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. Special cases of the risk measures of interest include the extreme Value-at-Risk as well as the recently introduced extreme Conditional Tail Moment. 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 technique on a set of real data.

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 dikes 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 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, if the true quantile at the chosen level exceeds the range covered by the available data, and this makes beyond-the-sample estimation impossible. Many of the aforementioned problems can actually be 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. The estimation of the tail index, an excellent overview of which is given in the recent monograph by de Haan and Ferreira (2006), is therefore the first step to understand the extremes of a random variable whose distribution is heavy-tailed. Once the tail index has been estimated, it becomes possible to estimate an extreme quantile, or Value-at-Risk (VaR) as it is known in the actuarial science literature, at an arbitrarily high level, for instance using the extrapolation method of Weissman (1978).

Of course, the estimation of a single extreme quantile 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 a 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) or Stop-loss Premium reinsurance risk measure; 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 framework.

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 like the Wang transform (Wang, 2000) which is very popular in finance 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.

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 a real dataset in Section 5. The proofs of our results 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. Note that if q is the quantile function of X, namely $q(\alpha) = \inf\{t \in \mathbb{R} \,|\, F(t) \geq \alpha\}$ for all $\alpha \in (0,1)$, a change of variables and an integration by parts make it possible to rewrite $R_g(X)$, when it is finite, as a Lebesgue-Stieltjes integral:

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

(the right-continuity of g makes this integral well-defined indeed). 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 $100\beta\%$ of cases, obtained by setting $g(x) = \mathbb{I}\{x \geq 1-\beta\}$, with $\mathbb{I}\{\cdot\}$ being the indicator function;

- the Tail Value-at-Risk TVaR(β) in the worst $100\beta\%$ cases, namely the average of all quantiles exceeding VaR(β), 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 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 spectral 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

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

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 or Conditional Tail Variance, which can then be obtained by combining a finite number of CTMs and the VaR; see El Methni *et al.* (2014) for further details.

We close this section by mentioning that according to Wirch and Hardy (2002), a DRM is a coherent risk measure (see Artzner *et al.*, 1999) if and only if the distortion function g is concave. The coherency of a risk measure is 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. 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

was already noted several times in the recent literature. More broadly, this result 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.

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 $g_\beta(y)$ which is defined by:

$$g_\beta(y) := \begin{cases} g\left(\dfrac{y}{1-\beta}\right) & \text{if} \quad y \leq 1-\beta \\ 1 & \text{otherwise.} \end{cases}$$

Such a function, which is deduced from g by a simple 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. Pick also a strictly increasing continuously differentiable function h and set

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

The distortion measure dg_β of this risk measure is concentrated on $[0, 1-\beta]$, so that it only takes into account those (high) quantiles of X whose order lies in $[\beta, 1]$. It is actually straightforward to obtain

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

which makes the interpretation of this risk measure easier: the high quantiles of X are transformed via the function h and then weighted using the function g. A further way to understand the action of g_β is the following:

Proposition 1. *Let $\beta \in [0,1)$. If the cdf of X is continuous, one has*

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

In many cases, the Wang DRM $R_{g,\beta}(h(X))$ is thus actually obtained by computing the Wang DRM of $h(X)$, with distortion function g, given that X exceeds $q(\beta)$. 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.

3.2 A first estimator

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 random variables having cdf F, and let (β_n) be a nondecreasing sequence of real numbers belonging to $(0, 1)$, which converges to 1. Denote by $\widehat{F}_n$ the empirical cdf related to this sample and by $\widehat{q}_n$ the related 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 $\lceil \cdot \rceil$ denotes the ceiling function. The empirical counterpart of $R_{g,\beta_n}(h(X))$ is then

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

In the case when $n(1 - \beta_n)$ is actually a positive integer, which is fairly common in practice (see Sections 4 and 5), this estimator can be conveniently rewritten as an L-statistic, namely:

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

or equivalently

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

If g is further assumed to be continuous on $[0, 1]$, this L-statistic takes the simpler form

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

Our aim is to examine the asymptotic properties of this estimator. Our framework is that of regular variation: 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 $t \to \infty$. We shall assume in this paper that the distribution of X is heavy-tailed, or in other words, that $1 - F$ is regularly varying with index $-1/\gamma < 0$, the parameter γ being the so-called tail index of the cdf F. It is in fact convenient to use an equivalent assumption on the left-continuous inverse U of $1/(1 - F)$, defined by $U(y) = \inf\{t \in \mathbb{R} \,|\, 1/(1-F(t)) \geq y\} = q(1-y^{-1})$. More precisely, we shall assume that U is regularly varying with index γ, and use the following second-order condition on U (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. Most standard examples in extreme-value theory satisfy this condition (see *e.g.* the examples pp.61–62 in de Haan and Ferreira, 2006).

