular variation property in a neighborhood of 0, and they finish their construction by plugging in an estimator of small exceedance probabilities and an external estimator of γ.

An appealing feature of the AE estimator is that it is easy to compute in many cases:

- in the case of the Conditional Tail moment of order a, *i.e.* $g(x)=x$, the estimator reads
$$\widehat{R}^{AE}_{g,\beta_n}(X^a) = X^a_{\lceil n\beta_n\rceil,n}\int_0^1 s^{-a\widehat{\gamma}_n}ds = \frac{X^a_{\lceil n\beta_n\rceil,n}}{1-a\widehat{\gamma}_n}$$
when $a\widehat{\gamma}_n<1$. In particular, this provides an estimator different from the sample average estimator of El Methni *et al.* (2014);

- in the case of the Dual Power risk measure, *i.e.* $g(x)=1-(1-x)^{1/\alpha}$ where $0<\alpha<1$ and $a=1$, then when $r:=1/\alpha$ is an integer, the estimator is
$$\widehat{R}^{AE}_{g,\beta_n}(X) = X_{\lceil n\beta_n\rceil,n}\int_0^1 rs^{-\widehat{\gamma}_n}(1-s)^{r-1}ds = \frac{r!\Gamma(1-\widehat{\gamma}_n)}{\Gamma(1-\widehat{\gamma}_n+r)}X_{\lceil n\beta_n\rceil,n}$$
provided $\widehat{\gamma}_n<1$. Here Γ is Euler's Gamma function, namely $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt$;

- in the case of the Proportional Hazard transform, *i.e.* $g(x)=x^\alpha$ where $0<\alpha<1$ and $a=1$, the estimator is
$$\widehat{R}^{AE}_{g,\beta_n}(X) = X_{\lceil n\beta_n\rceil,n}\int_0^1 \alpha s^{\alpha-\widehat{\gamma}_n-1}ds = \frac{\alpha X_{\lceil n\beta_n\rceil,n}}{\alpha-\widehat{\gamma}_n}$$
provided $\widehat{\gamma}_n<\alpha$.

The estimator $\widehat{R}^{AE}_{g,\beta_n}(X^a)$ does however require an external estimate of γ, and this should not be surprising because even in the simple Pareto case, estimating $R_{g,\beta_n}(X^a)$ requires in general more information about the tail of X than the knowledge of the single quantile $q(\beta_n)$. In a nutshell, we use the fact that the tail behavior of X is essentially known if a high quantile and the tail index can be consistently estimated simultaneously.

In order to examine the asymptotic properties of our estimator and precisely its asymptotic normality, it is necessary to compute the order of magnitude of its asymptotic bias. To do so, it is convenient to use an assumption on the left-continuous inverse U of $1/(1-F)$, defined by $U(t)=\inf\{x\in\mathbb{R}\,|\,1/(1-F(x))\geq t\}=q(1-t^{-1})$. Specifically, we assume that U is regularly varying with index γ and satisfies the following second-order condition (see de Haan and Ferreira, 2006):

Condition $\mathcal{C}_2(\gamma, \rho, A)$: for any $x > 0$, we have

$$\lim_{t\to\infty} \frac{1}{A(t)} \left(\frac{U(tx)}{U(t)} - x^\gamma \right) = x^\gamma \frac{x^\rho - 1}{\rho}$$

with $\gamma > 0$, $\rho \leq 0$ and A is a Borel measurable function which converges to 0 and has constant sign. When $\rho = 0$, the right-hand side is to be read as $x^\gamma \log x$.

We highlight that in condition $\mathcal{C}_2(\gamma, \rho, A)$, the function $|A|$ is necessarily regularly varying at infinity with index ρ (see Theorem 2.3.3 in de Haan and Ferreira, 2006). Such an assumption, which controls the rate of convergence of the ratio $U(tx)/U(t)$ to x^γ as $t \to \infty$, is classical when studying the rate of convergence of an estimator of a parameter describing the extremes of a random variable. All standard examples of heavy-tailed distributions in extreme-value theory satisfy this condition (see *e.g.* the examples pp.61–62 in de Haan and Ferreira, 2006).

The asymptotic properties of our estimator can be stated in this framework, as follows:

Theorem 1. *Assume that U is regularly varying with index $\gamma > 0$. Assume further that $\beta_n \to 1$ and $n(1-\beta_n) \to \infty$.*

1. *Pick a distortion function g and $a > 0$. If there is some $\eta > 0$ such that:*

$$\int_0^1 s^{-a\gamma-\eta} dg(s) < \infty$$

and $\widehat{\gamma}_n$ is a consistent estimator of γ, then

$$\frac{\widehat{R}^{AE}_{g,\beta_n}(X^a)}{R_{g,\beta_n}(X^a)} - 1 \xrightarrow{\mathbb{P}} 0 \quad as \quad n \to \infty.$$

2. *Assume moreover that U satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a $d-$tuple of distortion functions $(g_1, \ldots, g_d)$ and $a_1, \ldots, a_d > 0$. If for some $\eta > 0$,*

$$\forall j \in \{1, \ldots, d\}, \ \int_0^1 s^{-a_j\gamma-1/2-\eta} dg_j(s) < \infty,$$

then, provided we have the joint convergence:

$$\sqrt{n(1-\beta_n)} \left(\widehat{\gamma}_n - \gamma, \frac{X_{\lceil n\beta_n \rceil, n}}{q(\beta_n)} - 1 \right) \xrightarrow{d} (\Gamma, \Theta)$$

it holds that the random vector

$$\sqrt{n(1-\beta_n)} \left(\frac{\widehat{R}^{AE}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} - 1 \right)_{1\leq j\leq d}$$

asymptotically has the joint distribution of

$$\left(a_j\left[-\lambda\frac{\displaystyle\int_0^1 s^{-a_j\gamma}\frac{s^{-\rho}-1}{\rho}dg_j(s)}{\displaystyle\int_0^1 s^{-a_j\gamma}dg_j(s)}+\frac{\displaystyle\int_0^1 s^{-a_j\gamma}\log(1/s)dg_j(s)}{\displaystyle\int_0^1 s^{-a_j\gamma}dg_j(s)}\Gamma+\Theta\right]\right)_{1\le j\le d}.$$

It should be noted that Theorem 1 is obtained under the restriction $n(1-\beta_n)\to\infty$. Thus, it only ensures that the estimator consistently estimates so-called intermediate (*i.e.* not "too extreme") Wang DRMs, in the sense that the order of the smallest quantile that it takes into account must converge sufficiently slowly to 1. In other words, our first estimator should only be used to estimate those risk measures above a lower threshold $q(\beta_n)$ that belongs to the range covered by the available data. This restriction, which is undesirable from the practical point of view but standard nonetheless, will be lifted in Section 3.4 by the introduction of an estimator adapted to the extreme-value framework.

Another condition of this result, specific to the asymptotic normality statement, is that $(\widehat{\gamma}_n, X_{\lceil n\beta_n\rceil,n})$ should fulfill a joint convergence property; this is hardly a restrictive requirement in practice. For instance, if $\widehat{\gamma}_n=\widehat{\gamma}_{\beta_n}$ is the Hill estimator (Hill, 1975):

$$\widehat{\gamma}_{\beta_n}=H_n(\lceil n(1-\beta_n)\rceil)\ \text{ with }\ H_n(k)=\frac{1}{k}\sum_{i=1}^{k}\log\left(X_{n-i+1,n}\right)-\log\left(X_{n-k,n}\right), \tag{5}$$

then (see Theorem 2.4.8 and the proof of Theorem 3.2.5 in de Haan and Ferreira, 2006) we have the following joint convergence in distribution, under the bias condition $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1})\to\lambda$:

$$\sqrt{n(1-\beta_n)}\left(\widehat{\gamma}_n-\gamma,\frac{X_{\lceil n\beta_n\rceil,n}}{q(\beta_n)}-1\right)\xrightarrow{d}\left(\gamma\int_0^1[s^{-1}W(s)-W(1)]ds+\frac{\lambda}{1-\rho},\gamma W(1)\right)$$

where W is a standard Brownian motion, and therefore the right-hand side is a (possibly noncentered) Gaussian random pair. Similar joint convergence results can be found for other estimators of γ, such as the Pickands estimator (1975), the maximum likelihood estimator (Smith, 1987 and Drees *et al.*, 2004) and probability-weighted moment estimators (Hosking *et al.*, 1985, Diebolt *et al.*, 2007); see *e.g.* Sections 3 and 4 of de Haan and Ferreira (2006).

3.3 Estimation using a functional plug-in estimator

The simple estimator of Section 3.2 is not devoid of drawbacks though. For a start, the fact that it requires a plugged-in consistent estimator $\widehat{\gamma}_n$ of γ satisfying a certain integrability condition can indeed be a problem in small samples. When estimating an extreme Conditional Tail Expectation, for instance, the original requirement is $\gamma<1$, and what makes the estimator well-defined on a given sample is the condition $\widehat{\gamma}_n<1$. If the true value of γ is close to but below 1, the condition

$\widehat{\gamma}_n < 1$, while true with probability arbitrarily close to 1 as n grows to infinity, may fail to hold in a sizeable proportion of samples especially for small n. Even if the integrability requirement is fulfilled, if on a given sample the estimate of γ is a poor one then the resulting estimate of the Wang DRM is almost guaranteed to be a poor one as well even for moderate β_n. In theoretical terms, this means that the two successive approximations warranted by the heavy tails assumption on F,

$$q(1-(1-\beta_n)s) \approx q(\beta_n)s^{-\gamma} \approx X_{\lceil n\beta_n \rceil, n} s^{-\widehat{\gamma}_n},$$

which are at the heart of the construction of the AE estimator, may both introduce substantial errors.

Our idea now is to introduce an alternative estimator obtained by making a single approximation, which we can expect to perform better than the AE estimator. Recall that

$$R_{g,\beta_n}(h(X)) = \int_0^1 h \circ q(1-(1-\beta_n)s)dg(s).$$

We consider the statistic obtained by replacing the function $s \mapsto q(1-(1-\beta_n)s)$ by its empirical counterpart $s \mapsto \widehat{q}_n(1-(1-\beta_n)s) = X_{\lceil n(1-(1-\beta_n)s)\rceil, n}$. This yields the functional plug-in estimator

$$\widehat{R}^{PL}_{g,\beta_n}(h(X)) = \int_0^1 h \circ \widehat{q}_n(1-(1-\beta_n)s)dg(s) \tag{6}$$

which we call the PL estimator. Contrary to the AE estimator, the PL estimator is well-defined and finite with probability 1, and does not require an external estimator of γ. Its expression is a bit more involved though; in the case when $n(1-\beta_n)$ is actually a positive integer, which is fairly common in practice (see Sections 4 and 5), the PL estimator can actually be conveniently rewritten as an L-statistic, namely:

$$\begin{aligned}\widehat{R}^{PL}_{g,\beta_n}(h(X)) &= \sum_{i=1}^{n(1-\beta_n)} h(X_{n-i+1,n}) \int_0^1 \mathbb{I}\{x_{i-1,n}(\beta_n) \le s < x_{i,n}(\beta_n)\} dg(s) \\ &+ h(X_{n\beta_n,n}) \left[g(1) - \lim_{\substack{s\to 1\\ s<1}} g(s) \right] \text{ with } x_{i,n}(\beta_n) = \frac{i}{n(1-\beta_n)}\end{aligned}$$

or equivalently

$$\begin{aligned}\widehat{R}^{PL}_{g,\beta_n}(h(X)) &= \sum_{i=1}^{n(1-\beta_n)} h(X_{n-i+1,n}) \left[\lim_{\substack{s\to x_{i,n}(\beta_n)\\ s<x_{i,n}(\beta_n)}} g(s) - \lim_{\substack{s\to x_{i-1,n}(\beta_n)\\ s<x_{i-1,n}(\beta_n)}} g(s) \right] \\ &+ h(X_{n\beta_n,n}) \left[1 - \lim_{\substack{s\to 1\\ s<1}} g(s) \right].\end{aligned}$$

If g is further assumed to be continuous on $[0,1]$, a summation by parts shows that this L-statistic takes the simpler form

$$\widehat{R}^{PL}_{g,\beta_n}(h(X)) = h(X_{n\beta_n+1,n}) + \sum_{i=1}^{n(1-\beta_n)-1} g\left(\frac{i}{n(1-\beta_n)}\right) [h(X_{n-i+1,n}) - h(X_{n-i,n})].$$

Our aim is now to examine the asymptotic properties of the PL estimator. A technical complication comes from the fact that the level β_n is assumed to converge to 1 as n goes to infinity, and therefore our study is essentially different from what can be done for fixed β, see for example Jones and Zitikis (2003). While the above expressions of the estimator as an L-statistic are undoubtedly of practical value, we shall actually use in our proofs the basic expression (6) of our estimator as an integral of the tail quantile process $s \mapsto \widehat{q}_n(1-(1-\beta_n)s)$ and utilize the powerful distributional approximation of this process stated in Theorem 2.1 of Drees (1998), relating it to a standard Brownian motion up to a bias term.

