Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions

Jonathan El Methni, Gilles Stupfler

Improved estimators of extreme Wang distortion risk measures for very 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)] School of Mathematical Sciences, The University of Nottingham, University Park,
Nottingham NG7 2RD, United Kingdom

Abstract. A general way to study the extremes of a random variable is to consider the family of its Wang distortion risk measures. This class of risk measures encompasses several indicators such as the classical quantile/Value-at-Risk, the Tail-Value-at-Risk and Conditional Tail Moments. Trimmed and winsorised versions of the empirical counterparts of extreme analogues of Wang distortion risk measures are considered. Their asymptotic properties are analysed, and it is shown that it is possible to construct corrected versions of trimmed or winsorised estimators of extreme Wang distortion risk measures who appear to perform overall better than their standard empirical counterparts in difficult finite-sample situations when the underlying distribution has a very heavy right tail. This technique is showcased on a set of real fire insurance data.

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

Keywords: asymptotic normality, extreme value statistics, heavy-tailed distribution, trimming, Wang distortion risk measure, winsorising.

1 Introduction

Early developments of extreme value analysis focused on estimating a quantile at a level so high that the straightforward empirical quantile estimator could not be expected to be consistent. Motivating problems include estimating extreme rainfall at a given location (Koutsoyiannis, 2004) or extreme daily wind speeds (Beirlant et al., 1996), modeling large forest fires (Alvarado et al., 1998), analysing extreme log-returns of financial time series (Drees, 2003) and studying extreme risks related to large losses for an insurance company (Rootzén and Tajvidi, 1997). A large part of practical applications of extreme value theory can actually be modelled using heavy-tailed distributions, which shall be the focus of this paper. A distribution is said to be heavy-tailed if its survival function $1 - F$, where F is the related cumulative distribution function, roughly behaves like a power function with exponent $-1/\gamma$ at infinity where the positive parameter γ is the so-called tail index of the distribution. In such a model, the function $1 - F$ essentially satisfies a homogeneity property and it therefore becomes possible to use an extrapolation method (Weissman, 1978) to estimate quantiles at arbitrarily extreme levels, provided an estimate of γ is computed. Under appropriate stationarity assumptions this analysis can be used to draw predictive conclusions: extreme value analysis has been applied 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 (de Haan and Ferreira, 2006). It is also used nowadays by insurance companies operating in Europe so as to determine their own solvency capital necessary to meet 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.

Of course, the knowledge of a single high quantile is clearly not enough to characterise the behaviour of a random variable in its right tail, since two distributions may well share a quantile at some common level although their respective tail behaviours are different. This is why other quantities such as the Tail-Value-at-Risk, Conditional Value-at-Risk or Conditional Tail Moment (see El Methni et al., 2014) were developed and studied; a common feature of these indicators is that their computation takes into account the whole right tail of the random variable of interest. This, of course, also entails increased sensitivity to a change in tail behaviour compared to what is observed in quantiles, at the population level and at the finite-sample level alike. These measures are of great value in practice, especially in actuarial science: for instance, as mentioned in Dowd and Blake (2006), the Tail-Value-at-Risk would be used if one is interested in the average loss after a catastrophic event or to estimate the cover needed for an excess-of-loss reinsurance treaty. As shown in El Methni and Stupfler (2017), the aforementioned quantities can actually be written as simple combinations of Wang distortion risk measures of a power of the variable of interest (abbreviated by Wang DRMs hereafter; see Wang, 1996). Wang DRMs are weighted averages of the quantile function, the weighting scheme being specified by the so-called distortion function; on the practical side, Wang DRMs can, among others, be useful to price insurance premiums, bonds, and tackle capital allocation problems, see e.g. Wang et al. (1997), Wang (2004) and Belles-Sampera et al. (2014). It is therefore not surprising that the estimation of Wang DRMs above a *fixed* level of risk has been the subject of a number of papers: in particular, we refer the reader to Jones and Zitikis (2003), Necir and Meraghni (2009), Necir et al. (2010) and Deme et al. (2013, 2015).

To the best of our knowledge though, the only study providing estimators of *extreme* distortion risk measures is the recent work of El Methni and Stupfler (2017). More precisely, they show that a simple and efficient solution to estimate extreme Wang DRMs when the right tail of the underlying distribution is moderately heavy is to consider a so-called functional plug-in estimator. Two weaknesses of this study can be highlighted however. The first problem, a practical one, is that it is a consequence of the results in the simulation study of El Methni and Stupfler (2017) that the finite-sample performance of the suggested class of estimators decreases sharply in terms of mean squared error as the tail of the underlying distribution gets heavier. This is due to the propensity of heavier-tailed distributions to generate highly variable top order statistics and, therefore, to increase dramatically the variability of the estimates. No solution is put forward in El Methni and

Stupfler (2017) in order to tackle this issue. The second problem, which is theoretical, is that their asymptotic results about this class of estimators are restricted to asymptotic normality and are thus somewhat frustrating in the sense that they are stated under an integrability condition on the quantile function which is substantially stronger than the simple existence of the Wang DRM to be estimated. In particular, a consistency result under the latter condition, in the spirit of the one Jones and Zitikis (2003) obtained for the estimation of fixed-order Wang DRMs, is not provided in El Methni and Stupfler (2017).

Herein it is shown that robustifying the functional plug-in estimator of El Methni and Stupfler (2017) by deleting certain top order statistics and/or replacing them by lower order statistics, namely trimming or winsorising the estimator, enables one to obtain estimators with reduced variability, as well as to show a consistency result under weaker hypotheses and to retain the asymptotic normality result under the same technical conditions. Trimming and winsorising have both been (and arguably still are) the easiest and most intuitive ways to give a statistical technique some degree of robustness to high-value outliers. A historical account is given in Stigler (1973). The motivation here is rather that the integrability condition of El Methni and Stupfler (2017) depends solely on the behaviour of the quantile function around 1 and becomes more and more stringent as the rate of divergence of this function to infinity increases. At the sample level, this means that this integrability condition has to be fulfilled in order to control the highest order statistics. Deleting the most extreme part of the sample or replacing it by lower (but still high) order statistics can thus be thought of informally as a way to reduce the difficulty of the problem, both from the theoretical and practical point of view.