In all what follows, we also assume that the function h is a positive power of x. This choice allows us to consider estimators of a large class of risk measures of X, including the aforementioned CTM. Our first result is the following:

Theorem 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 $a_1, \ldots, a_d > 0$. 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:

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

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

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

Let us emphasize 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, will be tackled in the next section by the introduction of an estimator adapted to the extreme-value framework.

For now, let us draw a useful consequence from Theorem 1. For $b \in \mathbb{R}$, we introduce the class of functions

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

Roughly speaking, the classes $\mathcal{E}_b$ can be considered as the spaces of those functions g which are continuously differentiable on $[0,1]$ and whose first derivative behaves like a power of s in a neighborhood of 0. Especially, any polynomial function belongs to $\mathcal{E}_0([0,1])$. 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) greatly simplifies such an integrability hypothesis. 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:

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

with V as in Theorem 1.

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 Theorem 1 and Corollary 1 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 that paper.

3.3 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 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 which depends on the unknown tail index γ. 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 define:

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

where $\widehat{\gamma}$ is a consistent estimator of γ. This can be seen as a Weissman-type estimator of $R_{g,\delta_n}(X^a)$ (see Weissman, 1978); in fact, we exactly recover Weissman's estimator of an extreme quantile by setting $a = 1$ and $g(x) = 0$ if $x < 1$. Besides, taking $g(x) = x$, we also recover 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 the straightforward empirical estimator is known to be consistent, and then multiplied by an extrapolation factor which depends on an external estimate of the tail index γ.

Our second main result examines the asymptotic distribution of this estimator.

Theorem 2. *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 $\sqrt{n(1-\beta_n)}(\widehat{\gamma}-\gamma) \xrightarrow{d} \xi$ *then*

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

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 Theorem 2 of El Methni *et al.* (2014). 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 estimator is exactly determined by the estimator of γ which is used. Many estimators of γ can be considered: we mention the very popular Hill estimator (1975), the Pickands estimator (1975), the maximum likelihood estimator (Smith, 1987 and Drees *et al.*, 2004) and probability-weighted moment estimators (Hosking and Wallis, 1987, Diebolt *et al.*, 2007), among others. Such estimators 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 $\rho\leq 0$).

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 $c:=1/\alpha$ is a positive integer, the related DRM is the expectation of $\max(X_1,\ldots,X_c)$ for independent copies $X_1,\ldots,X_c$ of X;

- the Proportional Hazard (PH) transform function $g(x) = x^\alpha$ with $\alpha \in (0,1)$, which gives higher weight to large quantiles and is such that $g'(s) \uparrow \infty$ as $s \downarrow 0$.

Theorem 1 shall thus be valid for $\gamma < 1/2$ for either of the two first distortion functions, and $\gamma < \alpha - 1/2$ for the Proportional Hazard transform (note that in this case the related Wang DRM is actually well-defined and finite for $\gamma < \alpha$). In some sense, a suitable choice of α in the PH transform case will allow us to check if our technique is robust to a violation of the integrability condition of Theorems 1 and 2.

Our estimator 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 choice 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 (2014). In many practical cases though, the analysis of a dataset 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 idea here is to suggest an automatic such choice. We work with the popular Hill estimator (Hill, 1975):

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

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 similar to others in the extreme value literature (see *e.g.* Resnick and Stărică, 1997, Drees *et al.*, 2000 or de Sousa and Michailidis, 2004); closely related procedures when there

is random covariate information can be found in Stupfler (2013) and Gardes and Stupfler (2014). An illustration of this technique on a simulated dataset 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):

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

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 3–7. It appears on these examples that the performances of our estimators are globally satisfactory. Results deteriorate when γ increases: a possible explanation lies in the fact that the asymptotic distribution of our estimator is essentially that of $\widehat{\gamma}$ by Theorem 2, 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 estimator seems to be at least somewhat robust to a violation of the integrability condition in Theorems 1–2, as can be seen on the example of the PH(2/3) risk measure, although in this example, the results deteriorate as the rate of convergence of $g'(s)$ to ∞ as $s \downarrow 0$ increases. This likely comes from the fact that an increasing such rate amplifies the error made by the plug-in extreme quantile estimator, all the more so as the latter error itself increases when estimating quantiles whose order is very close to 1.