Our first result on the PL estimator is the following:

Theorem 2. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a $d-$tuple of distortion functions $(g_1,\ldots,g_d)$ and $a_1,\ldots,a_d > 0$. If for some $\eta > 0$,*

$$\forall j \in \{1,\ldots,d\}, \ \int_0^1 s^{-a_j\gamma-1/2-\eta} dg_j(s) < \infty,$$

then:

$$\sqrt{n(1-\beta_n)}\left(\frac{\widehat{R}^{PL}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} - 1\right)_{1\leq j\leq d} \xrightarrow{d} \mathcal{N}(0,V)$$

with V being the $d\times d$ matrix whose $(i,j)-$th entry is

$$V_{i,j} = a_i a_j \gamma^2 \frac{\int_{[0,1]^2} \min(s,t) s^{-a_i\gamma-1} t^{-a_j\gamma-1} dg_i(s) dg_j(t)}{\int_0^1 s^{-a_i\gamma} dg_i(s) \int_0^1 t^{-a_j\gamma} dg_j(t)}.$$

This asymptotic normality result, unsurprisingly, is also restricted to the case $n(1-\beta_n) \to \infty$, as Theorem 1 was. We may draw an interesting consequence from Theorem 2 though: it is compelling that the PL estimator is, for $g(x) = x$ and $h(x) = x^a$, identical to the sample average estimator introduced in El Methni *et al.* (2014), so one may wonder if Theorem 2 agrees with their result. More broadly, for $b \in \mathbb{R}$, let us consider the class of functions

$$\mathcal{E}_b([0,1]) := \left\{ g : [0,1] \to \mathbb{R} \,|\, g \text{ continuously differentiable on } (0,1) \text{ and } \limsup_{s\to 0} s^{-b}|g'(s)| < \infty \right\}.$$

Roughly speaking, the classes $\mathcal{E}_b$, $b > -1$, can be considered as the spaces of those functions g which are continuously differentiable on $(0,1)$ and whose first derivative behaves at most like a power of s in a neighborhood of 0. Especially, any polynomial function belongs to $\mathcal{E}_0([0,1])$, and the Proportional Hazard (Wang, 1996) distortion function $g(s) = s^\alpha$, $\alpha \in (0,1)$ belongs to $\mathcal{E}_{\alpha-1}([0,1])$. For a given distortion function g, the convergence condition $\int_0^1 s^{-a\gamma-1/2-\eta} dg(s) < \infty$ being determined by the behavior of g in a neighborhood of 0, it is obvious that checking the condition $g \in \mathcal{E}_b([0,1])$ (for some b) enables one to rewrite such an integrability hypothesis in a simpler fashion. Our next result focuses on this case:

Corollary 1. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a $d-$tuple of distortion functions $(g_1,\ldots,g_d)$ and $a_1,\ldots,a_d > 0$. Assume there are $b_1,\ldots,b_d \in \mathbb{R}$ such that for all $j \in \{1,\ldots,d\}$, we have $g_j \in \mathcal{E}_{b_j}([0,1])$. If*

$$\forall j \in \{1,\ldots,d\},\ \gamma < \frac{2b_j+1}{2a_j}$$

then:

$$\sqrt{n(1-\beta_n)}\left(\frac{\widehat{R}^{PL}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})}-1\right)_{1\le j\le d} \xrightarrow{d} \mathcal{N}(0,V)$$

with V as in Theorem 2.

In particular, the condition on γ we get for the asymptotic normality of the CTM of order a, obtained with $g(x)=x$ and thus $g \in \mathcal{E}_0([0,1])$, is $\gamma < 1/2a$, which is the condition obtained by El Methni *et al.* (2014). Since moreover

$$\forall a_1,a_2>0,\ \int_{[0,1]^2} \min(s,t)s^{-a_1\gamma-1}t^{-a_2\gamma-1}ds\,dt = \frac{2-(a_1+a_2)\gamma}{(1-a_1\gamma)(1-a_2\gamma)(1-(a_1+a_2)\gamma)}$$

when $\gamma < (2\max(a_1,a_2))^{-1}$, one may also readily check that the asymptotic variance is the same as in Theorem 1 there. We highlight however that the assumption $\mathcal{C}_2(\gamma,\rho,A)$ is somewhat stronger than the (conditional) assumptions made on F by El Methni *et al.* (2014). This is because the proofs of Theorem 2 and Corollary 1 are also designed to address the case of functions g_j that may be much more difficult to handle than the simple identity function, which is the only case addressed by Theorem 1 of El Methni *et al.* (2014).

Just like the AE estimator, the PL estimator is only consistent when (β_n) is an intermediate sequence. Our purpose is now to remove this restriction by showing how our extreme-value framework and the expression of our estimators make it possible to use the extrapolation methodology of Weissman (1978) in order to estimate proper extreme Wang DRMs.

3.4 Estimating extreme risk measures of arbitrary order

In order to design a consistent estimator of an arbitrarily extreme risk measure, we remark that for any $s \in (0,1)$ and $a>0$ we have:

$$[q(1-(1-\delta_n)s)]^a = \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\gamma}[q(1-(1-\beta_n)s)]^a(1+\mathrm{o}(1))$$

as $n\to\infty$, as a consequence of the regular variation property of U and provided that (β_n) is a sequence converging to 1 such that $(1-\delta_n)/(1-\beta_n)$ converges to a positive limit. In other words, the value of the quantile function at an arbitrarily extreme level is essentially its value at a much smaller level up to an extrapolation factor that depends on the unknown tail index γ. Integrating the above relationship with respect to the distortion measure dg therefore suggests that:

$$R_{g,\delta_n}(X^a) = \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\gamma} R_{g,\beta_n}(X^a)(1+\mathrm{o}(1)),$$

see Lemma 5 for a stronger and rigorous statement. A way to design an adapted estimator of the extreme risk measure $R_{g,\delta_n}(X^a)$, when $n(1-\delta_n) \to c < \infty$, is thus to take a sequence (β_n) such that $n(1-\beta_n) \to \infty$ and to plug the AE estimator and a consistent estimator $\widehat{\gamma}_n$ of γ in the right-hand side of the above equality: this yields the extrapolated AE estimator

$$\widehat{R}^{W,AE}_{g,\delta_n}(X^a;\beta_n) := \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\widehat{\gamma}_n} \widehat{R}^{AE}_{g,\beta_n}(X^a).$$

This principle can also be applied to the PL estimator, to obtain an extrapolated PL estimator:

$$\widehat{R}^{W,PL}_{g,\delta_n}(X^a;\beta_n) := \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\widehat{\gamma}_n} \widehat{R}^{PL}_{g,\beta_n}(X^a).$$

Both estimators can actually be seen as particular Weissman-type estimators of $R_{g,\delta_n}(X^a)$ (see Weissman, 1978):

$$\widehat{R}^{W}_{g,\delta_n}(X^a;\beta_n) := \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a\widehat{\gamma}_n} \widehat{R}_{g,\beta_n}(X^a)$$

where $\widehat{R}_{g,\beta_n}(X^a)$ is some relatively consistent estimator of the intermediate Wang DRM $R_{g,\beta_n}(X^a)$; in fact, in both cases, we exactly recover Weissman's estimator of an extreme quantile by setting $a=1$ and $g(x)=0$ if $x<1$. Besides, taking $g(x)=x$, the extrapolated PL estimator becomes the estimator of the extreme CTM of X introduced in El Methni *et al.* (2014). All the aforementioned estimators are based on the same idea: to estimate the quantity of interest at an arbitrarily extreme level, this quantity is estimated first at an intermediate level where an estimator is known to be consistent, and then multiplied by an extrapolation factor which depends on a consistent, external estimator of the tail index γ.

Our second main result examines the asymptotic distribution of this type of estimators.

Theorem 3. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$, with $\rho<0$. Assume further that $\beta_n,\delta_n \to 1$, $n(1-\beta_n)\to\infty$, $(1-\delta_n)/(1-\beta_n)\to 0$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a $d-$tuple of distortion functions $(g_1,\ldots,g_d)$ and $a_1,\ldots,a_d>0$. If for some $\eta>0$,*

$$\forall j \in \{1,\ldots,d\},\ \int_0^1 s^{-a_j\gamma-1/2-\eta}dg_j(s) < \infty$$

and $\sqrt{n(1-\beta_n)}(\widehat{\gamma}_n-\gamma) \xrightarrow{d} \xi$ then provided

$$\forall j \in \{1,\ldots,d\},\ \sqrt{n(1-\beta_n)}\left(\frac{\widehat{R}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})}-1\right) = \mathrm{O}_{\mathbb{P}}(1)$$

we have that

$$\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}\left(\frac{\widehat{R}^W_{g_j,\delta_n}(X^{a_j};\beta_n)}{R_{g_j,\delta_n}(X^{a_j})}-1\right)_{1\le j\le d}\xrightarrow{d}\begin{pmatrix}a_1\xi\\ \vdots\\ a_d\xi\end{pmatrix}.$$

Again, in the particular case $d=1$, $a=1$ and $g(x)=0$ if $x<1$, we recover the asymptotic result about Weissman's estimator, see Theorem 4.3.8 in de Haan and Ferreira (2006); for $g(x)=x$ and $d=1$, we recover a result similar to Theorem 2 of El Methni *et al.* (2014) if the intermediate estimator is the PL estimator. As far as practical situations are concerned, the estimation of the parameter γ is of course a central question, not least because the asymptotic distribution of our Weissman-type estimators is exactly determined by the estimator of γ which is used. Classical tail index estimators such as those mentioned at the end of Section 3.2 are computed using a number $k=k(n)\to\infty$ of order statistics of the sample (with $k/n\to 0$) and are $\sqrt{k}-$asymptotically normal under conditions akin to ours. It is then convenient to set $k=\lceil n(1-\beta_n)\rceil$, which ensures that the estimator of γ converges at the required rate $\sqrt{n(1-\beta_n)}$. The choice of the intermediate level β_n, which is crucial, is a difficult problem however, and we discuss a possible selection rule in our simulation study below.

4 Simulation study

The finite-sample performance of our estimators is illustrated on the following simulation study, where we consider a couple of classical heavy-tailed distributions and three different distortion functions g. The distributions studied are:

- the Fréchet distribution: $F(x)=\exp(-x^{-1/\gamma})$, $x>0$;
- the Burr distribution: $F(x)=1-(1+x^{-\rho/\gamma})^{1/\rho}$, $x>0$ (here $\rho\le 0$).

Both of these distributions have extreme value index γ and their respective second-order parameters are -1 and ρ, see *e.g.* Beirlant *et al.* (2004). We consider the following distortion functions:

- the Conditional Tail Expectation (CTE) function $g(x)=x$ which weights all quantiles equally;
- the Dual Power (DP) function $g(x)=1-(1-x)^{1/\alpha}$ with $\alpha\in(0,1)$, which gives higher weight to large quantiles. When $r:=1/\alpha$ is a positive integer, the related DRM is the expectation of $\max(X_1,\ldots,X_r)$ for independent copies $X_1,\ldots,X_r$ of X;
- the Proportional Hazard (PH) transform function $g(x)=x^\alpha$ with $\alpha\in(0,1)$, which gives higher and unbounded weight to large quantiles in the sense that $g'(s)\uparrow\infty$ as $s\downarrow 0$.

By Theorem 1, the AE estimator is consistent when $\gamma < 1$ for either of the two first distortion functions, and for $\gamma < \alpha$ when the Proportional Hazard transform is considered. By contrast, Theorems 2 and 3 ensure that the PL estimator shall be valid for $\gamma < 1/2$ for either of the two first distortion functions, and when $\gamma < \alpha - 1/2$ for the Proportional Hazard transform. In some sense, a suitable choice of α in the PH transform case will allow us to check if the PL technique is robust to a violation of the integrability condition of Theorems 2 and 3.

Our estimators being based on a preliminary estimation at level β_n where (β_n) is some intermediate sequence, we first discuss the choice of this level. This step is crucial: choosing β_n too close to 1 increases the variance of the estimator dramatically, while choosing β_n too far from 1 results in biased estimates. There has been a great amount of research carried out recently on this choice: an excellent overview of possible techniques, including bootstrap methods, Pareto quantile plots or procedures based on the analysis of finite-sample bias, is given in Section 5.4 of Gomes and Guillou (2015). In many practical cases though, the analysis of a data set from the point of view of extremes starts by drawing a plot of one or several tail index estimators, and then by selecting β_n in a region contained in the extremes of the sample where the estimation is "stable". Our purpose here is to suggest an automatic such choice. We work with the popular Hill estimator (Hill, 1975), see (5), which we shall also use to estimate the extreme value index γ. Our idea is to detect the last stability region in the Hill plot $\beta \mapsto \widehat{\gamma}_\beta$; choosing β in this region most often realizes a decent bias-variance trade-off. Specifically:

- choose $\beta_0 > 0$ and a window parameter $h > 1/n$;
- for $\beta_0 < \beta < 1 - h$, let $I(\beta, h) = [\beta, \beta + h]$ and compute the standard deviation $\sigma(\beta, h)$ of the set of estimates $\{\widehat{\gamma}_b,\, b \in I(\beta, h)\}$;
- if $\beta \mapsto \sigma(\beta, h)$ is monotonic, let β_{lm} be β_0 if it is increasing and $1 - h$ if it is decreasing;
- otherwise, denote by β_{lm} the last value of β such that $\sigma(\beta, h)$ is locally minimal and its value is less than the average value of the function $\beta \mapsto \sigma(\beta, h)$;
- choose β^* such that $\widehat{\gamma}_{\beta^*}$ is the median of $\{\widehat{\gamma}_b, b \in I(\beta_{lm}, h)\}$.

This procedure is somewhat related to others in the extreme value literature (see *e.g.* Resnick and Stărică, 1997, Drees *et al.*, 2000, de Sousa and Michailidis, 2004 and Frahm *et al.*, 2005); closely related procedures when there is random covariate information can be found in Stupfler (2013), Gardes and Stupfler (2014) and Stupfler (2016). An illustration of this technique on a simulated data set is given in Figure 1.