To be more specific, we shall essentially consider a Wang DRM of a random variable given that it lies between two high-level quantiles, instead of assuming that it simply lies above a high threshold like El Methni and Stupfler (2017) did. This is then estimated by its empirical counterpart, which leads to a trimmed estimator of a Wang DRM. The winsorised estimator, meanwhile, is obtained by considering the empirical counterpart of a Wang DRM given that the random variable lies above a high threshold and is clipped above yet another higher level. By construction, these two estimators do not depend on some of the highest observations, and therefore can be expected to suffer from less finite-sample variability than the original estimator of El Methni and Stupfler (2017) does. To ensure consistency, the highest level (that is, the trimming/winsorising level) is then made to increase to 1 faster than the lowest one does as the sample size increases. Both of these estimators can actually be embedded into a common class of estimators, whose consistency and asymptotic normality are studied. A somewhat surprising feature of this technique is that one can also obtain the consistency of the estimator using the full data above a high level by approximating it by such robustified estimators whose fraction of deleted data becomes smaller as the sample size increases; this argument is actually similar in spirit to a proof by Etemadi (1981) of the law of large numbers for independent copies of an integrable random variable, starting with the case when the variance is finite and concluding by a truncation argument.

These new estimators, for all their improved properties as far as variability is concerned, should be expected to suffer from finite-sample bias issues, since they are in fact sample counterparts of a different quantity than the originally targeted Wang DRM. The second step is then to devise a correction method which allows the estimator to have a bias intuitively similar to that of the basic functional plug-in estimator and therefore to be (almost) unbiased in practice, while retaining its low variability. The gist of the correction step is to note that the newly proposed estimators are in reality approximately equal to the Wang DRM to be estimated multiplied by a quantity converging to 1 and depending on the extremes of the sample only. This makes it possible to estimate the error made when using the purely trimmed or winsorised estimators and thus to design corrected estimators by multiplying them by a simple and intuitive correction factor. This correction step should therefore be viewed as closer to Bessel's correction method for the sample variance estimator when the mean is unknown, rather than to traditional bias-correction devices developed in second-order extreme value frameworks such as the estimators of Peng (1998) and Caeiro et al. (2005) which are based on the asymptotic distribution of an estimator to be corrected. Of course, while this approach should only be expected to be reasonable if the threshold above which the Wang DRM is computed may be consistently estimated by its empirical analogue, extreme Wang DRM estimators can be obtained afterwards by an extrapolation technique warranted by the extreme value framework.

The outline of this paper is the following. In Section 2, we recall what Wang DRMs are, as well as a definition of extreme analogues of Wang DRMs presented in El Methni and Stupfler (2017). Section 3 then considers their estimation, by introducing a two-stage improvement of the functional plug-in estimator of El Methni and Stupfler (2017), first in the intermediate case and then in the arbitrarily extreme case. The finite-sample performance of the estimators is examined on a simulation study in Section 4 and the method is applied on a real insurance data set in Section 5. Section 6 concludes and offers some perspective on future work. Proofs of all results and additional simulation results are deferred to an online supplementary material document.

2 Extreme Wang DRMs

It shall be said in all what follows that a function $g : [0,1] \to [0,1]$ is a distortion function if it is nondecreasing and right-continuous, with $g(0) = 0$ and $g(1) = 1$. Let X be a positive random variable with cumulative distribution function F. The Wang distortion risk measure (DRM) of X with distortion function g is (Wang, 1996):

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

An alternative, easily interpretable expression of $R_g(X)$ is actually available, and it shall be used extensively in what follows. 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 $M = \sup\{\alpha \in [0,1] \,|\, g(\alpha) < 1\}$, and assume 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

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

A Wang DRM is thus a Lebesgue-Stieltjes weighted version of the expectation of the random variable X, the weighting scheme being given by the measure $dg(\cdot)$. The above formula is actually true when g is continuous, with no condition at all on the distribution of X; when g is absolutely continuous, the weight is given by the Lebesgue derivative g' of g. Specific examples include

- the quantile (or Value-at-Risk) at level β for $g(x) = \mathbb{I}\{x \geq 1-\beta\}$, with $\mathbb{I}\{\cdot\}$ being the indicator function, in which case $dg(\cdot)$ is actually the Dirac measure at $1-\beta$;

- the Tail-Value-at-Risk TVaR(β) in the worst $100(1-\beta)\%$ cases, namely the average of all quantiles exceeding the quantile $q(\beta)$, for $g(x) = \min(x/(1-\beta), 1)$ and $dg(\cdot)$ being the Lebesgue measure on $[0, 1-\beta]$ up to a positive constant.

For more examples of DRMs, see Table 1 in El Methni and Stupfler (2017).

While the family of Wang DRMs of X already gives more information than a finite number of its quantiles, yet more information may be recovered by considering Wang DRMs of functions of X. More precisely, if $h : (0, \infty) \to (0, \infty)$ is a strictly increasing, continuously differentiable function then, under the aforementioned regularity conditions, the Wang DRM of $h(X)$ with distortion function g is

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

Since, when F is continuous, the Conditional Tail Moment (CTM) of order a of X (see El Methni et al., 2014) is

$$\mathbb{E}(X^a | X > q(\beta)) = \frac{1}{1-\beta} \int_0^{1-\beta} [q(1-\alpha)]^a d\alpha,$$

the CTM of order a may therefore be obtained by choosing $g(x) = \min(x/(1-\beta), 1)$, $\beta \in (0,1)$ and $h(x) = x^a$, with $a > 0$, and so may any risk measure obtained by combinations of CTMs; we refer the reader to Table 2 in El Methni and Stupfler (2017) for further examples.