5 Real data application

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, 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 a fraction of the square-root of the total sample size, we believe that the separate datasets are too small to carry out an analysis which is interpretable from the extreme-value point of view.

Since we work on time series data, there are particular concerns about independence and stationarity. These hypotheses are checked using the turning point test (see Kendall, 1973) 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 0.1 for a trend parameter of 0.103 and a lag parameter of 1 in the Newey-West variance estimator. This trend parameter being very small relatively to the values of our time series (its minimum, median and maximum are respectively 10, 895 and 6162), the KPSS test indicates that the data is essentially stationary indeed. 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 this random 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),

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

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 makes 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 8 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 $\widehat{\gamma} = 0.196$ for $\beta^* = 0.735$ and $\tau = 1/4$; especially, ρ is estimated by $\widehat{\rho} = -0.790$. Finally, Table 9 gives estimates of some risk measures for this dataset. From these results, it appears in particular that the maximal value in this dataset, corresponding to a losing streak costing 6.162 million USD, exceeds both our estimates of the 99% quantile and of the CTE in the 1% highest cases. It is also of the same order of magnitude as our estimate of the Dual Power(1/2) (resp. Dual Power(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 above the 99% quantile. In our opinion, this losing streak can thus be regarded as an extreme period of loss.

References

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.

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.

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.

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.

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.

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

Gomes, M.I. & Guillou, A. (2014). Extreme value theory and statistics of univariate extremes: a review. *Int. Stat. Rev.*, to appear.

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. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339–349.

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.

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.

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

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.

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.

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.

Contact information: Gilles STUPFLER, Aix Marseille Université, GREQAM & CNRS
Centre de La Vieille Charité, 2 rue de la Charité, 13002 Marseille, France
E-mail: `gilles.stupfler@univ-amu.fr`

Appendix

Proofs of the main results

Proof of Proposition 1. On the one hand, we have that

$$R_{g,\beta}(h(X)) = \int_0^1 h \circ q(1-\alpha) dg_\beta(\alpha) = \int_0^1 h \circ q(1-(1-\beta)s) dg(s). \tag{2}$$

On the other hand, it holds that for any x:

$$F_\beta(x) := \mathbb{P}(X \leq x | X > q(\beta)) = \begin{cases} 0 & \text{if } \ x \leq q(\beta), \\ \dfrac{F(x)-\beta}{1-\beta} & \text{if } \ x > q(\beta) \end{cases}$$

since F is continuous. The related quantile function is defined by

$$\forall \alpha \in (0,1),\ q_\beta(\alpha) = \inf\{t \in \mathbb{R} \,|\, F_\beta(t) \geq \alpha\} = \inf\{t \in \mathbb{R} \,|\, F(t) \geq \beta + (1-\beta)\alpha\} = q(\beta + (1-\beta)\alpha)$$

which yields

$$R_g(h(X_\beta)) = \int_0^1 h \circ q_\beta(1-\alpha) dg(\alpha) = \int_0^1 h \circ q(1-(1-\beta)\alpha) dg(\alpha). \tag{3}$$

Combining (2) and (3) completes the proof. ■

Proof of Theorem 1. 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:

$$\sqrt{k}\left(\frac{\int_0^1 X^{a_j}_{n-\lfloor ks \rfloor,n}\, dg_j(s) - R_{g_j,\beta_n}(X^{a_j})}{U_{a_j}(n/k)} \right)_{1 \leq j \leq d} \xrightarrow{d} \mathcal{N}(0,M) \tag{4}$$

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

$$\begin{aligned} \int_0^1 X^{a_j}_{n-\lfloor ks\rfloor,n}\, dg_j(s) - R_{g_j,\beta_n}(X^{a_j}) &= \zeta_{j,n} + \xi_{j,n} \qquad (5)\\ \text{with } \zeta_{j,n} &= \int_0^1 U_{a_j}(n/ks)\left(\frac{X^{a_j}_{n-\lfloor ks\rfloor,n}}{U_{a_j}(n/k)} - s^{-a_j\gamma}\right) s^{a_j\gamma} dg_j(s)\\ \text{and } \xi_{j,n} &= \int_0^1 U_{a_j}(n/ks)\frac{X^{a_j}_{n-\lfloor ks\rfloor,n}}{U_{a_j}(n/k)}\left(\frac{U_{a_j}(n/k)}{U_{a_j}(n/ks)} - s^{a_j\gamma}\right) dg_j(s). \end{aligned}$$