In each case, we carry out our computations on $N = 5000$ independent samples of $n \in \{100, 300\}$ independent copies of X; our choice procedure is conducted with $\beta_0 = 0.5$ and $h = 0.1$. We record relative mean squared errors (MSEs):

$$\mathrm{MSE}(\widehat{R}_{g,\delta}^W) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\widehat{R}_{g,\delta}^W(X;\beta_i^*)}{R_{g,\delta}(X)} - 1\right)^2$$

at $\delta = 0.99$, 0.995 and 0.999 (here β_i^* is the chosen intermediate level for the $i-$th sample). Our results are reported in Tables 3–5. It appears on these examples that the PL estimator performs at least as well as the AE estimator, as expected; besides, the PL estimator performs markedly better than the AE estimator both for smaller samples or when the condition γ tightens, as can be seen by comparing the results obtained when $n = 100$ for the DP(1/3) or PH(2/3) risk measure. Results deteriorate when γ increases: a possible explanation lies in the fact that the asymptotic distribution of our estimator is essentially that of $\widehat{\gamma}_n$ by Theorem 3, which is a Gaussian distribution with variance proportional to γ^2 (see Theorem 3.2.5 in de Haan and Ferreira, 2006). Results however improve when $|\rho|$ increases, which was expected since the larger is $|\rho|$, the smaller is the bias in the estimation.

Besides, the PL estimator seems to be at least somewhat robust to a violation of the integrability condition in Theorems 2–3, as can be seen on the example of the PH(2/3) risk measure with $\gamma = 1/4$. When comparing the results for the CTE and PH(2/3) risk measures, it can also be seen that results deteriorate as the limit of $g'(s)$ as $s \downarrow 0$ increases. This likely comes from the fact that an increasing such limit amplifies the error made by the empirical quantile function, all the more so as the latter error itself increases when estimating quantiles whose order is very close to 1; to put it differently, sample quantiles at extreme levels do not estimate the corresponding true quantiles consistently, which is why the most extreme values in a sample tend not to give a fair picture of the extremes of the underlying distribution (see *e.g.* Ghosh and Resnick, 2010). The AE estimator, meanwhile, could have been thought to provide additional robustness against this defect, since it only depends on a single intermediate order statistic and the Hill estimator, but it actually fails to improve upon the PL estimator, most likely because the multiplicative factor $(2/3 - \widehat{\gamma}_n)^{-1}$ makes it severely underperform in some cases, see the opening discussion of Section 3.3. In our opinion, improving the finite-sample performance of the AE and PL estimators in cases when the integrability condition is close to be violated is an important question and should be part of our future work.

5 Real data application

5.1 Analysis of extreme swings of the results curve of a professional poker player

We apply our method to the study of the results of high-stakes poker player Tom Dwan. The original data, extracted from results publicly available at `http://www.highstakesdb.com`, consists in his cumulative results on the Internet, aggregated over all poker variants and recorded approximately every five days from mid-October 2008 to April 2011. In this study, we focus on the sub-parts of the results curve where it is monotonic, namely, the periods of time when this player is either consistently winning or losing. The analysis of such timeframes, which may last from a few days to several weeks, is of great value to poker players since it helps them understand their own behavior (and possibly that of their opponents as well) during winning and losing streaks.

To this end, we record the values of the local minima and maxima of the results curve and we construct the differences between two such consecutive points. The data is now made of $n = 68$ observations, alternatively positive and negative, which represent the aggregated results during alternative winning and losing streaks. Our specific aim here is to analyze the extreme such streaks (also called "swings" in poker technical terms). Our data X_t, represented in Figure 2, is the absolute value of the 68 observations at our disposal, and the analysis will thus focus on the magnitude of the extreme swings of the results curve, irrespective of whether such a swing corresponds to a win or a loss. Of course, this leads to a loss of information and it would clearly be interesting to analyze the winning and losing streaks separately. Note though that the data on winning streaks is made of only 34 observations; the rate of convergence of our technique being only the square-root of a fraction of the total sample size, we believe that the separate data sets are too small to carry out an analysis which is interpretable from the extreme-value point of view. It should also be pointed out that a statistical analysis did not reveal a significant difference between the tail indices of winning and losing swings at the 5% error rate.

Since we work on time series data, there are particular concerns about independence and stationarity. In this particular context, these concerns are warranted because high-stakes poker players are part of a fairly small community (which may impact the independence assumption) and therefore have to often change their playing style so as to avoid displaying a distinctive playing pattern that makes them easily readable by experienced opponents (and this could violate the stationarity assumption). These hypotheses are checked using the turning point test (see Kendall and Stuart, 1968) contained in the R package `randtests`; the $p-$value of this test is 0.278 and thus we cannot reject the i.i.d.

assumption basing on this procedure. Since such a test is known to be suitable against cyclicity but poor against trends (Kendall, 1973), we also run the KPSS test for trend stationarity (Kwiatkowski *et al.*, 1992) contained in the R package `tseries`, whose $p-$value is greater than 0.1 for an estimated trend parameter of $\widehat{m} = -15.236$ (estimated via a linear regression) and a lag parameter of 1 in the Newey-West variance estimator. The stationarity assumption can therefore be assumed to be reasonable on the detrended time series $X_t - \widehat{m}t$, which is the sample of data we apply our procedures on in what follows; this idea is confirmed by the KPSS test for level stationarity, also part of the `tseries` package, whose $p-$value is greater than 0.1. Finally, let us note that the plot of the sample autocorrelation function (see Figure 3) does not indicate significant correlation in the data.

Our next aim is to estimate the extreme value index γ of the detrended sample. Since the sample size is fairly small, we use the Hill estimator together with a bias-reduced version inspired by the work of Peng (1998):

$$\widehat{\gamma}_\beta^{RB}(\tau) = \frac{1}{\widehat{\rho}_{\beta_1}(\tau)}\widehat{\gamma}_\beta + \left(1 - \frac{1}{\widehat{\rho}_{\beta_1}(\tau)}\right)\frac{\widehat{\gamma}_\beta^S}{2\widehat{\gamma}_\beta},$$

with

$$\widehat{\gamma}_\beta^S = \frac{1}{\lceil n(1-\beta)\rceil}\sum_{i=1}^{\lceil n(1-\beta)\rceil}\left(\log X_{n-i+1,n} - \log X_{n-\lceil n(1-\beta)\rceil,n}\right)^2$$

and $\widehat{\rho}_{\beta_1}(\tau)$ is the consistent estimator of ρ presented in equation (2.18) of Fraga Alves *et al.* (2003) which depends on a different sample fraction $1-\beta_1$ and a tuning parameter $\tau \leq 0$. By Theorem 2.1 in Peng (1998),

$$\sqrt{n(1-\beta_n)}(\widehat{\gamma}_{\beta_n}^{RB}(\tau) - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\frac{1-2\rho+2\rho^2}{\rho^2}\right) \tag{7}$$

provided (β_n) is an intermediate sequence. The generalized jackknife estimator $\widehat{\gamma}_\beta^{RB}(\tau)$ is thus essentially a suitably weighted combination of the Hill estimator and a similar estimator, the coefficients being estimates of those which make the asymptotic biases cancel out. We take $\beta_1 = 1 - \lceil n^{0.975}\rceil/n \approx 0.0882$, as recommended by Caeiro *et al.* (2009).

Some estimates of γ are given in Table 6 and Hill plots are represented in Figure 4. The Hill estimator seems to drift away fairly quickly due to the finite-sample bias, and we decide to drop it for our analysis. We then estimate γ by the median of the bias-reduced estimates obtained for $\tau \in \{0, 1/4, 1/2, 3/4, 1\}$: in each case, the estimate is obtained by a straightforward adaptation of the selection procedure detailed in Section 4. We get $\widehat{\gamma} = 0.158$ for $\beta^* = 0.912$ and $\tau = 1/2$; especially, ρ is estimated by $\widehat{\rho} = -1.130$. Finally, Table 8 gives estimates of some risk measures for the detrended data set and Figure 5 represents the estimates of some extreme quantile lines for the time series X_t, obtained by re-adding the trend component $\widehat{m}t$ to our estimates of the VaR. From these results, it appears in particular that the maximal value in this data set, corresponding to a losing streak costing more than 6.1 million USD, exceeds our estimate of the 99% quantile. It is also of the same order of magnitude as our estimates of the CTE and DP(1/2) (resp. DP(1/3)) risk measure in the 1% highest cases, which corresponds to the average value of the maximum of two (resp. three) consecutive extreme results. In our opinion, this losing streak can thus be regarded as an extreme period of loss indeed.

5.2 The Secura Belgian Re actuarial data set on automobile claims

We consider here the Secura Belgian Re data set on automobile claims from 1998 until 2001, introduced in Beirlant *et al.* (2004) and further analyzed in Vandewalle and Beirlant (2006) from the extreme-value perspective. The data set consists of $n = 371$ claims which were at least as large as 1.2 million Euros and were corrected for inflation. Our aim here is to revisit this data set and show how we can recover results essentially equivalent to those of Vandewalle and Beirlant (2006) although they worked in a different context.

We start as in Section 5.1 by estimating the extreme value index γ. We use again the Hill estimator and some of its bias-reduced versions: Hill plots are represented in Figure 6, on which we can see that all our selected estimators give very close estimates. Results using our selection procedure are given in Table 7. Retaining the median estimate of γ yields $\widehat{\gamma} = 0.261$ for $\beta^* = 0.792$ and $\tau = 1/2$, with $\widehat{\rho} = -1.064$. Table 9 gives estimates of some risk measures for this data set.

The main example of excess-of-loss reinsurance policy that Vandewalle and Beirlant (2006) considered, namely the net premium principle, can actually be recovered from these estimates. Indeed, according to Vandewalle and Beirlant (2006), the risk premium for a reinsurance policy in excess of a high retention level R is $\int_R^\infty g(1-F(x))dx$ and the net premium $\mathrm{NP}(R)$ is obtained with $g(x) = x$, that is

$$\mathrm{NP}(R) = \int_R^\infty [1 - F(x)]dx.$$

Besides, we have from equation 2 that

$$R_{g,\beta}(X) = q(\beta) + \int_{q(\beta)}^\infty g\left(\frac{1 - F(x)}{1 - \beta}\right) dx.$$

Setting $g(x) = x$ and rearranging gives the identity

$$\mathrm{NP}(q(\beta)) = (1 - \beta)(R_{g,\beta}(X) - \mathrm{VaR}(\beta))$$

and in particular, the right-hand side is actually $\mathrm{SP}(\beta)$. We can therefore estimate the net premium above a high level of retention R, provided that we can match the level R with an estimated quantile of the underlying distribution *i.e.* estimate the exceedance probability $\mathbb{P}(X > R)$. When R is equal to 5 million Euros, as considered in Vandewalle and Beirlant (2006), it can be seen that this exceedance probability $1 - \beta$ is estimated to be approximately 0.02, or in other words that R is the estimated VaR at the 98% level. Estimates of our risk measures at this level are provided in Table 9; in particular, the net premium is estimated to be approximately 36,000 Euros, which is in line with the 41,798 Euros that Vandewalle and Beirlant (2006) obtained, with our estimate being slightly lower partly because a bias-reduced estimate of γ was used in the present work, whereas Vandewalle and Beirlant (2006) computed a simple Hill estimate.

6 Discussion

In the application of statistics to insurance and finance, the study of extreme risk is of prime importance, especially in view of the recent European Union Solvency II directive. A way of studying risk above a high level, by the means of Wang distortion risk measures (DRMs), is introduced here, and we believe that a major part of the value of our work lies in the flexibility and generality of the proposed class of extreme Wang DRMs. In addition to the numerous risk measures that it contains, such as the (extreme versions of) Value-at-Risk or Conditional Tail Moments, it should be noted that many other interesting quantities such as the Conditional Value-at-Risk, Conditional Tail Variance or Stop-loss Premium can be obtained by suitable finite linear combinations of elements of this class.

We also provide two simple classes of estimators to estimate our concept of extreme Wang DRM when the underlying distribution is heavy-tailed. Because our asymptotic results about our estimators of extreme Wang DRMs are joint convergence results for finitely many of them, our work makes it theoretically possible to give a detailed picture of extreme risk in practical situations; the finite-sample procedure we introduce, which is completely data-driven and has decent performance when the tail index is moderate, is a step towards achieving this goal in practice. Our methodology is applied to two sets of real data, one financial and one actuarial. Regarding this latter data set, our results essentially match earlier results of Vandewalle and Beirlant (2006), although our class of extreme Wang DRMs and our estimation results can arguably be considered as more general than theirs, not least because we recover the usual Value-at-Risk as a particular case although they do not.

Because the proposed class of extreme Wang DRMs allows for almost total freedom in choosing how to weight quantiles above a high level, it should be highlighted that it allows for yet many other interesting problems to be tackled. A possibility opened by the present study is to look for instance at extreme conditional versions of Dual Power (DP) distortion risk measures; when the parameter of this risk measure is the inverse of a positive integer r, the DP risk measure is actually the expectation of the maximum $M_r = \max(X_1, \ldots, X_r)$ of independent copies of the random variable of interest X above a high threshold. This is of course interesting in financial contexts, as our real data application to the results curve of high-stakes poker player Tom Dwan shows.

As far as actuarial applications are concerned, a possible situation is the following: when insurance firms have to cover against flood risk, then assuming that r floods occur in a given year, a catastrophic event occurs when the maximum $M_r = \max(X_1, \ldots, X_r)$ of water levels during these flood episodes exceeds a given extreme level (for instance, one meter plus the height of the dykes protecting the shores). If flood heights can reasonably be thought to be independent then such a problem can be examined as a simple application of the devices developed in this paper. Another appealing perspective lies in the fact that our class of extreme Wang DRMs contains (if finite linear combinations are allowed) certain reinsurance objects such as the Stop-loss Premium and therefore, similarly to what is outlined in our application to the Secura Belgian Re data set, our framework may also be applied to certain reinsurance calculations.

References

Adrian, T., Brunnermeier, M.K. (2008). CoVaR. *Federal Reserve of New York Staff Reports* **348**.

Alvarado, E., Sandberg, D.V., Pickford, S.G. (1998). Modeling large forest fires as extreme events. *Northwest Science* **72**, 66–75.

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

Basel Committee on Banking Supervision (2006). *International convergence of capital measurement and capital standards: a revised framework (comprehensive version).*

Basel Committee on Banking Supervision (2011). *Revisions to the Basel II market risk framework.*

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L. (2004). *Statistics of extremes – Theory and applications.* Wiley Series in Probability and Statistics.