The idea developed in El Methni and Stupfler (2017) in order to obtain Wang DRMs of the extremes of X is to consider

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

A similar, if slightly different, idea is Yang (2015), while a construction adapted to stop-loss risk measures is Vandewalle and Beirlant (2006). In the remainder of this paper, it is assumed that the quantile function q of X is continuous and strictly increasing in a neighbourhood of infinity; it can then be shown (see Proposition 1 in El Methni and Stupfler, 2017) that $R_{g,\beta}(h(X))$ is actually, for β large enough, the Wang DRM R_g of $h(X_\beta)$, where $X_\beta \stackrel{d}{=} X | X > q(\beta)$. In other words, $R_{g,\beta}(h(X))$ is the Wang DRM of $h(X)$ given that X lies above a (high) level. Using this construction, it is very easy to recover several extreme parameters such as an extreme quantile/Value-at-Risk, an extreme Tail-Value-at-Risk or extreme versions of the CTMs.

An important question is then to consider the estimation of such extreme Wang DRMs. An idea to tackle this problem is that of El Methni and Stupfler (2017): consider a sample of independent random variables $(X_1, \ldots, X_n)$ having cumulative distribution function F and (β_n) a nondecreasing sequence of real numbers belonging to $(0, 1)$ which converges to 1. Moreover, denote by $\widehat{F}_n$ the empirical cumulative distribution function related to this sample and by $\widehat{q}_n$ the related empirical quantile function:

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

in which $X_{1,n} \le \cdots \le X_{n,n}$ are the order statistics of the sample $(X_1, \ldots, X_n)$ and $\lceil \cdot \rceil$ denotes the ceiling function. A first step is to estimate the Wang DRM $R_{g,\beta_n}(h(X))$ by its empirical, functional plug-in counterpart:

$$\widehat{R}^{\mathrm{PL}}_{g,\beta_n}(h(X)) := \int_0^1 h \circ \widehat{q}_n(1 - (1 - \beta_n)s)dg(s) = \int_0^1 h(X_{\lceil n(1-(1-\beta_n)s) \rceil, n})dg(s).$$

When h is a power function, which is enough to recover all Wang DRMs as well as the class of Conditional Tail Moments, this estimator, which shall be referred to as the PL estimator hereafter, is consistent and asymptotically normal when (β_n) is an intermediate sequence, namely $n(1-\beta_n) \to \infty$ as $n \to \infty$, within an extreme value framework which will be introduced shortly (see El Methni and Stupfler, 2017). This is a usual and well-understood restriction in extreme value theory: to estimate the Wang DRM above level $q(\beta_n)$ by an empirical estimator in a consistent fashion, then $q(\beta_n)$ should be asymptotically within the range of the data, or equivalently, there should be a growing number of data points above $q(\beta_n)$ to ensure that its empirical estimator $X_{\lceil n\beta_n \rceil, n}$ is relatively consistent. The case when $n(1 - \beta_n) \to \lambda < \infty$, corresponding to proper extreme quantiles, is then handled by the classical extrapolation argument of Weissman (1978).

3 Extreme Wang DRM estimation

3.1 Heavy tails, top order statistics and finite-sample variability

A problem with the use of the PL estimator in practice can arise when g is strictly increasing in a left neighbourhood of 1, which is for instance the case for the Tail-Value-at-Risk, Dual Power and Proportional Hazard risk measures considered in El Methni and Stupfler (2017). In that case, the PL estimator takes into account all the data above level $X_{\lceil n\beta_n \rceil, n}$ in the sample; in any sample where some of the highest order statistics are far from their population counterparts, this will result in inappropriate estimates. Such situations appear regularly: suppose here that X has a Pareto distribution with parameter $1/\gamma$, i.e.

$$\forall x \ge 1, \; F(x) = 1 - x^{-1/\gamma} \quad \text{so that} \quad \forall \alpha \in (0, 1), \; q(\alpha) = (1 - \alpha)^{-\gamma}.$$

The probability that the sample maximum $X_{n,n}$ exceeds a multiple of its population counterpart, namely the quantile $q(1-n^{-1})$, is then

$$\begin{aligned}\mathbb{P}(X_{n,n} > Kq(1-n^{-1})) &= 1-\left[1-\mathbb{P}(X > Kq(1-n^{-1}))\right]^n \\ &= 1-\left[1-\frac{K^{-1/\gamma}}{n}\right]^n \\ &\approx 1-\exp(-K^{-1/\gamma}) \ \text{ for large enough } n.\end{aligned}$$

This result is, of course, linked to the fact that sample quantiles at extreme levels do not estimate the corresponding true quantiles consistently; a related point is that, sample-wise, the most extreme values tend not to give a fair picture of the extremes of the underlying distribution (see e.g. Ghosh and Resnick, 2010). Carrying on with this example, it follows that in the case $\gamma = 0.49$, $K = 3$, and $n = 1000$, the sample maximum is larger than 88.54, which is three times the quantile at level 0.999, with probability approximately equal to 0.101. In this sense, approximately 10% of samples feature at least one unusually high value. Besides, as the above calculation shows, the probability that a sample features one or several very large values increases as γ increases, i.e. as the tail gets heavier. The influence of such values on extreme Wang DRM estimates should of course not be underestimated. In the case of the Tail-Value-at-Risk, obtained for $g(s) = s$, namely:

$$\text{TVaR}(\beta) = R_{\text{Id},\beta}(X) = \frac{1}{1-\beta}\int_0^{1-\beta} q(1-\alpha)d\alpha = \frac{(1-\beta)^{-\gamma}}{1-\gamma},$$

a simulation study shows that, on 5000 replicates of a sample of 1000 independent random copies of the aforementioned Pareto distribution conditioned on the fact that the sample maximum is larger than 88.54, the relative bias of the Tail-Value-at-Risk PL estimator,

$$\widehat{R}^{\text{PL}}_{\text{Id},\beta}(X) = \frac{1}{1-\beta}\int_0^{1-\beta} \widehat{q}_n(1-\alpha)d\alpha = \frac{1}{n(1-\beta)}\sum_{j=1}^{n(1-\beta)} X_{n-j+1,n},$$