According to Lemma 3, we have:

$$\left(\sqrt{k}\frac{\zeta_{j,n}}{U_{a_j}(n/k)}\right)_{1\le j\le n} \xrightarrow{d} \mathcal{N}(\lambda C, M) \qquad (6)$$

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 (12), there exists Borel measurable functions $B_{a_1},\ldots,B_{a_d}$, respectively asymptotically equivalent to $a_1A_1,\ldots,a_dA_d$ and having constant sign, such that for any $\varepsilon > 0$:

$$\forall s \in (0,1],\ \left|\frac{1}{B_{a_j}(n/ks)}\left(\frac{U_{a_j}(n/k)}{U_{a_j}(n/ks)} - s^{a_j\gamma}\right) - s^{a_j\gamma}\frac{s^\rho-1}{\rho}\right| \le \varepsilon s^{a_j\gamma+\rho-\varepsilon} \qquad (7)$$

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

$$\xi_{j,n} = \xi^{(1)}_{j,n} + \xi^{(2)}_{j,n} \qquad (8)$$

with

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

Writing

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

we get by Lemma 3:

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

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

$$\begin{aligned}\sqrt{k}\frac{\xi_{j,n}^{(1)}}{U_{a_j}(n/k)} &= \sqrt{k}B_{a_j}(n/k)\int_0^1 s^{-a_j\gamma}\frac{1-s^{-\rho}}{\rho}dg_j(s) + \mathrm{o}_{\mathbb{P}}(1) \\ &= -a_j\lambda\int_0^1 s^{-a_j\gamma}\frac{s^{-\rho}-1}{\rho}dg_j(s) + \mathrm{o}_{\mathbb{P}}(1) = -\lambda C_j + \mathrm{o}_{\mathbb{P}}(1) \qquad (9)\end{aligned}$$

since B_{a_j} is equivalent to a_jA. The quantity $\xi_{j,n}^{(2)}$ is controlled by applying inequality (7): 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

$$\sqrt{k}\left|\frac{\xi_{j,n}^{(2)}}{U_{a_j}(n/k)}\right| = \mathrm{o}_{\mathbb{P}}(1). \qquad (10)$$

Combining (8), (9) and (10) entails

$$\sqrt{k}\left(\frac{\xi_{j,n}}{U_{a_j}(n/k)}\right)_{1\leq j\leq d} \xrightarrow{\mathbb{P}} -\lambda C. \qquad (11)$$

Combine finally (5), (6) and (11) to obtain (4): the proof is complete. ■

Proof of Theorem 2. We start by writing, for any j:

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

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 4 with $Y = X^a$, we get

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

A use of Theorem 1, together with the delta-method, entails

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

Recall that $\log([1-\beta_n]/[1-\delta_n]) \to \infty$; a Taylor expansion and the hypothesis on $\widehat{\gamma}$ 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 $t, tx \geq t_0$:

$$\left|\frac{1}{B_a(t)}\left(\frac{U_a(tx)}{U_a(t)} - x^{a\gamma}\right) - x^{a\gamma}\frac{x^\rho - 1}{\rho}\right| \leq \varepsilon x^{a\gamma+\rho}\max(x^\varepsilon, x^{-\varepsilon}). \tag{12}$$

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 $k = k(n) \to \infty$ and $k/n \to 0$:

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

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

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

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):

$$\sup_{0<s\leq 1} s^{b+\varepsilon} \left| \frac{f(n/ks)}{f(n/k)} - s^{-b} \right| = \sup_{0<s\leq 1} \left| \frac{f_1(n/ks)}{f_1(n/k)} - s^{\varepsilon} \right| = \sup_{t\geq 1} \left| \frac{f_1(tn/k)}{f_1(n/k)} - t^{-\varepsilon} \right| \to 0.$$

As a consequence

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

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

The third lemma is the key to the proof of Theorem 1. It examines the asymptotic behavior of some weighted integrals of the empirical tail quantile process.