Beirlant, J., Vynckier, P., Teugels, J.L. (1996). Excess functions and estimation of the extreme value index. *Bernoulli* **2**(4), 293–318.

Belles-Sampera, J., Guillén, M., Santolino, M. (2014). Beyond Value-at-Risk: GlueVaR distortion risk measures. *Risk Analysis* **34**(1), 121–134.

Bingham, N.H., Goldie, C.M., Teugels, J.L. (1987). *Regular Variation.* Cambridge, U.K.: Cambridge University Press.

Caeiro, F., Gomes, M.I., Rodrigues, L.H. (2009). Reduced-bias tail index estimators under a third-order framework. *Comm. Statist. Theory Methods* **38**, 1019–1040.

Cherny, A., Madan, D. (2009). New measures for performance evaluation. *Review of Financial Studies* **22**(7), 2571–2606.

Cont, R., Deguest, R., Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance* **10**(6), 593–606.

Cotter, J., Dowd, K. (2006). Extreme spectral risk measures: an application to futures clearinghouse margin requirements. *Journal of Banking and Finance* **30**(12), 3469–3485.

Daníelsson, J., Jorgensen, B.N., Samorodnitsky, G., Sarma, M., de Vries, C.G. (2013). Fat tails, VaR and subadditivity. *J. Econometrics* **172**(2), 283–291.

Davis, M.H.A. (2013). Consistency of risk measure estimates, preprint, available at
`http://ssrn.com/abstract=2342279`.

Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R. (2005). *Actuarial theory for dependent risks: measures, orders and models.* John Wiley & Sons, Chichester.

Diebolt, J., Guillou, A., Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method. *J. Statist. Plann. Inference* **137**(3), 841–857.

Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9**(1), 617–657.

Drees, H., Ferreira, A., de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14**(3), 1179–1201.

Drees, H., de Haan, L., Resnick, S. (2000). How to make a Hill plot. *Ann. Statist.* **28**(1), 254–274.

El Methni, J., Gardes, L., Girard, S. (2014). Nonparametric estimation of extreme risks from conditional heavy-tailed distributions. *Scand. J. Stat.* **41**(4), 988–1012.

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

Embrechts, P., Wang, B., Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics* **19**(4), 763–790.

Fraga Alves, M.I., Gomes, M.I., de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Port. Math.* **60**, 193–213.

Frahm, G., Junker, M., Schmidt, R. (2005). Estimating the tail-dependence coefficient: properties and pitfalls. *Insurance Math. Econom.* **37**(1), 80–100.

Furman, E., Landsman, Z. (2006). On some risk-adjusted tail-based premium calculation principles. *Journal of Actuarial Practice* **13**, 175–190.

Gardes, L., Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes* **17**(1), 45–75.

Ghosh, S., Resnick, S. (2010). A discussion on mean excess plots. *Stoch. Proc. Appl.* **120**(8), 1492–1517.

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

Gomes, M.I., Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review. *Int. Stat. Rev.* **83**(2), 263–292.

Gordy, M.B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation* **12**, 199–232.

Guegan, D., Hassani, B. (2015). Distortion risk measures or the transformation of unimodal distributions into multimodal functions, in *Future Perspectives in Risk Models and Finance* (eds. A. Bensoussan, D. Guegan, C.S. Tapiero), 71–88. Springer.

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

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

Hosking, J.R.M., Wallis, J.R., Wood, E.F. (1987). Estimation of the generalized extreme value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251–261.

Jones, B.L., Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal* **7**(4), 44–54.

Kendall, M.G., Stuart, A. (1968). *The advanced theory of statistics, Volume 3: Design and analysis, and time-series* (second edition). Griffin, London.

Keppo, J., Kofman, L., Meng, X. (2010). Unintended consequences of the market risk requirement in banking regulation. *Journal of Economic Dynamics and Control* **34**(10), 2192–2214.

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

Kou, S.G., Peng, X. (2014). On the measurement of economic tail risk, preprint, available at `http://arxiv.org/pdf/1401.4787v2.pdf`.

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrological Sciences Journal* **49**(4), 591–610.

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

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econometrics* **54**, 159–178.

Linton, O., Xiao, Z. (2013). Estimation of and inference about the expected shortfall for time series with infinite variance. *Econometric Theory* **29**(4), 771–807.

McNeil, A.J., Frey, R., Embrechts, P. (2005). *Quantitative risk management: concepts, techniques, and tools.* Princeton University Press.

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

Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statist. Probab. Lett.* **38**, 107–115.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**(1), 119–131.

Resnick, S., Stărică, C. (1997). Smoothing the Hill estimator. *Adv. in Appl. Probab.* **29**, 271–293.

Rootzén, H., Tajvidi, N. (1997). Extreme value statistics and wind storm losses: a case study. *Scand. Actuar. J.* **1**, 70–94.

Sereda, E.N., Bronshtein, E.M., Rachev, S.T., Fabozzi, F.J., Sun, W., Stoyanov, S.V. (2010). Distortion risk measures in portfolio optimization, in *Handbook of Portfolio Construction* (ed. J.B. Guerard), 649–673.

Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174–1207.

de Sousa, B., Michailidis, G. (2004). A diagnostic plot for estimating the tail index of a distribution. *J. Comput. Graph. Statist.* **13**(4), 974–995.

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

Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *J. Multivariate Anal.* **144**, 1–24.

Vandewalle, B., Beirlant, J. (2006). On univariate extreme value statistics and the estimation of reinsurance premiums. *Insurance Math. Econom.* **38**, 441–459.

Wang, S.S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bull.*

26, 71–92.

Wang, S.S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance* **67**(1), 15–36.

Wang, R., Ziegel, J.F. (2015). Elicitable distortion risk measures: A concise proof. *Statist. Probab. Lett.* **100**, 172–175.

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

Wirch, J.L., Hardy, M. R. (1999). A synthesis of risk measures for capital adequacy. *Insurance Math. Econom.* **25**(3), 337–347.

Wirch, J.L., Hardy, M. R. (2002). Distortion risk measures: coherence and stochastic dominance, in *International Congress on Insurance: Mathematics and Economics*, Lisbon.

Ziegel, J.F. (2015). Coherence and elicitability. *Math. Finance*, to appear, available at `http://arxiv.org/pdf/1303.1690v3.pdf`.

Appendix

Proofs of the main results

Proof of Theorem 1. Write for any j:

$$\frac{\widehat{R}^{AE}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} = \left[\frac{X_{\lceil n\beta_n\rceil,n}}{q(\beta_n)}\right]^a \times \frac{\int_0^1 s^{-a_j\widehat{\gamma}_n} dg_j(s)}{\int_0^1 s^{-a_j\gamma} dg_j(s)} \times \frac{[q(\beta_n)]^a \int_0^1 s^{-a_j\gamma} dg_j(s)}{R_{g_j,\beta_n}(X^{a_j})}.$$

We start by showing the consistency statement: from Lemma 3(i) and the continuity of the maps $t \mapsto \int_0^1 s^{-a_j t} dg_j(s)$, $1 \leq j \leq d$ at the point γ, we obtain

$$\frac{\widehat{R}^{AE}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} = \left[\frac{X_{\lceil n\beta_n\rceil,n}}{q(\beta_n)}\right]^a (1 + o_{\mathbb{P}}(1))$$

Write now $X_{\lceil n\beta_n\rceil,n} = U(Y_{\lceil n\beta_n\rceil,n})$ where Y has a standard Pareto distribution, and use Corollary 2.2.2 in de Haan and Ferreira (2006) together with the regular variation property of U to get

$$\frac{\widehat{R}^{AE}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} \xrightarrow{\mathbb{P}} 1.$$

To show the asymptotic normality of the estimator, use first the hypothesis on $X_{\lceil n\beta_n\rceil,n}$ and Lemma 3(ii) together with a Taylor expansion to get

$$\frac{\widehat{R}^{AE}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} = \frac{\int_0^1 s^{-a_j\widehat{\gamma}_n} dg_j(s)}{\int_0^1 s^{-a_j\gamma} dg_j(s)} \left[1 + \frac{a_j}{\sqrt{n(1-\beta_n)}} \left\{\Theta - \lambda \frac{\int_0^1 s^{-a_j\gamma} \frac{s^{-\rho}-1}{\rho} dg_j(s)}{\int_0^1 s^{-a_j\gamma} dg_j(s)} + o_{\mathbb{P}}(1)\right\}\right]. \quad (8)$$

Set then $\kappa(x) = e^x - 1 - x$ and notice that

$$\frac{\int_0^1 s^{-a_j\widehat{\gamma}_n} dg_j(s)}{\int_0^1 s^{-a_j\gamma} dg_j(s)} = 1 + a_j(\widehat{\gamma}_n - \gamma)\frac{\int_0^1 s^{-a_j\gamma}\log(1/s)dg_j(s)}{\int_0^1 s^{-a_j\gamma}dg_j(s)} + \frac{\int_0^1 s^{-a_j\gamma}\kappa(a_j(\widehat{\gamma}_n - \gamma)\log(1/s))dg_j(s)}{\int_0^1 s^{-a_j\gamma}dg_j(s)}.$$

A Taylor inequality for the exponential function at order 2 gives $|\kappa(x)| \leq x^2 e^{|x|}/2$ and thus

$$\left|\int_0^1 s^{-a_j\gamma}\kappa(a_j(\widehat{\gamma}_n - \gamma)\log(1/s))dg_j(s)\right| \leq \frac{a_j^2}{2}(\widehat{\gamma}_n - \gamma)^2 \int_0^1 s^{-a_j\gamma}\log^2(1/s)s^{-a_j|\widehat{\gamma}_n - \gamma|}dg_j(s).$$

Since $\int_0^1 s^{-a_j\gamma-\eta}dg_j(s) < \infty$, it follows by the $\sqrt{n(1-\beta_n)}-$consistency of $\widehat{\gamma}_n$ that

$$\left|\int_0^1 s^{-a_j\gamma}\kappa(a_j(\widehat{\gamma}_n - \gamma)\log(1/s))dg_j(s)\right| = \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)$$

and thus

$$\frac{\int_0^1 s^{-a_j\widehat{\gamma}_n} dg_j(s)}{\int_0^1 s^{-a_j\gamma} dg_j(s)} = 1 + \frac{a_j}{\sqrt{n(1-\beta_n)}}\frac{\int_0^1 s^{-a_j\gamma}\log(1/s)dg_j(s)}{\int_0^1 s^{-a_j\gamma}dg_j(s)}\Gamma + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right). \quad (9)$$

Combining (8) and (9) completes the proof. ■

Proof of Theorem 2. First, recall that for any $t \in \mathbb{R}$ we have $\lfloor t \rfloor + \lceil -t \rceil = 0$, where $\lfloor \cdot \rfloor$ denotes the floor function. Whence the equality

$$\widehat{R}_{g_j,\beta_n}(X^{a_j}) = \int_0^1 X^{a_j}_{n-\lfloor ls \rfloor,n}\, dg_j(s)$$

with $l = l(n) = n(1-\beta_n) \to \infty$. Clearly:

$$\forall s \in [0,1],\; X_{n-\lfloor(\lfloor l \rfloor+1)s\rfloor,n} \leq X_{n-\lfloor ls \rfloor,n} \leq X_{n-\lfloor \lfloor l \rfloor s\rfloor,n},$$

and thus it is enough to prove that, for any sequence of integers $k = k(n)$ such that $k(n)/l(n) \to 1$, we have:

$$\sqrt{k}\left(\frac{\int_0^1 X^{a_j}_{n-\lfloor ks \rfloor,n}\, dg_j(s)}{R_{g_j,\beta_n}(X^{a_j})} - 1\right)_{1\leq j\leq d} \xrightarrow{d} \mathcal{N}(0,V).$$

For any $a > 0$, let $U_a(x) := [U(x)]^a$ denote the left-continuous inverse of $1/(1-F_a)$, where F_a is the cdf of X^a. By Lemma 2:

$$\frac{R_{g_j,\beta_n}(X^{a_j})}{U_{a_j}(n/k)} = \int_0^1 \frac{U_{a_j}(n/ks)}{U_{a_j}(n/k)}dg_j(s) \to \int_0^1 s^{-a_j\gamma}dg_j(s).$$

It is therefore enough to prove that:

$$\sqrt{k}\left(\frac{\int_0^1 X^{a_j}_{n-\lfloor ks \rfloor,n}\, dg_j(s) - R_{g_j,\beta_n}(X^{a_j})}{U_{a_j}(n/k)}\right)_{1\leq j\leq d} \xrightarrow{d} \mathcal{N}(0,M) \quad (10)$$

where M is the $d \times d$ matrix with $(i,j)-$th entry

$$M_{i,j} = a_i a_j \gamma^2 \int_{[0,1]^2} \min(s,t)s^{-a_i\gamma-1}t^{-a_j\gamma-1}dg_i(s)dg_j(t).$$

Pick now $j \in \{1, \ldots, d\}$ and write

$$\begin{aligned}
\int_0^1 X_{n-\lfloor ks \rfloor, n}^{a_j} \, dg_j(s) - R_{g_j, \beta_n}(X^{a_j}) &= \zeta_{j,n} + \xi_{j,n} \qquad (11)\\
\text{with } \zeta_{j,n} &= \int_0^1 U_{a_j}(n/ks) \left(\frac{X_{n-\lfloor ks \rfloor, n}^{a_j}}{U_{a_j}(n/k)} - s^{-a_j\gamma} \right) s^{a_j\gamma} dg_j(s)\\
\text{and } \xi_{j,n} &= \int_0^1 U_{a_j}(n/ks) \frac{X_{n-\lfloor ks \rfloor, n}^{a_j}}{U_{a_j}(n/k)} \left(\frac{U_{a_j}(n/k)}{U_{a_j}(n/ks)} - s^{a_j\gamma} \right) dg_j(s).
\end{aligned}$$