at level $\beta = 0.95$ is approximately 0.389. In other words, the PL estimator is, on such samples, on average a little less than 40% higher than it should be. Of course, this could have been expected since it is straightforward to see that the above PL estimator is adversely affected by high values of $X_{n,n}$ (just as the sample mean is). The concern here is rather that problematic cases, through the apparition of very high values of the sample maximum and more generally of the highest order statistics, appear more and more frequently as γ increases, even when γ is such that the estimator $\widehat{R}^{\text{PL}}_{\text{Id},\beta}(X)$ is asymptotically Gaussian (which is the case here for the extreme Tail-Value-at-Risk estimator since $\gamma = 0.49 < 1/2$, see Theorem 2 in El Methni and Stupfler, 2017). This should therefore mean increased variability of the estimates as γ gets larger and indeed, at the finite-sample level, MSEs become higher (up to unsustainably high levels) when the tail of X gets heavier, as the simulation study in El Methni and Stupfler (2017) tends to show. Our first objective is to introduce estimators which deal with this variability issue.

3.2 First step of improvement: reducing finite-sample variability

A simple idea to tackle the problem highlighted in Section 3.1 is to delete the highest problematic values altogether, namely to trim the PL estimator, by considering the statistic

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

where (t_n) is a sequence of trimming levels, i.e. a sequence such that $\beta_n < t_n \leq 1$. This is the empirical estimator of

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

which in many cases is actually the Wang DRM of $h(X)$ given that X lies between $q(\beta_n)$ and $q(t_n)$, as the following result shows.

Proposition 1. *Let $\beta \in (0,1)$ and $t \in (0,1]$ such that $t > \beta$. If q is continuous and strictly increasing on an open interval containing $[\beta,1)$ then:*

$$R^{\text{Trim}}_{g,\beta,t}(h(X)) = R_g(h(X^{\text{Trim}}_{\beta,t})) \quad \textit{with} \quad X^{\text{Trim}}_{\beta,t} \stackrel{d}{=} X | X \in [q(\beta), q(t)].$$

In practice, it is very often the case that nt_n and $n(t_n - \beta_n)$ are positive integers (see Sections 4 and 5). In that particular case, the trimmed estimator $\widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(h(X))$, which we shall call the Trim-PL estimator, can be conveniently rewritten as a generalised L-statistic:

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

When the function g is moreover continuous on $[0,1]$, this can be further simplified as

$$\begin{aligned}
&\widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(h(X)) \\
&= h(X_{n\beta_n+1,n}) + \sum_{i=1}^{n(t_n-\beta_n)-1} g\left(\frac{i}{n(t_n-\beta_n)}\right) [h(X_{nt_n-i+1,n}) - h(X_{nt_n-i,n})].
\end{aligned}$$

It should thus be clear at this stage that the Trim-PL estimator $\widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(h(X))$ is both the empirical counterpart of $R^{\text{Trim}}_{g,\beta_n,t_n}(h(X))$ and a trimmed estimator of $R_{g,\beta_n}(h(X))$ in the sense that the top order statistics $X_{nt_n+1,n},\ldots,X_{n,n}$ are discarded for the estimation. This amounts to a trimming percentage equal to $100(1-t_n)\%$ in the highest values of the sample. The intermediate PL estimator of El Methni and Stupfler (2017) is recovered for $t_n=1$.

Although the idea of trimming seems appealing because it is expected to curb the estimator's variability, it may not be the best method available in that it effectively reduces the available sample size. The overall bias of the estimator, meanwhile, would be negatively affected as well, since despite their high variability, the highest order statistics in the sample are those who carry the least bias about the extremes of the underlying distribution. One could try reducing the loss of information that trimming entails by winsorising the estimator $\widehat{R}^{\mathrm{PL}}_{g,\beta_n}(h(X))$ instead, which amounts to considering the following so-called Wins-PL estimator:

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

When nt_n and $n(t_n-\beta_n)$ are positive integers it is easy to see that, contrary to the trimmed estimator, the winsorised estimator replaces the data points $X_{nt_n+1,n},\ldots,X_{n,n}$ by $X_{nt_n,n}$. This estimator can of course also be written as a generalised L-statistic, viz.

$$\begin{aligned}
&\widehat{R}^{\mathrm{Wins}}_{g,\beta_n,t_n}(h(X))\\
=\;& h(X_{nt_n,n})\int_0^1 \mathbb{I}\{0\le s<x_{n(1-t_n),n}(\beta_n,1)\}dg(s)\\
+\;& \sum_{i=n(1-t_n)+1}^{n(1-\beta_n)} h(X_{n-i+1,n})\int_0^1 \mathbb{I}\{x_{i-1,n}(\beta_n,1)\le s<x_{i,n}(\beta_n,1)\}dg(s)\\
+\;& h(X_{n\beta_n,n})\left[g(1)-\lim_{\substack{s\to1\\ s<1}} g(s)\right]\\
=\;& \sum_{i=n(1-t_n)+1}^{n(1-\beta_n)} h(X_{n-i+1,n})\left[\lim_{\substack{s\to x_{i,n}(\beta_n,1)\\ s<x_{i,n}(\beta_n,1)}} g(s)-\lim_{\substack{s\to x_{i-1,n}(\beta_n,1)\\ s<x_{i-1,n}(\beta_n,1)}} g(s)\right]\\
+\;& h(X_{n\beta_n,n})\left[g(1)-\lim_{\substack{s\to1\\ s<1}} g(s)\right]+h(X_{nt_n,n})\lim_{\substack{s\to x_{n(1-t_n),n}(\beta_n,1)\\ s<x_{n(1-t_n),n}(\beta_n,1)}} g(s).
\end{aligned}$$

If g is continuous on $[0,1]$, this reads:

$$\begin{aligned}
&\widehat{R}^{\mathrm{Wins}}_{g,\beta_n,t_n}(h(X))\\
=\;& h(X_{n\beta_n+1,n})+\sum_{i=n(1-t_n)+1}^{n(1-\beta_n)-1} g\left(\frac{i}{n(1-\beta_n)}\right)[h(X_{n-i+1,n})-h(X_{n-i,n})].
\end{aligned}$$