Lemma 3. *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 $b_j \leq a_j\gamma$ 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

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

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

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

Proof of Lemma 3. Define $\varepsilon := \min_{1\leq j\leq d}(\eta - \delta_j)/2 > 0$, so that $\delta_j + \varepsilon < \eta$ for all j, and let $\varepsilon' > 0$ be so small that

$$\forall j \in \{1, \ldots, d\}, \ a_j\gamma + \frac{1}{2} + \frac{\varepsilon}{2} - \frac{1-\varepsilon'}{1+2\varepsilon'} \left(a_j\gamma + \frac{1}{2} + \varepsilon \right) < 0. \tag{13}$$

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

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

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

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

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

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

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

$$\sup_{0<s\leq 1} s^{\gamma+1/2+\varepsilon'}\left|\sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor,n}}{U(n/k)} - s^{-\gamma}\right) - \gamma s^{-\gamma-1}W_n(s) - \sqrt{k}A_0(n/k)s^{-\gamma}\frac{s^{-\rho}-1}{\rho}\right| \xrightarrow{\mathbb{P}} 0 \quad (14)$$

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, $W_n \overset{d}{=} W$ 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:

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

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

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

By Lemma 2, we obtain

$$I_{j,n}=\zeta_{j,n}+\xi_{j,n}+\mathrm{o}_{\mathbb{P}}(1). \tag{15}$$

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

$$\xi_{j,n}=a_j\lambda\int_0^1\frac{s^{-\rho}-1}{\rho}s^{-b_j-\delta_j}dg_j(s)(1+\mathrm{o}(1))\to\lambda C_j. \tag{16}$$

Notice now that

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

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

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

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

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

is Gaussian centered and has variance

$$\gamma^2\,\mathrm{Var}\left(\sum_{j=1}^d u_ja_j\int_0^1 s^{-1-b_j-\delta_j}W(s)dg_j(s)\right)=\sum_{i,j=1}^d u_iu_j\Sigma_{i,j} \tag{17}$$

by Fubini's theorem. It remains to combine Equations (15), (16) and (17), and to use the Cramér-Wold theorem to complete the proof. ■

The fourth and final lemma shall be useful to control the bias term in Theorem 2.

Lemma 4. *Assume that Y_i, $i \geq 1$ 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

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

Proof of Lemma 4. 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

$$R_{g,\delta_n}(Y) = r_n^{\gamma_Y} R_{g,\beta_n}(Y) + u_{1,n} + u_{2,n} \tag{18}$$

where

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

with the notation of (14). By Lemma 2 and the convergence $\sqrt{k_1} A_0(n/k_1) \to \lambda$,

$$\begin{aligned} \sqrt{k_1} \frac{u_{1,n}}{U_Y(n/k_1)} &= \lambda r_n^{\gamma_Y} \frac{r_n^{\rho_Y} - 1}{\rho_Y} \int_0^1 s^{-\gamma_Y - \rho_Y} dg(s)(1 + \mathrm{o}(1)) \\ &= -\frac{\lambda}{\rho_Y} r_n^{\gamma_Y} \int_0^1 s^{-\gamma_Y - \rho_Y} dg(s)(1 + \mathrm{o}(1)) \end{aligned} \tag{19}$$

because $r_n \to \infty$ and $\rho_Y < 0$. The sequence $u_{2,n}$ is controlled by using first inequality (12) and Lemma 2: for any $\varepsilon \in (0, -\rho_Y)$, we have if n is large enough,

$$\begin{aligned} \sqrt{k_1} \frac{|u_{2,n}|}{U_Y(n/k_1)} &\leq \varepsilon r_n^{\gamma_Y + \rho_Y + \varepsilon} |\sqrt{k_1} A_0(n/k_1)| \int_0^1 \frac{U_Y(n/k_1 s)|A_0(n/k_1 s)|}{U_Y(n/k_1)|A_0(n/k_1)|} dg(s) \\ &= \varepsilon |\lambda| r_n^{\gamma_Y + \rho_Y + \varepsilon} \int_0^1 s^{-\gamma_Y - \rho_Y} dg(s)(1 + \mathrm{o}(1)) \\ &= \mathrm{o}(r_n^{\gamma_Y}). \end{aligned} \tag{20}$$

Combining (19) and (20) entails

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

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

$$\frac{\sqrt{k_1}}{R_{g,\beta_n}(Y)}(u_{1,n} + u_{2,n}) = -\frac{\lambda}{\rho_Y} r_n^{\gamma_Y} \frac{\int_0^1 s^{-\gamma_Y - \rho_Y} dg(s)}{\int_0^1 s^{-\gamma_Y} dg(s)}(1 + \mathrm{o}(1)). \tag{21}$$