According to Lemma 4, we have:

$$\sqrt{k} \left(\frac{\zeta_{j,n}}{U_{a_j}(n/k)} \right)_{1 \le j \le n} \xrightarrow{d} \mathcal{N}(\lambda C, M) \qquad (12)$$

where C is the column vector whose $j-$th entry is

$$C_j = a_j \int_0^1 \frac{s^{-\rho} - 1}{\rho} s^{-a_j\gamma} dg_j(s).$$

To examine the convergence of $\xi_{j,n}$, we note that according to (18), there exists Borel measurable functions $B_{a_1}, \ldots, B_{a_d}$, respectively asymptotically equivalent to $a_1 A_1, \ldots, a_d A_d$ and having constant sign, such that for any $\varepsilon > 0$:

$$\forall s \in (0, 1], \ \left| \frac{1}{B_{a_j}(n/ks)} \left(\frac{U_{a_j}(n/k)}{U_{a_j}(n/ks)} - s^{a_j\gamma} \right) - s^{a_j\gamma} \frac{s^\rho - 1}{\rho} \right| \le \varepsilon s^{a_j\gamma + \rho - \varepsilon} \qquad (13)$$

for n sufficiently large. Consider then the following decomposition of $\xi_{j,n}$:

$$\xi_{j,n} = \xi_{j,n}^{(1)} + \xi_{j,n}^{(2)} \qquad (14)$$

with

$$\begin{aligned}
\xi_{j,n}^{(1)} &= \int_0^1 U_{a_j}(n/ks) B_{a_j}(n/ks) \frac{X_{n-\lfloor ks \rfloor, n}^{a_j}}{U_{a_j}(n/k)} s^{a_j\gamma} \frac{s^\rho - 1}{\rho} dg_j(s),\\
\xi_{j,n}^{(2)} &= \int_0^1 U_{a_j}(n/ks) B_{a_j}(n/ks) \frac{X_{n-\lfloor ks \rfloor, n}^{a_j}}{U_{a_j}(n/k)} \left(\frac{1}{B_{a_j}(n/ks)} \left[\frac{U_{a_j}(n/k)}{U_{a_j}(n/ks)} - s^{a_j\gamma} \right] - s^{a_j\gamma} \frac{s^\rho - 1}{\rho} \right) dg_j(s).
\end{aligned}$$

Writing

$$\frac{X_{n-\lfloor ks \rfloor, n}^{a_j}}{U_{a_j}(n/k)} s^{a_j\gamma} = 1 + \left(\frac{X_{n-\lfloor ks \rfloor, n}^{a_j}}{U_{a_j}(n/k)} - s^{-a_j\gamma} \right) s^{a_j\gamma},$$

we get by Lemma 4:

$$\xi_{j,n}^{(1)} = \int_0^1 U_{a_j}(n/ks) B_{a_j}(n/ks) \frac{s^\rho - 1}{\rho} dg_j(s) + \mathrm{O}_{\mathbb{P}} \left(\frac{U_{a_j}(n/k) B_{a_j}(n/k)}{\sqrt{k}} \right).$$

Applying Lemma 2 to the regularly varying functions $t \mapsto U_{a_j}(t)|B_{a_j}(t)|$ and $t \mapsto t^{-\rho} U_{a_j}(t)|B_{a_j}(t)|$, which have respective regular variation indices $a_j\gamma + \rho$ and $a_j\gamma$, we get

$$\begin{aligned}
\sqrt{k} \frac{\xi_{j,n}^{(1)}}{U_{a_j}(n/k)} &= \sqrt{k} B_{a_j}(n/k) \int_0^1 s^{-a_j\gamma} \frac{1 - s^{-\rho}}{\rho} dg_j(s) + \mathrm{o}_{\mathbb{P}}(1)\\
&= -a_j \lambda \int_0^1 s^{-a_j\gamma} \frac{s^{-\rho} - 1}{\rho} dg_j(s) + \mathrm{o}_{\mathbb{P}}(1) = -\lambda C_j + \mathrm{o}_{\mathbb{P}}(1) \qquad (15)
\end{aligned}$$

since B_{a_j} is equivalent to $a_j A$. The quantity $\xi_{j,n}^{(2)}$ is controlled by applying inequality (13): for any $\varepsilon \in (0,\eta)$, we have for sufficiently large n that:

$$|\xi_{j,n}^{(2)}| \leq \varepsilon \int_0^1 U_{a_j}(n/ks)|B_{a_j}(n/ks)| \frac{X_{n-\lfloor ks \rfloor,n}^{a_j}}{U_{a_j}(n/k)} s^{a_j\gamma+\rho-\varepsilon} dg_j(s).$$

The ideas used to control $\xi_{j,n}^{(1)}$ yield for n large enough:

$$\sqrt{k}\left|\frac{\xi_{j,n}^{(2)}}{U_{a_j}(n/k)}\right| \leq \varepsilon a_j |\lambda| \int_0^1 s^{-a_j\gamma-\varepsilon} dg_j(s) + \mathrm{o}_{\mathbb{P}}(1) \leq \varepsilon a_j |\lambda| \int_0^1 s^{-a_j\gamma-\eta} dg_j(s) + \mathrm{o}_{\mathbb{P}}(1)$$

which, since ε is arbitrary, entails

$$\sqrt{k}\left|\frac{\xi_{j,n}^{(2)}}{U_{a_j}(n/k)}\right| = \mathrm{o}_{\mathbb{P}}(1). \tag{16}$$

Combining (14), (15) and (16) entails

$$\sqrt{k}\left(\frac{\xi_{j,n}}{U_{a_j}(n/k)}\right)_{1\leq j\leq d} \xrightarrow{\mathbb{P}} -\lambda C. \tag{17}$$

Combine finally (11), (12) and (17) to obtain (10): the proof is complete. ■

Proof of Theorem 3. We start by writing, for any j:

$$\frac{\widehat{R}_{g_j,\delta_n}^W(X^{a_j};\beta_n)}{R_{g_j,\delta_n}(X^{a_j})} = \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a_j(\widehat{\gamma}_n-\gamma)} \frac{\widehat{R}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})} \times \frac{R_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\delta_n}(X^{a_j})}\left(\frac{1-\beta_n}{1-\delta_n}\right)^{a_j\gamma}.$$

Recall that for any $a>0$, U_a satisfies condition $\mathcal{C}_2(a\gamma,\rho,aA)$ by Lemma 1. Taking logarithms and applying Lemma 5 with $Y = X^{a_j}$, we get

$$\log\left(\frac{\widehat{R}_{g_j,\delta_n}^W(X^{a_j};\beta_n)}{R_{g_j,\delta_n}(X^{a_j})}\right) = a_j(\widehat{\gamma}_n-\gamma)\log\left(\frac{1-\beta_n}{1-\delta_n}\right) + \log\left(\frac{\widehat{R}_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\beta_n}(X^{a_j})}\right) + \mathrm{O}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).$$

The $\sqrt{n(1-\beta_n)}-$relative consistency of $\widehat{R}_{g_j,\beta_n}(X^{a_j})$ entails

$$\log\left(\frac{\widehat{R}_{g_j,\delta_n}^W(X^{a_j};\beta_n)}{R_{g_j,\delta_n}(X^{a_j})}\right) = a_j(\widehat{\gamma}_n-\gamma)\log\left(\frac{1-\beta_n}{1-\delta_n}\right) + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).$$

Recall that $\log([1-\beta_n]/[1-\delta_n]) \to \infty$; a Taylor expansion and the hypothesis on $\widehat{\gamma}_n$ now make it clear that

$$\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}\left(\frac{\widehat{R}_{g_j,\delta_n}^W(X^{a_j};\beta_n)}{R_{g_j,\delta_n}(X^{a_j})} - 1\right) = a_j\xi(1+\mathrm{o}_{\mathbb{P}}(1))$$

which completes the proof. ■

Preliminary results and their proofs

The first result is a very useful fact which we shall use several times in our proofs.

Lemma 1. *Assume that condition $\mathcal{C}_2(\gamma,\rho,A)$ is satisfied. Pick $a>0$ and define $U_a(x) := [U(x)]^a$. Then U_a satisfies condition $\mathcal{C}_2(a\gamma,\rho,aA)$.*

Proof of Lemma 1. Pick $x>0$. The function U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$ which is equivalent to:

$$U(tx) = U(t)\left(x^\gamma + A(t)\left[\frac{x^\gamma(x^\rho-1)}{\rho} + \mathrm{o}(1)\right]\right) \quad \text{as} \quad t\to\infty.$$

Thus

$$U_a(tx) = U_a(t)x^{a\gamma}\left(1 + A(t)\left[\frac{x^\rho-1}{\rho} + \mathrm{o}(1)\right]\right)^a \quad \text{as} \quad t\to\infty.$$

Using a Taylor expansion and rearranging terms, we get:

$$U_a(tx) = U_a(t)\left(x^{a\gamma} + aA(t)\left[\frac{x^{a\gamma}(x^\rho-1)}{\rho} + \mathrm{o}(1)\right]\right) \quad \text{as} \quad t\to\infty,$$

which is the result. ■

This result yields an important inequality which is actually contained in Theorem 2.3.9 in de Haan and Ferreira (2006): for any $a>0$, one may find a Borel measurable function B_a, asymptotically equivalent to aA and having constant sign, such that for any $\varepsilon>0$, there is $t_0>0$ such that for $t, tx \geq t_0$:

$$\left|\frac{1}{B_a(t)}\left(\frac{U_a(tx)}{U_a(t)} - x^{a\gamma}\right) - x^{a\gamma}\frac{x^\rho-1}{\rho}\right| \leq \varepsilon x^{a\gamma+\rho}\max(x^\varepsilon, x^{-\varepsilon}). \tag{18}$$

The second preliminary result is a technical lemma on some integrals, which we shall use frequently in our proofs.

Lemma 2. *Let g be a nondecreasing right-continuous function on $[0,1]$. Assume that f is a Borel measurable regularly varying function with index $b\in\mathbb{R}$. If for some $\eta>0$:*

$$\int_0^1 s^{-b-\eta}dg(s) < \infty,$$

then for any $\delta\in\mathbb{R}$ such that $\delta<\eta$ and any continuous and bounded function φ on $(0,1]$ we have, provided (u_n) is a positive sequence tending to infinity:

$$\int_0^1 \frac{f(u_n/s)}{f(u_n)}s^{-\delta}\varphi(s)dg(s) \to \int_0^1 s^{-b-\delta}\varphi(s)dg(s).$$

Proof of Lemma 2. Pick $\delta<\eta$ and define $\varepsilon := (\eta-\delta)/2 > 0$, so that $\delta+\varepsilon<\eta$. We have

$$\left|\int_0^1 \frac{f(u_n/s)}{f(u_n)}s^{-\delta}\varphi(s)dg(s) - \int_0^1 s^{-b-\delta}\varphi(s)dg(s)\right| \leq \int_0^1 s^{b+\varepsilon}\left|\frac{f(u_n/s)}{f(u_n)} - s^{-b}\right| s^{-b-\delta-\varepsilon}|\varphi(s)|dg(s).$$

Notice that the function $f_1 : y \mapsto y^{-b-\varepsilon} f(y)$ is regularly varying with index $-\varepsilon < 0$. By a uniform convergence result for regularly varying functions (see *e.g.* Theorem 1.5.2 in Bingham *et al.*, 1987):

$$\sup_{0<s\le 1} s^{b+\varepsilon} \left| \frac{f(u_n/s)}{f(u_n)} - s^{-b} \right| = \sup_{0<s\le 1} \left| \frac{f_1(u_n/s)}{f_1(u_n)} - s^{\varepsilon} \right| = \sup_{t\ge 1} \left| \frac{f_1(u_n t)}{f_1(u_n)} - t^{-\varepsilon} \right| \to 0.$$

As a consequence

$$\left| \int_0^1 \frac{f(u_n/s)}{f(u_n)} s^{-\delta} \varphi(s) dg(s) - \int_0^1 s^{-b-\delta} \varphi(s) dg(s) \right| = \mathrm{O}\left(\sup_{0<s\le 1} s^{b+\varepsilon} \left| \frac{f(u_n/s)}{f(u_n)} - s^{-b} \right| \right)$$

and the right-hand side converges to 0. The proof is complete. ■

The third lemma gives an asymptotic expansion of a Wang DRM that is in particular the key to the construction of our first family of estimators.