Like the Trim-PL estimator, the Wins-PL estimator is a direct empirical estimator, of the quantity

$$R^{\mathrm{Wins}}_{g,\beta_n,t_n}(h(X))=\int_0^1 h\circ q(\min(t_n,1-(1-\beta_n)s))dg(s),$$

which is actually essentially the Wang DRM of $h(X)$ given that X is larger than $q(\beta_n)$ and clipped above level $q(t_n)$:

Proposition 2. *Let $\beta \in (0,1)$ and $t \in (0,1]$ such that $t > \beta$. If q is continuous and strictly increasing on an open interval containing $[\beta, 1)$ then:*

$$R_{g,\beta,t}^{\text{Wins}}(h(X)) = R_g(h(X_{\beta,t}^{\text{Wins}}))$$
$$\textit{with} \qquad X_{\beta,t}^{\text{Wins}} \stackrel{d}{=} X\mathbb{I}\{q(\beta) \le X < q(t)\} + q(t)\mathbb{I}\{X \ge q(t)\}.$$

The focus of this paper is to study the merits of trimming/winsorising in the context of the estimation of extreme Wang DRMs, both theoretically and at the finite-sample level. While it would be straightforward to obtain the asymptotic properties of both estimators for fixed orders β and t through L-statistic techniques (see e.g. Jones and Zitikis, 2003), a difficulty here lies in the fact that $\beta = \beta_n \uparrow 1$. As a consequence, theoretical developments involve knowing the weak behaviour of the quantile process $s \mapsto \widehat{q}_n(s)$ on $[\beta_n, 1]$. The crucial tool is a corollary of the powerful distributional approximation stated in Theorem 2.1 of Drees (1998), relating this tail quantile process to a standard Brownian motion up to a bias term. The relevant framework for this result 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 on the half-line $(0, \infty)$ and for any $x > 0$, $f(tx)/f(t) \to x^b$ as $t \to \infty$. In this paper, the distribution of X is heavy-tailed, namely, $1 - F$ is regularly varying with index $-1/\gamma < 0$, the parameter γ being the so-called tail index of the cumulative distribution function F. We shall actually use an equivalent assumption on the left-continuous inverse U of $1/(1-F)$, defined for $y \ge 1$ by $U(y) = \inf\{t \in \mathbb{R} \,|\, 1/(1-F(t)) \ge y\} = q(1 - y^{-1})$, and called the tail quantile function. More precisely, the main hypothesis is that the tail quantile function is regularly varying with index γ and satisfies a 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 $\gamma > 0$, $\rho \le 0$ 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 $x^\gamma \log x$.

In condition $\mathcal{C}_2(\gamma, \rho, A)$, the function $|A|$ must be regularly varying at infinity with index ρ (see Theorem 2.3.3 in de Haan and Ferreira, 2006). Such a condition is classical when studying estimators of extreme parameters of a heavy-tailed distribution, because it makes it possible, through the function A, to measure the deviation of the distribution of the random variable of interest from the Pareto distribution, the latter being the simplest case of a heavy-tailed distribution. The function A thus typically appears in bias conditions. Most standard examples of heavy-tailed distributions used in extreme value theory satisfy assumption $\mathcal{C}_2(\gamma, \rho, A)$, see e.g. the examples p.59 in Beirlant et al. (2004) and pp.61–62 in de Haan and Ferreira (2006).

Our next step is to highlight that the Trim-PL and Wins-PL estimators are actually part of a common class of estimators. For $0 < \beta < t \le 1$, let $\mathcal{F}(\beta, t)$ be the set of those nonincreasing Borel measurable functions ψ taking values in $[0,1]$ such that

$$\psi(0) = t, \ \psi(1) = \beta \ \text{ and } \ \forall s \in [0,1], \ 0 \le 1 - (1-\beta)s - \psi(s) \le 1 - t.$$

Let now (ψ_n) be a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n, t_n)$, and set

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

whose empirical counterpart is the estimator

$$\widehat{R}_{g,\beta_n}(h(X);\psi_n) = \int_0^1 h \circ \widehat{q}_n \circ \psi_n(s) dg(s) = \int_0^1 h(X_{\lceil n\psi_n(s)\rceil,n}) dg(s).$$

All estimators in this class only take into account data points among the $X_{i,n}$, $\lceil n\beta_n \rceil \le i \le \lceil nt_n \rceil$, and can therefore be considered robust with respect to change in the most extreme values in the sample when $n(1-t_n) \ge 1$. The class of estimators $\widehat{R}_{g,\beta_n}(h(X);\psi_n)$ is a reasonable, unifying framework for our purpose: indeed, particular examples of the sequence (ψ_n) are $s \mapsto t_n - (t_n - \beta_n)s$ which appears as the argument of the empirical quantile function in the Trim-PL estimator, and $s \mapsto \min(t_n, 1-(1-\beta_n)s)$ giving rise to the Wins-PL estimator. These two examples should be those coming to mind when reading the asymptotic results below. Finally, the case $\psi_n(s) = 1-(1-\beta_n)s$, corresponding to the original PL estimator of El Methni and Stupfler (2017), is recovered by setting $t_n = 1$.

At the technical level, because $\psi_n(s) \approx 1-(1-\beta_n)s$ in a certain sense when t_n is close enough to 1, the quantity $R_{g,\beta_n}(h(X);\psi_n)$ should be expected to be close to $R_{g,\beta_n}(h(X))$ and therefore, $\widehat{R}_{g,\beta_n}(h(X);\psi_n)$ should be thought to be a consistent estimator of $R_{g,\beta_n}(h(X))$. The first result below shows that $\widehat{R}_{g,\beta_n}(h(X);\psi_n)$ is a relatively consistent and $\sqrt{n(1-\beta_n)}$−asymptotically normal estimator of $R_{g,\beta_n}(h(X))$ when h is a power function, under suitable conditions on β_n and t_n.