Combining (18) and (21) 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 $0 \leq \alpha \leq 1$
Dual Power	$g(x) = 1-(1-x)^{1/\alpha}$ where $0 \leq \alpha \leq 1$
Gini's principle	$g(x) = (1+\alpha)x - \alpha x^2$ where $0 \leq \alpha \leq 1$
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 $0 \leq \alpha \leq 1$
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 $a = \left(\frac{1}{6} - \frac{\delta}{2} + \frac{\delta^2}{2} + \beta\right)$ 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 $\omega_1 = h_1 - \frac{(h_2-h_1)(1-\beta)}{\beta-\alpha}$, $\omega_2 = \frac{(h_2-h_1)(1-\alpha)}{\beta-\alpha}$ and $\omega_3 = 1-\omega_1-\omega_2 = 1-h_2$, with $h_1 \in [0,1]$, $h_2 \in [h_1,1]$ and $\alpha < \beta$
$\mathrm{SP}(\beta)$	$(1-\beta)(\mathrm{CTM}_1(\beta) - \mathrm{VaR}(\beta))$
$\mathrm{CTV}(\beta)$	$\mathrm{CTM}_2(\beta) - \mathrm{CTM}_1^2(\beta)$
$\mathrm{CTS}(\beta)$	$\mathrm{CTM}_3(\beta)/(\mathrm{CTM}_2(\beta) - \mathrm{CTM}_1^2(\beta))^{3/2}$

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

MSE		Fréchet		Burr $\rho = -1$		Burr $\rho = -2$	
CTE	δ	$n = 100$	$n = 300$	$n = 100$	$n = 300$	$n = 100$	$n = 300$
	0.99	0.0307	0.0106	0.0366	0.0129	0.0294	0.00967
$\gamma = 1/6$	0.995	0.0432	0.0148	0.0532	0.0188	0.0407	0.0132
	0.999	0.0843	0.0282	0.111	0.0382	0.0767	0.0240
	0.99	0.0901	0.0260	0.105	0.0337	0.0762	0.0241
$\gamma = 1/4$	0.995	0.133	0.0365	0.158	0.0491	0.107	0.0331
	0.999	0.302	0.0711	0.365	0.103	0.216	0.0612
	0.99	0.210	0.0594	0.231	0.0750	0.213	0.0481
$\gamma = 1/3$	0.995	0.315	0.0840	0.359	0.110	0.317	0.0660
	0.999	0.750	0.169	0.918	0.240	0.763	0.125

Table 3: Relative MSE, case of the CTE.

MSE		Fréchet		Burr $\rho = -1$		Burr $\rho = -2$	
DP(1/2)	δ	$n = 100$	$n = 300$	$n = 100$	$n = 300$	$n = 100$	$n = 300$
	0.99	0.0422	0.0135	0.0469	0.0163	0.0367	0.0127
$\gamma = 1/6$	0.995	0.0578	0.0182	0.0658	0.0227	0.0493	0.0168
	0.999	0.109	0.0328	0.130	0.0437	0.0887	0.0290
	0.99	0.109	0.0368	0.144	0.0445	0.105	0.0319
$\gamma = 1/4$	0.995	0.153	0.0496	0.211	0.0628	0.145	0.0421
	0.999	0.311	0.0907	0.476	0.125	0.287	0.0736
	0.99	0.313	0.0753	0.326	0.0999	0.229	0.0682
$\gamma = 1/3$	0.995	0.458	0.103	0.491	0.142	0.320	0.0900
	0.999	1.058	0.194	1.218	0.296	0.668	0.160

Table 4: Relative MSE, case DP(1/2).

MSE		Fréchet		Burr $\rho=-1$		Burr $\rho=-2$	
DP(1/3)	δ	$n=100$	$n=300$	$n=100$	$n=300$	$n=100$	$n=300$
	0.99	0.0475	0.0154	0.0545	0.0198	0.0434	0.0145
$\gamma=1/6$	0.995	0.0642	0.0203	0.0751	0.0271	0.0571	0.0188
	0.999	0.120	0.0353	0.145	0.0505	0.100	0.0312
	0.99	0.142	0.0426	0.155	0.0527	0.120	0.0370
$\gamma=1/4$	0.995	0.197	0.0564	0.220	0.0724	0.162	0.0476
	0.999	0.398	0.100	0.463	0.139	0.311	0.0802
	0.99	0.309	0.0970	0.372	0.116	0.270	0.0803
$\gamma=1/3$	0.995	0.450	0.129	0.551	0.160	0.378	0.104
	0.999	1.058	0.239	1.319	0.320	0.813	0.178

Table 5: Relative MSE, case DP(1/3).