Lemma 3. *Let g be a distortion function on $[0,1]$ and $a > 0$. Pick a sequence (β_n) such that $\beta_n \to 1$.*

(i) If U is regularly varying with index $\gamma > 0$ and there is $\eta > 0$ such that

$$\int_0^1 s^{-a\gamma-\eta} dg(s) < \infty$$

then we have that:

$$\frac{R_{g,\beta_n}(X^a)}{U_a([1-\beta_n]^{-1})} \to \int_0^1 s^{-a\gamma} dg(s) \quad \text{as} \quad n \to \infty.$$

(ii) If furthermore condition $\mathcal{C}_2(\gamma, \rho, A)$ is satisfied and $n(1-\beta_n) \to \infty$, $\sqrt{n(1-\beta_n)} A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$ then provided

$$\int_0^1 s^{-a\gamma-1/2-\eta} dg(s) < \infty$$

for some $\eta > 0$, we have that:

$$\frac{R_{g,\beta_n}(X^a)}{U_a([1-\beta_n]^{-1})} = \int_0^1 s^{-a\gamma} dg(s) + \frac{a\lambda}{\sqrt{n(1-\beta_n)}} \int_0^1 \frac{s^{-\rho}-1}{\rho} s^{-a\gamma} dg(s) + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right).$$

Proof of Lemma 3. The first statement is proven by applying Lemma 2:

$$\frac{R_{g,\beta_n}(X^a)}{U_a([1-\beta_n]^{-1})} = \int_0^1 \frac{U_a([1-\beta_n]^{-1}/s)}{U_a([1-\beta_n]^{-1})} dg(s) = \int_0^1 s^{-a\gamma} dg(s)(1+\mathrm{o}(1)). \tag{19}$$

To show the second statement, use (18) to get:

$$\begin{aligned} \frac{R_{g,\beta_n}(X^a)}{U_a([1-\beta_n]^{-1})} &- \int_0^1 \left(1 + B_a([1-\beta_n]^{-1}) \frac{s^{-\rho}-1}{\rho} \right) s^{-a\gamma} dg(s) \\ &= \mathrm{o}\left(B_a([1-\beta_n]^{-1}) \int_0^1 s^{-a\gamma-\rho-\eta} dg(s) \right). \end{aligned}$$

Rearranging and using the convergence $\sqrt{n(1-\beta_n)} B_a((1-\beta_n)^{-1}) \to a\lambda \in \mathbb{R}$, we obtain

$$\frac{R_{g,\beta_n}(X^a)}{U_a([1-\beta_n]^{-1})} = \int_0^1 s^{-a\gamma} dg(s) + \frac{a\lambda}{\sqrt{n(1-\beta_n)}} \int_0^1 \frac{s^{-\rho}-1}{\rho} s^{-a\gamma} dg(s) + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right) \tag{20}$$

which completes the proof. ■

The fourth lemma is the key to the proof of Theorem 2. It examines the asymptotic behavior of some weighted integrals of the empirical tail quantile process.

Lemma 4. *Assume that condition $\mathcal{C}_2(\gamma,\rho,A)$ is satisfied. Let $a_1,\ldots,a_d>0$, $f_1,\ldots,f_d$ be Borel measurable regularly varying functions with respective indices $b_j \le a_j\gamma$ and $g_1,\ldots,g_d$ be distortion functions. Assume that $k=k(n)\to\infty$, $k/n\to 0$, $\sqrt{k}A(n/k)\to\lambda\in\mathbb{R}$ and for some $\eta>0$:*

$$\forall j\in\{1,\ldots,d\},\ \int_0^1 s^{-a_j\gamma-1/2-\eta}dg_j(s)<\infty.$$

Pick $\delta_1,\ldots,\delta_d\in\mathbb{R}$ such that $\delta_j<(a_j\gamma-b_j)+\eta$, and set

$$I_{j,n}:=\frac{1}{f_j(n/k)}\int_0^1 f_j(n/ks)\sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor,n}^{a_j}}{U_{a_j}(n/k)}-s^{-a_j\gamma}\right)s^{a_j\gamma-\delta_j}dg_j(s).$$

Then we have:

$$(I_{1,n},\ldots,I_{d,n})\xrightarrow{d}\mathcal{N}(\lambda C,\Sigma)$$

with C being the column vector with $j-$th entry

$$C_j=a_j\int_0^1\frac{s^{-\rho}-1}{\rho}s^{-b_j-\delta_j}dg_j(s)$$

and Σ being the $d\times d$ matrix with $(i,j)-$th entry

$$\Sigma_{i,j}=a_ia_j\gamma^2\int_{[0,1]^2}\min(s,t)s^{-b_i-\delta_i-1}t^{-b_j-\delta_j-1}dg_i(s)dg_j(t).$$

Proof of Lemma 4. Define $\varepsilon:=\min_{1\le j\le d}(\eta-\delta_j)/2>0$, so that $\delta_j+\varepsilon<\eta$ for all j, and let $\varepsilon'>0$ be so small that

$$\forall j\in\{1,\ldots,d\},\ a_j\gamma+\frac{1}{2}+\frac{\varepsilon}{2}-\frac{1-\varepsilon'}{1+2\varepsilon'}\left(a_j\gamma+\frac{1}{2}+\varepsilon\right)<0. \tag{21}$$

Set $s_n=k^{-(1-\varepsilon')/(1+2\varepsilon')}$. Pick $j\in\{1,\ldots,d\}$ and use the triangle inequality to get:

$$\left|\frac{1}{f_j(n/k)}\int_0^{s_n} f_j(n/ks)\sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor,n}^{a_j}}{U_{a_j}(n/k)}-s^{-a_j\gamma}\right)s^{a_j\gamma-\delta_j}dg_j(s)\right|\le E_{j,n}^{(1)}+E_{j,n}^{(2)},$$

with

$$\begin{aligned} E_{j,n}^{(1)} &= \sqrt{k}\frac{X_{n,n}^{a_j}}{U_{a_j}(n/k)}\int_0^{s_n}\frac{f_j(n/ks)}{f_j(n/k)}s^{a_j\gamma-\delta_j}dg_j(s)\\ \text{and }\ E_{j,n}^{(2)} &= \sqrt{k}\int_0^{s_n}\frac{f_j(n/ks)}{f_j(n/k)}s^{-\delta_j}dg_j(s).\end{aligned}$$

Since the distribution of X is heavy-tailed it follows from Theorem 1.1.6, Theorem 1.2.1 and Lemma 1.2.9 in de Haan and Ferreira (2006) that $X_{n,n}=\mathrm{O}_{\mathbb{P}}(U(n))$. Thus

$$E_{j,n}^{(1)}=\mathrm{O}_{\mathbb{P}}\left(\sqrt{k}\frac{U_{a_j}(n)}{U_{a_j}(n/k)}\int_0^{s_n}\frac{f_j(n/ks)}{f_j(n/k)}s^{a_j\gamma-\delta_j}dg_j(s)\right)$$

Use now Potter bounds for U (see *e.g.* Theorem 1.5.6 in Bingham *et al.*, 1987) to get

$$\begin{aligned} E_{j,n}^{(1)} &= \mathrm{O}_{\mathbb{P}}\left(k^{a_j\gamma+1/2+\varepsilon/2}\int_0^{s_n}\frac{f_j(n/ks)}{f_j(n/k)}s^{a_j\gamma-\delta_j}dg_j(s)\right) \\ &= \mathrm{O}_{\mathbb{P}}\left(k^{a_j\gamma+1/2+\varepsilon/2}s_n^{a_j\gamma}\int_0^{s_n}\frac{f_j(n/ks)}{f_j(n/k)}s^{-\delta_j}dg_j(s)\right). \end{aligned}$$

Besides, note that

$$\int_0^{s_n}\frac{f_j(n/ks)}{f_j(n/k)}s^{-\delta_j}dg_j(s) \quad\leq\quad s_n^{1/2+\varepsilon}\int_0^{s_n}\frac{f_j(n/ks)}{f_j(n/k)}s^{-1/2-\delta_j-\varepsilon}dg_j(s)=\mathrm{o}\left(s_n^{1/2+\varepsilon}\right),$$

by Lemma 2. Thus

$$E_{j,n}^{(1)}=\mathrm{o}_{\mathbb{P}}(k^{a_j\gamma+1/2+\varepsilon/2}s_n^{a_j\gamma+1/2+\varepsilon})=\mathrm{o}_{\mathbb{P}}(1)\ \text{ and }\ E_{j,n}^{(2)}=\mathrm{o}_{\mathbb{P}}(k^{1/2+\varepsilon/2}s_n^{1/2+\varepsilon})=\mathrm{o}_{\mathbb{P}}(1)$$

by (21) and the fact that $s_n=k^{-(1-\varepsilon')/(1+2\varepsilon')}$. From this we deduce that for any $j\in\{1,\dots,d\}$:

$$I_{j,n}=\frac{1}{f_j(n/k)}\int_{s_n}^1 f_j(n/ks)\sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor,n}^{a_j}}{U_{a_j}(n/k)}-s^{-a_j\gamma}\right)s^{a_j\gamma-\delta_j}dg_j(s)+\mathrm{o}_{\mathbb{P}}(1).$$

Now, by Theorem 2.4.8 in de Haan and Ferreira (2006), we may find a Borel measurable function A_0 which has constant sign and is asymptotically equivalent to A at infinity such that for any $\varepsilon'>0$, we have

$$\sup_{0<s\leq 1}s^{\gamma+1/2+\varepsilon'}\left|\sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor,n}}{U(n/k)}-s^{-\gamma}\right)-\gamma s^{-\gamma-1}W_n(s)-\sqrt{k}A_0(n/k)s^{-\gamma}\frac{s^{-\rho}-1}{\rho}\right|\xrightarrow{\mathbb{P}}0 \quad (22)$$

where W_n is an appropriate sequence of standard Brownian motions. In other words:

$$\frac{X_{n-\lfloor ks\rfloor,n}}{U(n/k)}=s^{-\gamma}\left(1+\frac{1}{\sqrt{k}}\gamma s^{-1}W_n(s)+A_0(n/k)\frac{s^{-\rho}-1}{\rho}+\frac{1}{\sqrt{k}}s^{-1/2-\varepsilon'}\mathrm{o}_{\mathbb{P}}(1)\right)$$

with the $\mathrm{o}_{\mathbb{P}}(1)$ being uniform in $s\in(0,1]$. Now for any n, $W_n\stackrel{d}{=}W$ where W is a standard Brownian motion, and the random process W has continuous sample paths and $s^{-1/2+\varepsilon'}W(s)\to 0$ almost surely as $s\to 0$. Moreover, for $s\in[s_n,1]$, $s^{-1/2-\varepsilon'}\leq s_n^{-1/2-\varepsilon'}=\sqrt{k^{1-\varepsilon'}}=\mathrm{o}(\sqrt{k})$. Finally, $(s^{-\rho}-1)/\rho$ is bounded by a constant on $[s_n,1]$ when $\rho<0$, and is equal to $-\log(s)$ for $\rho=0$ and thus dominated by $s^{-1/2-\varepsilon'}$ in a neighborhood of 0. A Taylor expansion therefore yields:

$$\begin{aligned} \frac{X_{n-\lfloor ks\rfloor,n}^{a_j}}{U_{a_j}(n/k)} &= s^{-a_j\gamma}\left(1+\frac{1}{\sqrt{k}}\gamma s^{-1}W_n(s)+A_0(n/k)\frac{s^{-\rho}-1}{\rho}+\frac{1}{\sqrt{k}}s^{-1/2-\varepsilon'}\mathrm{o}_{\mathbb{P}}(1)\right)^{a_j} \\ &= s^{-a_j\gamma}\left(1+\frac{1}{\sqrt{k}}a_j\gamma s^{-1}W_n(s)+a_jA_0(n/k)\frac{s^{-\rho}-1}{\rho}+\frac{1}{\sqrt{k}}s^{-1/2-\varepsilon'}\mathrm{o}_{\mathbb{P}}(1)\right) \end{aligned}$$

where the $\mathrm{o}_{\mathbb{P}}(1)$ is uniform in $s\in[s_n,1]$. We deduce from this convergence that

$$\begin{aligned} I_{j,n} &= \zeta_{j,n}+\xi_{j,n}+\mathrm{o}_{\mathbb{P}}\left(\int_0^1\frac{f_j(n/ks)}{f_j(n/k)}s^{-1/2-\delta_j-\varepsilon'}dg_j(s)\right)+\mathrm{o}_{\mathbb{P}}(1) \\ \text{with}\quad \zeta_{j,n} &= a_j\gamma\int_0^1\frac{f_j(n/ks)}{f_j(n/k)}s^{-1-\delta_j}W_n(s)dg_j(s) \\ \text{and}\quad \xi_{j,n} &= a_j\sqrt{k}A_0(n/k)\int_0^1\frac{f_j(n/ks)}{f_j(n/k)}\frac{s^{-\rho}-1}{\rho}s^{-\delta_j}dg_j(s). \end{aligned}$$

By Lemma 2, we obtain

$$I_{j,n} = \zeta_{j,n} + \xi_{j,n} + \mathrm{o}_{\mathbb{P}}(1). \tag{23}$$

The bias term $\xi_{j,n}$ is controlled by applying Lemma 2:

$$\xi_{j,n} = a_j \lambda \int_0^1 \frac{s^{-\rho} - 1}{\rho} s^{-b_j - \delta_j} dg_j(s) + \mathrm{o}(1) \to \lambda C_j. \tag{24}$$

Notice now that

$$(\zeta_{1,n}, \ldots, \zeta_{d,n}) \stackrel{d}{=} \left(a_j \gamma \int_0^1 \frac{f_j(n/ks)}{f_j(n/k)} s^{-1-\delta_j} W(s) dg_j(s) \right)_{1 \le j \le d}$$

where W is a standard Brownian motion. Since W has continuous sample paths and $s^{-1/2+\varepsilon'} W(s) \to 0$ almost surely as $s \to 0$, we get by Lemma 2 that

$$\begin{aligned} (\zeta_{1,n}, \ldots, \zeta_{d,n}) & \stackrel{d}{=} \left(a_j \gamma \int_0^1 \frac{f_j(n/ks)}{f_j(n/k)} s^{-1/2-\delta_j-\varepsilon'} (s^{-1/2+\varepsilon'} W(s)) dg_j(s) \right)_{1 \le j \le d} \\ & \xrightarrow{d} \left(a_j \gamma \int_0^1 s^{-1-b_j-\delta_j} W(s) dg_j(s) \right)_{1 \le j \le d}. \end{aligned}$$

The entries of this random vector are almost surely finite. Let us recall that W is a centered Gaussian process with covariance function $\mathrm{Cov}(W(s), W(t)) = \min(s,t)$; consequently, for all $(u_1, \ldots, u_d) \in \mathbb{R}^d$, the random variable

$$\sum_{j=1}^d u_j a_j \gamma \int_0^1 s^{-1-b_j-\delta_j} W(s) dg_j(s)$$

is Gaussian centered and has variance

$$\gamma^2 \,\mathrm{Var} \left(\sum_{j=1}^d u_j a_j \int_0^1 s^{-1-b_j-\delta_j} W(s) dg_j(s) \right) = \sum_{i,j=1}^d u_i u_j \Sigma_{i,j} \tag{25}$$

by Fubini's theorem. It remains to combine Equations (23), (24) and (25), and to use the Cramér-Wold theorem to complete the proof. ∎

The fifth and final lemma shall be useful to control the bias term in Theorem 3.