Theorem 1. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that (ψ_n) is a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n,t_n)$, with $0 < \beta_n < t_n \le 1$, $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $(1-t_n)/(1-\beta_n) \to 0$.*

(i) Pick a distortion function g and $a > 0$, and assume that for some $\eta > 0$, we have

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

If furthermore $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) = \mathrm{O}(1)$ then:

$$\frac{\widehat{R}_{g,\beta_n}(X^a;\psi_n)}{R_{g,\beta_n}(X^a)} - 1 \xrightarrow{\mathbb{P}} 0.$$

(ii) Pick distortion functions $g_1,\ldots,g_d$ and $a_1,\ldots,a_d > 0$, and assume that for some $\eta > 0$, we have

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

$$\text{and} \qquad \sqrt{n(1-t_n)}\left(\frac{1-t_n}{1-\beta_n}\right)^{\varepsilon} \to 0,$$

for some $\varepsilon \in (0, \min(1/2,\eta))$. If furthermore

$$\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R},$$

then:

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

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

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

A particularly appealing consequence of Theorem 1 is the consistency of the estimators $\widehat{R}_{g,\beta_n}(X^a;\psi_n)$ under mild conditions. Especially, the integrability condition $\int_0^1 s^{-a\gamma-\eta}dg(s)<\infty$ needed to ensure consistency is weaker than the integrability condition $\int_0^1 s^{-a\gamma-1/2-\eta}dg(s)<\infty$ required in El Methni and Stupfler (2017). Broadly speaking, the former condition is essentially the one required for the existence of the Wang DRM $R_{g,\beta_n}(X^a)$ to be estimated, while the latter is needed to write a weak approximation of the estimator by an integral of a standard Brownian motion. We may in fact choose $t_n=1$, for which the consistency of the PL estimator of El Methni and Stupfler (2017), which is not shown therein, is obtained; let us point out that the proof of Theorem 1(i) consists of two steps, the first one being to prove that any estimator $\widehat{R}_{g,\beta_n}(X^a;\psi_n)$ is asymptotically equivalent to another estimator in this class for which $n(1-t_n)\ge 1$, and the second one being to show the consistency of the latter estimator. In particular, the consistency of proper trimmed/winsorised estimators can be used together with an approximation argument to obtain the consistency of the estimator using all the data above a high threshold.

A second property of the estimators $\widehat{R}_{g,\beta_n}(X^a;\psi_n)$ is that they share the same limiting Gaussian distribution under the classical bias condition

$$\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1})\to\lambda\in\mathbb{R},$$

and provided hypothesis $\sqrt{n(1-t_n)}[(1-t_n)/(1-\beta_n)]^{\varepsilon}\to 0$, relating the order t_n to the intermediate level β_n, holds true. This condition implies that t_n should converge to 1 quickly enough, or, in other words, that not too many values should be deleted from the sample for asymptotic unbiasedness to hold. The necessity of such a condition appears in the earlier works of Csörgő et al. (1986a) and Csörgő et al. (1986b) in the context of mean estimation by the trimmed sample mean: in the former paper, it is shown that discarding a fixed number of order statistics does not create asymptotic bias, while the latter paper states that this may not be true for more severe trimmings. It should be noted that the present assumption is clearly satisfied for $t_n=1-c/n$, with c being a fixed nonnegative integer, corresponding to the case when the top c order statistics are discarded and the trimming/winsorising percentage across the whole sample is $100c/n\%$. Finally, taking $t_n=1$ in Theorem 1(ii) yields the original asymptotic normality result for the PL estimator in El Methni and Stupfler (2017).

As noted therein, the integrability conditions of Theorem 1 can be difficult to grasp. They are, however, determined by the behaviour of g in a neighbourhood of 0, which motivates the introduction of the classes of functions

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

The classes $\mathcal{E}_b[0,1]$, $b > -1$ can be considered as the spaces of those continuously differentiable functions g on $(0,1)$ whose first derivative behaves like a power of s in a neighbourhood of 0. Especially, any polynomial function belongs to $\mathcal{E}_0[0,1]$, and the Proportional Hazard (Wang, 1995) distortion function $g(s) = s^\alpha$, $\alpha \in (0,1)$ belongs to $\mathcal{E}_{\alpha-1}[0,1]$. The next result sums up what can be said when g belongs to such a space.

Corollary 1. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that (ψ_n) is a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n, t_n)$, with $0 < \beta_n < t_n \le 1$, $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $(1-t_n)/(1-\beta_n) \to 0$.*

(i) Pick a distortion function g and $a > 0$. Assume that g belongs to some $\mathcal{E}_b[0,1]$ with $b > -1$. If $\gamma < (b+1)/a$ and $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) = \mathrm{O}(1)$ then:

$$\frac{\widehat{R}_{g,\beta_n}(X^a;\psi_n)}{R_{g,\beta_n}(X^a)} - 1 \xrightarrow{\mathbb{P}} 0.$$

(ii) Pick distortion functions $g_1,\ldots,g_d$ and $a_1,\ldots,a_d > 0$. Assume there are $b_1,\ldots,b_d > -1$ such that $g_j \in \mathcal{E}_{b_j}[0,1]$ for all $j \in \{1,\ldots,d\}$. If $\gamma < (2b_j+1)/(2a_j)$ for all $j \in \{1,\ldots,d\}$ and

$$\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R} \text{ and } \sqrt{n(1-t_n)}\left(\frac{1-t_n}{1-\beta_n}\right)^{\varepsilon} \to 0,$$

for some $\varepsilon \in (0, \min(0, b_1 - a_1\gamma, \ldots, b_d - a_d\gamma) + 1/2)$ then:

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

with V as in Theorem 1.