MSE		Fréchet		Burr $\rho=-1$		Burr $\rho=-2$	
PH(2/3)	δ	$n=100$	$n=300$	$n=100$	$n=300$	$n=100$	$n=300$
	0.99	0.0381	0.0140	0.0437	0.0170	0.0362	0.0130
$\gamma=1/6$	0.995	0.0515	0.0184	0.0606	0.0230	0.0483	0.0169
	0.999	0.0959	0.0322	0.119	0.0430	0.0865	0.0283
	0.99	0.107	0.0387	0.118	0.0433	0.106	0.0332
$\gamma=1/4$	0.995	0.148	0.0504	0.168	0.0587	0.143	0.0422
	0.999	0.300	0.0885	0.359	0.112	0.272	0.0696
	0.99	0.242	0.0841	0.344	0.0959	0.235	0.0748
$\gamma=1/3$	0.995	0.346	0.107	0.511	0.131	0.324	0.0939
	0.999	0.792	0.188	1.271	0.260	0.691	0.156

Table 6: Relative MSE, case PH(2/3).

MSE		Fréchet		Burr $\rho=-1$		Burr $\rho=-2$	
PH(1/2)	δ	$n=100$	$n=300$	$n=100$	$n=300$	$n=100$	$n=300$
	0.99	0.0450	0.0192	0.0486	0.0202	0.0442	0.0197
$\gamma=1/6$	0.995	0.0562	0.0230	0.0629	0.0253	0.0549	0.0234
	0.999	0.0926	0.0352	0.113	0.0423	0.0883	0.0345
	0.99	0.119	0.0564	0.132	0.0581	0.119	0.0565
$\gamma=1/4$	0.995	0.149	0.0648	0.172	0.0689	0.146	0.0642
	0.999	0.264	0.0924	0.334	0.108	0.247	0.0878
	0.99	0.253	0.128	0.408	0.149	0.260	0.133
$\gamma=1/3$	0.995	0.307	0.139	0.551	0.169	0.307	0.143
	0.999	0.540	0.182	1.217	0.252	0.495	0.175

Table 7: Relative MSE, case PH(1/2).

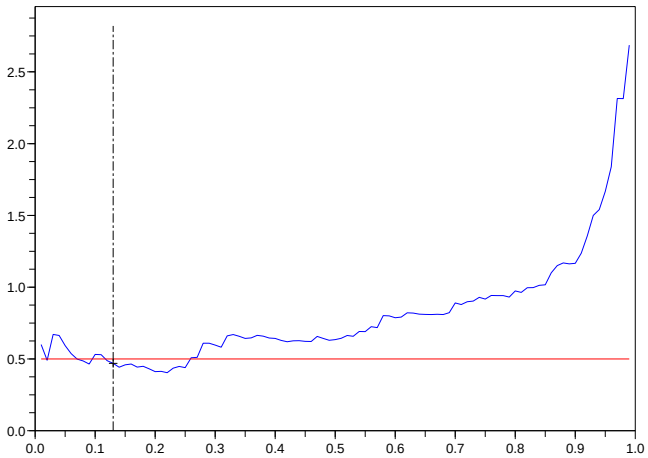


Figure 1: Choosing β on a random sample of $n=100$ Burr observations with $\gamma=1/2$ and $\rho=-1$; $x-$axis: $1-\beta$. The choice procedure is conducted with $\beta_0=0.5$ and $h=0.1$. The blue line is the Hill estimator; we obtain $\beta^*=0.86$ and $\widehat{\gamma}=0.475$.