Lemma 5. *Assume that Y_i, $i \ge 1$ are independent random variables with common cdf F_Y, such that the left-continuous inverse U_Y of $1/(1-F_Y)$ satisfies condition $\mathcal{C}_2(\gamma_Y, \rho_Y, A_Y)$, with $\rho_Y < 0$. Assume further that $\beta_n, \delta_n \to 1$, $n(1-\beta_n) \to \infty$, $(1-\delta_n)/(1-\beta_n) \to 0$ and $\sqrt{n(1-\beta_n)} A_Y((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R}$. Pick a distortion function g. If for some $\eta > 0$,*

$$\int_0^1 s^{-\gamma_Y - \eta} dg(s) < \infty,$$

then

$$\frac{R_{g,\delta_n}(Y)}{R_{g,\beta_n}(Y)} \left(\frac{1-\beta_n}{1-\delta_n} \right)^{-\gamma_Y} = 1 - \frac{\lambda/\rho_Y}{\sqrt{n(1-\beta_n)}} \frac{\int_0^1 s^{-\gamma_Y - \rho_Y} dg(s)}{\int_0^1 s^{-\gamma_Y} dg(s)} + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right).$$

Proof of Lemma 5. Set $k_1 = k_1(n) = n(1-\beta_n)$, $r_n = (1-\beta_n)/(1-\delta_n)$, $k_2 = k_2(n) = k_1/r_n$. Since for any $b \in (0,1)$,

$$R_{g,b}(Y) = \int_0^1 U_Y([(1-b)s]^{-1})dg(s),$$

we may write

$$R_{g,\delta_n}(Y) = r_n^{\gamma_Y} R_{g,\beta_n}(Y) + u_{1,n} + u_{2,n} \tag{26}$$

where

$$\begin{aligned} u_{1,n} &= r_n^{\gamma_Y} \frac{r_n^{\rho_Y}-1}{\rho_Y} \int_0^1 U_Y(n/k_1 s) A_0(n/k_1 s) dg(s) \\ \text{and} \quad u_{2,n} &= \int_0^1 U_Y(n/k_1 s) A_0(n/k_1 s) \left(\frac{1}{A_0(n/k_1 s)} \left[\frac{U_Y(n/k_2 s)}{U_Y(n/k_1 s)} - r_n^{\gamma_Y} \right] - r_n^{\gamma_Y} \frac{r_n^{\rho_Y}-1}{\rho_Y} \right) dg(s) \end{aligned}$$

with the notation of (22). By Lemma 2 and the convergence $\sqrt{k_1}A_0(n/k_1) \to \lambda$,

$$\begin{aligned} \sqrt{k_1}\frac{u_{1,n}}{U_Y(n/k_1)} &= \lambda r_n^{\gamma_Y} \frac{r_n^{\rho_Y}-1}{\rho_Y} \int_0^1 s^{-\gamma_Y-\rho_Y} dg(s) + \mathrm{o}(r_n^{\gamma_Y}) \\ &= -\frac{\lambda}{\rho_Y} r_n^{\gamma_Y} \int_0^1 s^{-\gamma_Y-\rho_Y} dg(s) + \mathrm{o}(r_n^{\gamma_Y}) \end{aligned} \tag{27}$$

because $r_n \to \infty$ and $\rho_Y < 0$. The sequence $u_{2,n}$ is controlled by using first inequality (18) and Lemma 2: for any $\varepsilon \in (0, -\rho_Y)$, we have if n is large enough,

$$\begin{aligned} \sqrt{k_1}\frac{|u_{2,n}|}{U_Y(n/k_1)} &\le \varepsilon r_n^{\gamma_Y+\rho_Y+\varepsilon} |\sqrt{k_1}A_0(n/k_1)| \int_0^1 \frac{U_Y(n/k_1 s)|A_0(n/k_1 s)|}{U_Y(n/k_1)|A_0(n/k_1)|} dg(s) \\ &= \varepsilon|\lambda| r_n^{\gamma_Y+\rho_Y+\varepsilon} \int_0^1 s^{-\gamma_Y-\rho_Y} dg(s) + \mathrm{o}(r_n^{\gamma_Y+\rho_Y+\varepsilon}) \\ &= \mathrm{o}(r_n^{\gamma_Y}). \end{aligned} \tag{28}$$

Combining (27) and (28) entails

$$\frac{\sqrt{k_1}}{U_Y(n/k_1)}(u_{1,n}+u_{2,n}) = -\frac{\lambda}{\rho_Y} r_n^{\gamma_Y} \int_0^1 s^{-\gamma_Y-\rho_Y} dg(s) + \mathrm{o}(r_n^{\gamma_Y}).$$

Use once more Lemma 2 to get

$$\frac{R_{g,\beta_n}(Y)}{U_Y(n/k_1)} = \int_0^1 \frac{U_Y(n/k_1 s)}{U_Y(n/k_1)} dg(s) \to \int_0^1 s^{-\gamma_Y} dg(s),$$

which yields

$$\frac{\sqrt{k_1}}{R_{g,\beta_n}(Y)}(u_{1,n}+u_{2,n}) = -\frac{\lambda}{\rho_Y} r_n^{\gamma_Y} \frac{\int_0^1 s^{-\gamma_Y-\rho_Y} dg(s)}{\int_0^1 s^{-\gamma_Y} dg(s)} + \mathrm{o}(r_n^{\gamma_Y}). \tag{29}$$

Combining (26) and (29) completes the proof. ∎

Risk measure $R_g(X)$	Distortion function g
VaR at level β	$g(x) = \mathbb{I}\{x \geq 1-\beta\}$ where $0 \leq \beta < 1$
TVaR above level β	$g(x) = \min\left\{\frac{x}{1-\beta}, 1\right\}$ where $0 \leq \beta < 1$
Proportional Hazard transform	$g(x) = x^\alpha$ where $0 < \alpha < 1$
Dual Power	$g(x) = 1 - (1-x)^{1/\alpha}$ where $0 < \alpha < 1$
MAXMINVAR	$g(x) = (1-(1-x)^\alpha)^{1/\alpha}$ where $0 < \alpha < 1$
MINMAXVAR	$g(x) = 1-(1-x^{1/\alpha})^\alpha$ where $0 < \alpha < 1$
Gini's principle	$g(x) = (1+\alpha)x - \alpha x^2$ where $0 < \alpha \leq 1$
Denneberg's absolute deviation	$g(x) = \begin{cases} (1+\alpha)x & \text{if} \quad 0 \leq x \leq 1/2 \\ \alpha + (1-\alpha)x & \text{if} \quad 1/2 \leq x \leq 1 \end{cases}$ where $0 < \alpha \leq 1$
Exponential transform	$g(x) = \begin{cases} (1-\exp(-rx))/(1-\exp(-r)) & \text{if} \quad r > 0 \\ x & \text{if} \quad r = 0 \end{cases}$
Logarithmic transform	$g(x) = \begin{cases} (\log(1+rx))/(\log(1+r)) & \text{if} \quad r > 0 \\ x & \text{if} \quad r = 0 \end{cases}$
Square-root transform	$g(x) = \begin{cases} (\sqrt{1+rx}-1)/(\sqrt{1+r}-1) & \text{if} \quad r > 0 \\ x & \text{if} \quad r = 0 \end{cases}$
S-inverse shaped transform	$g(x) = a\left(\frac{x^3}{6} - \frac{\delta}{2}x^2 + \left(\frac{\delta^2}{2} + \beta\right)x\right)$ where $a = \left(\frac{1}{6} - \frac{\delta}{2} + \frac{\delta^2}{2} + \beta\right)^{-1}$ with $0 \leq \delta \leq 1$ and $\beta \in \mathbb{R}$
Wang's transform	$g(x) = \Phi(\Phi^{-1}(x) + \Phi^{-1}(\alpha))$ where Φ is the standard Gaussian cdf and $0 \leq \alpha \leq 1$
Beta's transform	$g(x) = \int_0^x \frac{1}{\beta(a,b)} t^{a-1}(1-t)^{b-1} dt$ where $\beta(a,b)$ is the Beta function with parameters a, $b > 0$

Table 1: Some risk measures and their distortion functions.

Risk measure	Expression as a combination of $\mathrm{CTM}_a(\beta)$ and $\mathrm{VaR}(\beta)$
$\mathrm{CTE}(\beta)$	$\mathrm{CTM}_1(\beta)$
$\mathrm{CVaR}_\lambda(\beta)$	$\lambda\mathrm{VaR}(\beta) + (1-\lambda)\mathrm{CTM}_1(\beta)$ where $\lambda \in [0,1]$
$\mathrm{GlueVaR}_{\beta,\alpha}^{h_1,h_2}$	$\omega_1\mathrm{CTM}_1(\beta) + \omega_2\mathrm{CTM}_1(\alpha) + \omega_3\mathrm{VaR}(\alpha)$ where $\omega_1 = h_1 - \dfrac{(h_2-h_1)(1-\beta)}{\beta-\alpha}$, $\omega_2 = \dfrac{(h_2-h_1)(1-\alpha)}{\beta-\alpha}$ and $\omega_3 = 1-\omega_1-\omega_2 = 1-h_2$, with $h_1 \in [0,1]$, $h_2 \in [h_1,1]$ and $\alpha < \beta$
$\mathrm{SP}(\beta)$	$(1-\beta)(\mathrm{CTM}_1(\beta) - \mathrm{VaR}(\beta))$
$\mathrm{CTV}(\beta)$	$\mathrm{CTM}_2(\beta) - \mathrm{CTM}_1^2(\beta)$
$\mathrm{TSD}_\lambda(\beta)$	$\mathrm{CTM}_1(\beta) + \lambda\sqrt{\mathrm{CTM}_2(\beta) - \mathrm{CTM}_1^2(\beta)}$ where $\lambda \geq 0$
$\mathrm{CTS}(\beta)$	$\mathrm{CTM}_3(\beta)/(\mathrm{CTM}_2(\beta) - \mathrm{CTM}_1^2(\beta))^{3/2}$

Table 2: Link between the CTM and some risk measures when the cdf of X is continuous.

Value of γ	δ	Estimator	Fréchet		Burr $\rho=-1$		Burr $\rho=-2$	
			$n=100$	$n=300$	$n=100$	$n=300$	$n=100$	$n=300$
$\gamma=1/6$	0.99	AE	0.0325	0.0098	0.0374	0.0133	0.0291	0.0095
		PL	0.0317	0.0097	0.0357	0.0127	0.0286	0.0094
	0.995	AE	0.0457	0.0137	0.0540	0.0191	0.0401	0.0130
		PL	0.0446	0.0135	0.0518	0.0184	0.0395	0.0129
	0.999	AE	0.0891	0.0258	0.1115	0.0386	0.0752	0.0236
		PL	0.0871	0.0255	0.1073	0.0375	0.0741	0.0235
$\gamma=1/5$	0.99	AE	0.0519	0.0164	0.0627	0.0199	0.0472	0.0140
		PL	0.0502	0.0161	0.0588	0.0191	0.0461	0.0138
	0.995	AE	0.0739	0.0229	0.0915	0.0289	0.0657	0.0191
		PL	0.0717	0.0225	0.0862	0.0277	0.0643	0.0189
	0.999	AE	0.1500	0.0437	0.1952	0.0589	0.1266	0.0349
		PL	0.1461	0.0430	0.1850	0.0569	0.1239	0.0344
$\gamma=1/4$	0.99	AE	0.0973	0.0285	0.1028	0.0349	0.0834	0.0248
		PL	0.0900	0.0278	0.0944	0.0332	0.0835	0.0246
	0.995	AE	0.1411	0.0402	0.1515	0.0509	0.1190	0.0341
		PL	0.1305	0.0392	0.1395	0.0484	0.1202	0.0337
	0.999	AE	0.3039	0.0787	0.3350	0.1063	0.2492	0.0631
		PL	0.2807	0.0768	0.3102	0.1017	0.2604	0.0622

Table 3: Relative MSE for both estimators, case of the CTE.

Value of γ	δ	Estimator	Fréchet		Burr $\rho = -1$		Burr $\rho = -2$	
			$n = 100$	$n = 300$	$n = 100$	$n = 300$	$n = 100$	$n = 300$
$\gamma = 1/6$	0.99	AE	0.0487	0.0169	0.0629	0.0215	0.0458	0.0140
		PL	0.0448	0.0160	0.0549	0.0194	0.0443	0.0142
	0.995	AE	0.0653	0.0225	0.0866	0.0295	0.0609	0.0182
		PL	0.0597	0.0212	0.0757	0.0267	0.0586	0.0184
	0.999	AE	0.1177	0.0394	0.1658	0.0549	0.1084	0.0307
		PL	0.1073	0.0371	0.1456	0.0499	0.1033	0.0306
$\gamma = 1/5$	0.99	AE	0.0808	0.0261	0.0988	0.0336	0.0680	0.0211
		PL	0.0743	0.0256	0.0852	0.0304	0.0652	0.0217
	0.995	AE	0.1100	0.0349	0.1376	0.0463	0.0907	0.0276
		PL	0.1004	0.0339	0.1187	0.0417	0.0862	0.0281
	0.999	AE	0.2078	0.0620	0.2723	0.0870	0.1630	0.0468
		PL	0.1879	0.0598	0.2362	0.0785	0.1535	0.0468
$\gamma = 1/4$	0.99	AE	0.1558	0.0449	0.2175	0.0570	0.1327	0.0376
		PL	0.1397	0.0439	0.1707	0.0501	0.1252	0.0388
	0.995	AE	0.2182	0.0602	0.3161	0.0787	0.1818	0.0494
		PL	0.1932	0.0582	0.2471	0.0690	0.1698	0.0503
	0.999	AE	0.4485	0.1086	0.7089	0.1508	0.3561	0.0854
		PL	0.3899	0.1038	0.5482	0.1323	0.3279	0.0852

Table 4: Relative MSE for both estimators, case of the DP(1/3).