As previously noted, the integrability condition for the asymptotic normality of our class of estimators is that of El Methni and Stupfler (2017), which was already obtained by El Methni et al. (2014) in the case of the CTM of order a (for which $b = 0$). In this case, Corollary 1 shows that the condition $\gamma < 1/a$, which is exactly the condition needed to ensure that the CTM of order a exists, is sufficient to make sure that the estimator $\widehat{R}_{g,\beta_n}(X^a;\psi_n)$ is consistent. For instance, in the case $a = 1$, corresponding to the estimation of the extreme Tail-Value-at-Risk, this condition is $\gamma < 1$, which is exactly the condition required for the existence of a finite mean, instead of the more restrictive condition $\gamma < 1/2$ which would be required for the existence of a finite second moment. By contrast, El Methni et al. (2014), working in a model with random covariates, always require $\gamma < 1/(2a)$ in their asymptotic results. They do, however, only assume that a first-order condition holds instead of second-order condition $\mathcal{C}_2(\gamma,\rho,A)$, which is made possible since their CTM estimator can be written as a sum of independent and identically distributed random variables and is thus much easier to handle than the generalised L-statistic $\widehat{R}_{g,\beta_n}(X^a;\psi_n)$.

3.3 Second step of improvement: finite-sample bias correction

The estimators introduced above have been shown to be asymptotically normal estimators of Wang DRMs. It should be noted that on finite-sample situations, such estimators can be expected to carry some (negative) bias, all the more so as the trimming/winsorising order t_n increases. An intuitive justification for this behaviour is that the estimator $\widehat{R}_{g,\beta_n}(X^a;\psi_n)$ is actually the empirical counterpart of $R_{g,\beta_n}(X^a;\psi_n)$, which is in general different from, and especially less than, the target DRM $R_{g,\beta_n}(X^a)$. For instance, in the case of extreme Tail-Value-at-Risk estimation for the Pareto distribution with tail index γ, then the Tail-Value-at-Risk of X in the worst $100(1-\beta_n)\%$ cases is

$$R_{g,\beta_n}(X) = \frac{(1-\beta_n)^{-\gamma}}{1-\gamma},$$

see Section 3.1. By contrast, the trimmed Tail-Value-at-Risk given that X lies between levels $q(\beta_n)$ and $q(t_n)$ is obtained for $\psi_n(s) = t_n - (t_n - \beta_n)s$ and is

$$\begin{aligned} R^{\text{Trim}}_{g,\beta_n,t_n}(X) = \int_0^1 q(t_n - (t_n-\beta_n)s)dg(s) &= \int_0^1 [1 - t_n + (t_n - \beta_n)s]^{-\gamma} ds \\ &= \frac{(1-\beta_n)^{1-\gamma} - (1-t_n)^{1-\gamma}}{(1-\gamma)(t_n-\beta_n)}. \end{aligned}$$

Rewriting this as

$$R^{\text{Trim}}_{g,\beta_n,t_n}(X) = R_{g,\beta_n}(X)\left\{\frac{(1-\beta_n)^{1-\gamma} - (1-t_n)^{1-\gamma}}{(1-\beta_n)^{-\gamma}(t_n-\beta_n)}\right\},$$

results in an expression of $R^{\text{Trim}}_{g,\beta_n,t_n}(X)$ as $R_{g,\beta_n}(X)$ multiplied by a quantity depending on β_n, t_n and γ and smaller than 1. In the case $n = 1000$, $\beta_n = 0.9$, $t_n = 0.99$ and $\gamma = 1/2$, namely the top 100 observations are selected and the top 10 observations among them are eliminated, the reduction factor is actually 0.760, i.e. the expected relative bias is -0.240. When the number of observations removed is halved ($t_n = 0.995$) this factor becomes 0.817 for an expected relative bias of -0.183. The smallest trimming percentage, obtained when $t_n = 0.999$, for removal of the sample maximum only, results in a reduction factor of 0.909, which is still an expected relative bias of -0.091.

To retain the reduction in variability brought by the estimator $\widehat{R}_{g,\beta_n}(X;\psi_n)$ and at the same time obtain an estimator with acceptable finite-sample bias, we design a new estimator based on the previous calculation. More precisely, in the case of Tail-Value-at-Risk estimation, estimating γ by a consistent estimator $\widehat{\gamma}_n$ and plugging in the previous estimator $\widehat{R}_{g,\beta_n}(X;\psi_n) = \widehat{R}^{\text{Trim}}_{g,\beta_n,t_n}(X)$ in the left-hand side of the above equality gives the corrected estimator

$$\widetilde{R}_{g,\beta_n}(X;\psi_n) = \widehat{R}_{g,\beta_n}(X;\psi_n)\left\{\frac{(1-\beta_n)^{1-\widehat{\gamma}_n} - (1-t_n)^{1-\widehat{\gamma}_n}}{(1-\beta_n)^{-\widehat{\gamma}_n}(t_n-\beta_n)}\right\}^{-1}.$$

Note that the correction factor is in fact

$$\left\{\frac{(1-\beta_n)^{1-\widehat{\gamma}_n} - (1-t_n)^{1-\widehat{\gamma}_n}}{(1-\beta_n)^{-\widehat{\gamma}_n}(t_n-\beta_n)}\right\}^{-1} = \frac{R_{g,\beta_n}(Y_{\widehat{\gamma}_n})}{R_{g,\beta_n}(Y_{\widehat{\gamma}_n};\psi_n)},$$

where Y_γ has a Pareto distribution with tail index γ. There is an abundant literature on consistent estimation of the parameter γ: we refer, among others, to the very popular Hill estimator (Hill, 1975), the Pickands estimator (Pickands, 1975), the maximum likelihood estimator (Smith, 1987 and Drees et al., 2004) and probability-weighted moment estimators (Hosking et al., 1985 and Diebolt et al., 2007). A comprehensive review is contained in Section 5 of Gomes and Guillou (2015).