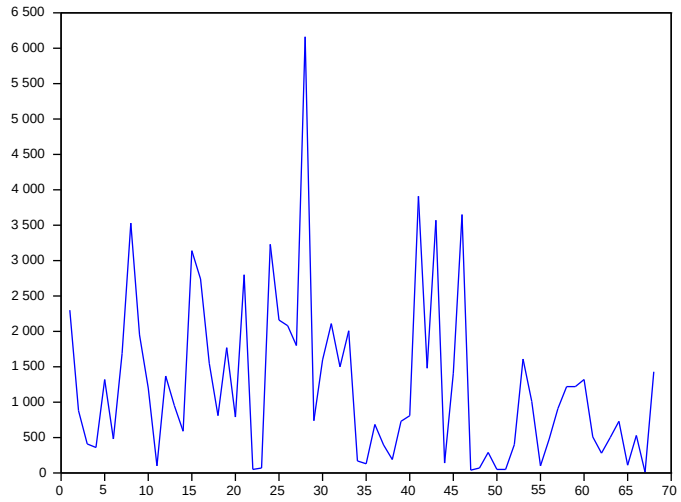


Figure 2: Poker dataset (measurement unit: thousands of USD).

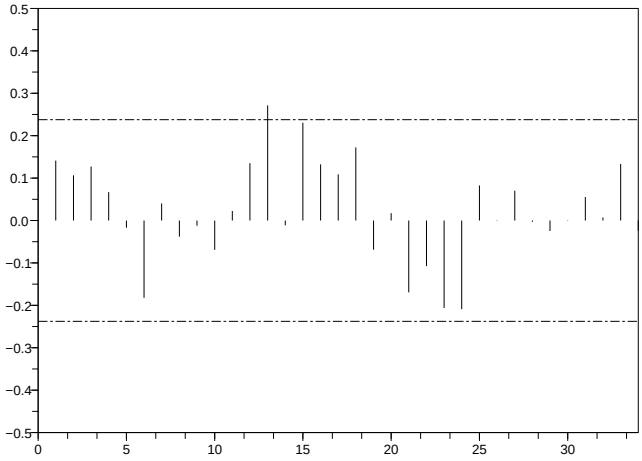


Figure 3: Poker dataset: sample autocorrelation function until lag 34. Dashed line: 95% significance level.

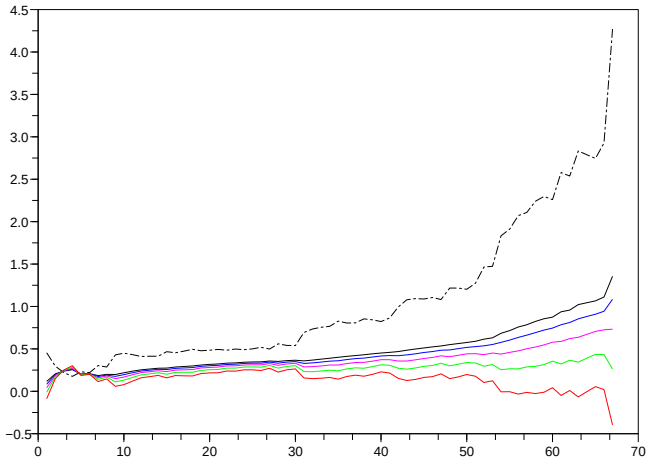


Figure 4: Poker dataset: Hill estimators; $x-$axis: $1-\beta$. Dashed line: standard Hill estimator, black line: estimator $\widehat{\gamma}_\beta^{RB}(1)$, blue line: estimator $\widehat{\gamma}_\beta^{RB}(3/4)$, purple line: estimator $\widehat{\gamma}_\beta^{RB}(1/2)$, green line: estimator $\widehat{\gamma}_\beta^{RB}(1/4)$,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

Estimator $\widehat{\gamma}$	β^*	Estimate of γ
Standard Hill	0.868	0.430
Bias-reduced Hill, $\tau = 1$	0.853	0.197
Bias-reduced Hill, $\tau = 3/4$	0.882	0.182
Bias-reduced Hill, $\tau = 1/2$	0.794	0.225
Bias-reduced Hill, $\tau = 1/4$	0.735	0.196
Bias-reduced Hill, $\tau = 0$	0.735	0.145

Table 8: Poker dataset: estimates of γ.

δ	$\widehat{\text{VaR}}$	$\widehat{\text{CTE}}$	$\widehat{\text{CVaR}_{1/2}}$	$\widehat{\text{DP}(1/2)}$	$\widehat{\text{DP}(1/3)}$
0.95	2234	3900	3067	4668	5155
0.97	2470	4312	3391	5161	5699
0.99	3065	5351	4208	6404	7072

Table 9: Poker dataset: estimating some risk measures (measurement unit: thousands of USD).