Value of γ	δ	Estimator	Fréchet		Burr $\rho = -1$		Burr $\rho = -2$	
			$n = 100$	$n = 300$	$n = 100$	$n = 300$	$n = 100$	$n = 300$
$\gamma = 1/6$	0.99	AE	0.0517	0.0162	0.0618	0.0207	0.0487	0.0141
		PL	0.0395	0.0145	0.0421	0.0157	0.0382	0.0133
	0.995	AE	0.0699	0.0216	0.0848	0.0282	0.0654	0.0184
		PL	0.0534	0.0191	0.0584	0.0215	0.0511	0.0172
	0.999	AE	0.1290	0.0383	0.1612	0.0523	0.1196	0.0311
		PL	0.0993	0.0334	0.1143	0.0406	0.0932	0.0286
$\gamma = 1/5$	0.99	AE	0.0800	0.0272	0.1116	0.0335	0.0756	0.0204
		PL	0.0579	0.0221	0.0670	0.0240	0.0583	0.0186
	0.995	AE	0.1083	0.0363	0.1549	0.0455	0.1010	0.0267
		PL	0.0780	0.0291	0.0941	0.0327	0.0776	0.0239
	0.999	AE	0.2020	0.0644	0.3067	0.0843	0.1829	0.0454
		PL	0.1457	0.0515	0.1916	0.0619	0.1401	0.0397
$\gamma = 1/4$	0.99	AE	0.1920	0.0461	0.2432	0.0678	0.1516	0.0405
		PL	0.1008	0.0347	0.1122	0.0438	0.0927	0.0355
	0.995	AE	0.2669	0.0613	0.3421	0.0921	0.2055	0.0529
		PL	0.1384	0.0453	0.1595	0.0594	0.1242	0.0452
	0.999	AE	0.5454	0.1088	0.7137	0.1727	0.3928	0.0906
		PL	0.2760	0.0796	0.3409	0.1136	0.2330	0.0748

Table 5: Relative MSE for both estimators, case of the PH(2/3).

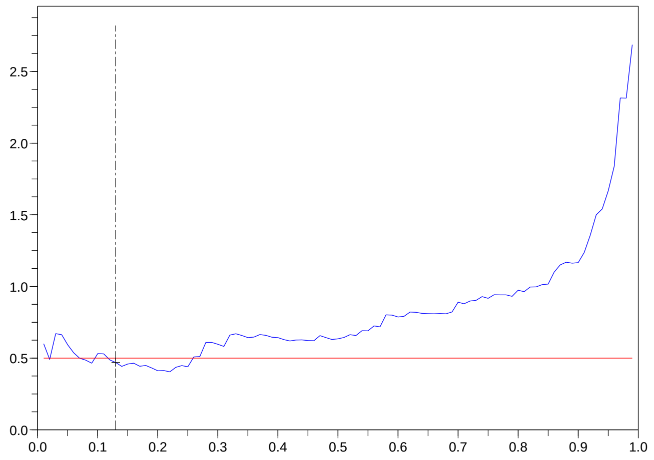

Figure 1: Choosing β on a random sample of $n = 100$ Burr observations with $\gamma = 1/2$ and $\rho = -1$; $x-$axis: $1 - \beta$. The choice procedure is conducted with $\beta_0 = 0.5$ and $h = 0.1$. The blue line is the Hill estimator; we obtain $\beta^* = 0.86$ and $\widehat{\gamma} = 0.475$.

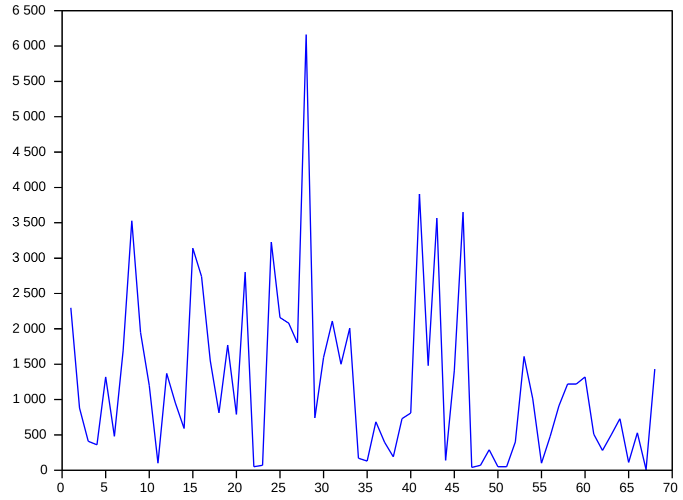

Figure 2: Poker data set: values of the consecutive swings of poker player Tom Dwan (absolute value of the aggregated results during alternative winning and losing streaks). Measurement unit: thousands of USD.

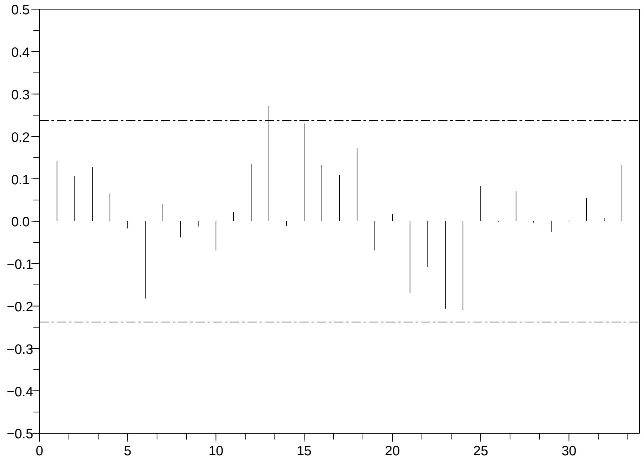

Figure 3: Poker data set: sample autocorrelation function until lag 34. Dashed line: 95% significance level.

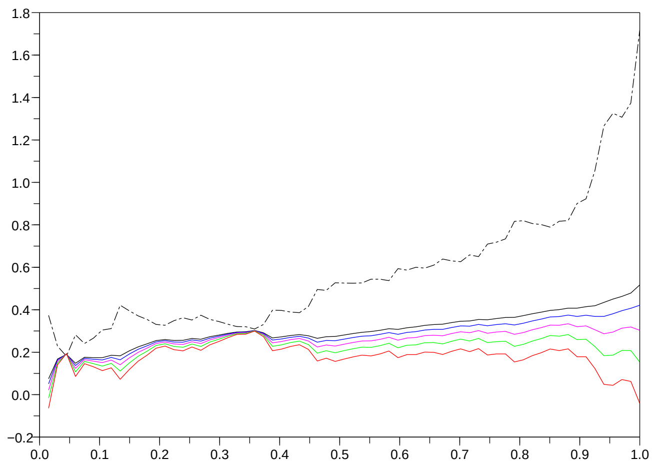

Figure 4: Poker data set, detrended data: Hill estimators; $x-$axis: $1-\beta$. Dashed line: standard Hill estimator, black line: estimator $\widehat{\gamma}_{\beta}^{RB}(1)$, blue line: estimator $\widehat{\gamma}_{\beta}^{RB}(3/4)$, purple line: estimator $\widehat{\gamma}_{\beta}^{RB}(1/2)$, green line: estimator $\widehat{\gamma}_{\beta}^{RB}(1/4)$, red line: estimator $\widehat{\gamma}_{\beta}^{RB}(0)$.

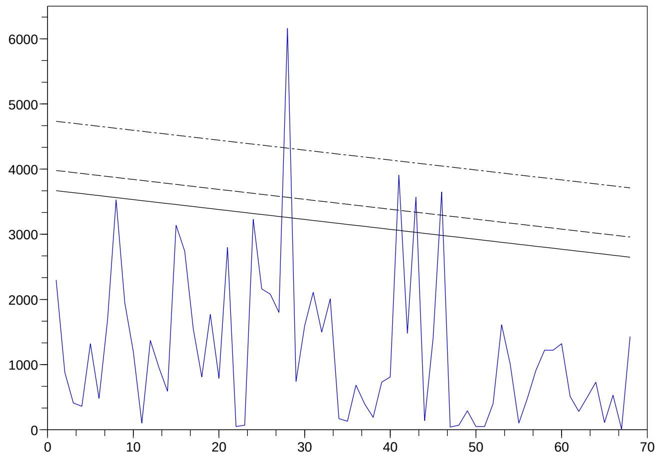

Figure 5: Poker data set (measurement unit: thousands of USD). Full line: 95% quantile line, dashed line: 97% quantile line, dashed-dotted line: 99% quantile line.

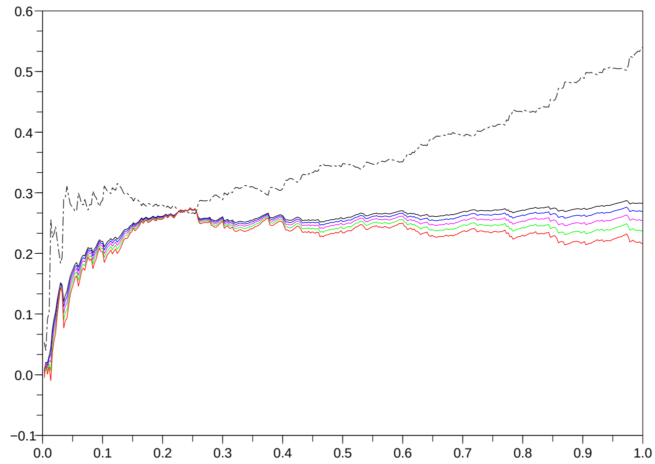

Figure 6: Secura Belgian Re data set: Hill estimators; $x-$axis: $1-\beta$. Dashed line: standard Hill estimator, black line: estimator $\widehat{\gamma}_{\beta}^{RB}(1)$, blue line: estimator $\widehat{\gamma}_{\beta}^{RB}(3/4)$, purple line: estimator $\widehat{\gamma}_{\beta}^{RB}(1/2)$, green line: estimator $\widehat{\gamma}_{\beta}^{RB}(1/4)$, red line: estimator $\widehat{\gamma}_{\beta}^{RB}(0)$.

Estimator $\widehat{\gamma}$	β^*	Estimate of γ
Standard Hill	0.75	0.351
Bias-reduced Hill, $\tau = 1$	0.794	0.260
Bias-reduced Hill, $\tau = 3/4$	0.912	0.167
Bias-reduced Hill, $\tau = 1/2$	0.912	0.158
Bias-reduced Hill, $\tau = 1/4$	0.912	0.146
Bias-reduced Hill, $\tau = 0$	0.853	0.118

Table 6: Poker data set: estimates of γ.

Estimator $\widehat{\gamma}$	β^*	Estimate of γ
Standard Hill	0.854	0.292
Bias-reduced Hill, $\tau = 1$	0.782	0.263
Bias-reduced Hill, $\tau = 3/4$	0.792	0.262
Bias-reduced Hill, $\tau = 1/2$	0.792	0.261
Bias-reduced Hill, $\tau = 1/4$	0.792	0.260
Bias-reduced Hill, $\tau = 0$	0.792	0.258

Table 7: Secura Belgian Re data set: estimates of γ.

δ	Estimator	$\widehat{\text{VaR}}$	$\widehat{\text{CTE}}$	$\widehat{\text{DP}(1/2)}$	$\widehat{\text{DP}(1/3)}$
0.95	AE	3684 $[3121, 4247]$	4373 $[3705, 5041]$	4747 $[4022, 5472]$	5010 $[4245, 5775]$
	PL	3684 $[3121, 4247]$	4911 $[4161, 5661]$	5450 $[4618, 6282]$	5805 $[4918, 6692]$
0.97	AE	3993 $[2832, 5154]$	4740 $[3362, 6118]$	5145 $[3649, 6641]$	5430 $[3851, 7009]$
	PL	3993 $[2832, 5154]$	5323 $[3775, 6871]$	5907 $[4190, 7624]$	6291 $[4462, 8120]$
0.99	AE	4748 $[1958, 7538]$	5636 $[2325, 8947]$	6118 $[2524, 9712]$	6457 $[2663, 10251]$
	PL	4748 $[1958, 7538]$	6329 $[2611, 10047]$	7023 $[2897, 11149]$	7480 $[3085, 11875]$

Table 8: Poker data set, detrended data: estimating some risk measures (measurement unit: thousands of USD). Between square brackets: asymptotic 95% confidence intervals.

δ	Estimator	$\widehat{\text{VaR}}$	$\widehat{\text{CTE}}$	$\widehat{\text{SP}}$
0.98	AE	4989 $[3505, 6473]$	6750 $[4742, 8758]$	35.220 $[24.744, 45.696]$
	PL	4989 $[3505, 6473]$	6864 $[4822, 8906]$	37.500 $[26.346, 48.654]$
0.99	AE	5978 $[3673, 8283]$	8087 $[4969, 11205]$	21.092 $[12.960, 29.224]$
	PL	5978 $[3673, 8283]$	8224 $[5053, 11395]$	22.459 $[13.800, 31.118]$
0.995	AE	7163 $[3770, 10556]$	9690 $[5100, 14280]$	12.636 $[6.6506, 18.621]$
	PL	7163 $[3770, 10556]$	9854 $[5186, 14522]$	13.455 $[7.0817, 19.828]$
0.999	AE	10899 $[3506, 18291]$	14744 $[4743, 24745]$	3.8452 $[1.2371, 6.4533]$
	PL	10899 $[3506, 18292]$	14993 $[4823, 25163]$	4.0944 $[1.3172, 6.8716]$

Table 9: Insurance data set: estimating some risk measures (measurement unit: thousands of Euros). Between square brackets: asymptotic 95% confidence intervals.