Of course, in practice the underlying distribution of X is not known, but in many cases the Pareto distribution (or a multiple of it) still provides a decent approximation for X in its right tail. It can thus be expected that in a wide range of situations and for n large enough,

$$R_{g,\beta_n}(X^a) = R_{g,\beta_n}(X^a;\psi_n)\frac{R_{g,\beta_n}(X^a)}{R_{g,\beta_n}(X^a;\psi_n)} \approx R_{g,\beta_n}(X^a;\psi_n)\frac{R_{g,\beta_n}(Y_{a\gamma})}{R_{g,\beta_n}(Y_{a\gamma};\psi_n)}.$$

This motivates the following class of corrected estimators:

$$\begin{aligned}
\widetilde{R}_{g,\beta_n}(X^a;\psi_n) &= \widehat{R}_{g,\beta_n}(X^a;\psi_n)\frac{R_{g,\beta_n}(Y_{a\widehat{\gamma}_n})}{R_{g,\beta_n}(Y_{a\widehat{\gamma}_n};\psi_n)} \\
&= \widehat{R}_{g,\beta_n}(X^a;\psi_n)\frac{\int_0^1[(1-\beta_n)s]^{-a\widehat{\gamma}_n}dg(s)}{\int_0^1[1-\psi_n(s)]^{-a\widehat{\gamma}_n}dg(s)}.
\end{aligned}$$

This estimator should be seen as the result of a two-stage procedure:

- first, compute an estimator of the target extreme Wang DRM using a trimmed/winsorised sample, thus reducing variability;
- then, use what can be found on the tail behaviour of the sample to shift the previous estimate back to an essentially bias-neutral position.

Let us emphasise that this bias-correction procedure is a simple one, much closer in spirit to the construction of the corrected sample variance estimator when the population mean is unknown than to bias-reduction methods based on asymptotic results in a second-order extreme value framework, of which an excellent summary is Section 5.3 in Gomes and Guillou (2015) again. In particular, the multiplicative correction factor introduced here only depends on the tail index γ, but not on the second-order parameter ρ. Finally, note that the correction factor might depend on the top values in the sample, but can only actually do so through the estimator $\widehat{\gamma}_n$. For instance, the Hill estimator of γ,

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

of which a bias-reduced version is considered in the simulation study below, depends on the top values only through their logarithms, which sharply reduces their contribution to the variability of our final estimator.

The next result shows that any member of this new class of corrected estimators shares the asymptotic properties of its uncorrected version. Our preference shall thus be driven by finite-sample considerations.

Theorem 2. *Assume that U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$. Assume further that (ψ_n) is a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n, t_n)$, with $0 < \beta_n < t_n \leq 1$, $\beta_n \to 1$, $n(1-\beta_n) \to \infty$ and $(1-t_n)/(1-\beta_n) \to 0$.*

(i) Pick a distortion function g and $a > 0$, and assume that for some $\eta > 0$, we have

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

If furthermore $\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) = \mathrm{O}(1)$ then, provided $\widehat{\gamma}_n$ is a consistent estimator of γ, it holds that:

$$\frac{\widetilde{R}_{g,\beta_n}(X^a;\psi_n)}{\widehat{R}_{g,\beta_n}(X^a;\psi_n)} - 1 \quad \textit{and therefore} \quad \frac{\widetilde{R}_{g,\beta_n}(X^a;\psi_n)}{R_{g,\beta_n}(X^a)} - 1 \xrightarrow{\mathbb{P}} 0.$$

(ii) Pick distortion functions $g_1, \ldots, g_d$ and $a_1, \ldots, a_d > 0$, and assume that for some $\eta > 0$, we have

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

$$\textit{and} \qquad \sqrt{n(1-t_n)}\left(\frac{1-t_n}{1-\beta_n}\right)^{\varepsilon} \to 0,$$

for some $\varepsilon \in (0, \min(1/2, \eta))$. If furthermore

$$\sqrt{n(1-\beta_n)}A((1-\beta_n)^{-1}) \to \lambda \in \mathbb{R} \quad \textit{and} \quad \sqrt{n(1-\beta_n)}(\widehat{\gamma}_n - \gamma) = \mathrm{O}_{\mathbb{P}}(1),$$

then:

$$\forall j \in \{1,\ldots,d\}, \ \sqrt{n(1-\beta_n)}\left(\frac{\widetilde{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}{\widehat{R}_{g_j,\beta_n}(X^{a_j};\psi_n)} - 1\right) \xrightarrow{\mathbb{P}} 0,$$

and therefore

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

with V as in Theorem 1.

It should be noted here that the requirement $\sqrt{n(1-\beta_n)}(\widehat{\gamma}_n - \gamma) = \mathrm{O}_{\mathbb{P}}(1)$ is hardly a restrictive one, for all the aforementioned tail index estimators satisfy such a property in their respective domains of validity under second-order condition $\mathcal{C}_2(\gamma,\rho,A)$, see Sections 3 and 4 in de Haan and Ferreira (2006).

As we mentioned at the end of Section 2, the empirical estimators developed so far only work provided β_n is an intermediate level, namely $n(1-\beta_n) \to \infty$. The next and final step is to design an estimator working for arbitrarily extreme levels as well.

3.4 Final step: estimation in the extreme case

A consistent estimator of an arbitrarily extreme risk measure is now designed by using an extrapolation property of the tail quantile function U. Let (δ_n) be a sequence converging to 1 such that $(1-\delta_n)/(1-\beta_n)$ converges to a positive and finite limit, and remark that for any $s \in (0,1)$ and $a > 0$ it holds that:

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

as $n \to \infty$, as a consequence of the regular variation property of $U(y) = q(1 - y^{-1})$. Integrating the above relationship with respect to the distortion measure dg yields:

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

To put it differently, the extreme risk measure $R_{g,\delta_n}(X^a)$ is essentially obtained by multiplying the intermediate risk measure $R_{g,\beta_n}(X^a)$ by an extrapolation factor depending on the unknown tail index γ. To estimate the left-hand side, suppose then that $n(1 - \delta_n) \to c < \infty$, take a sequence (β_n) such that $n(1 - \beta_n) \to \infty$ and define

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

where $\widehat{\gamma}_n$ is the consistent estimator of γ appearing in $\widetilde{R}_{g,\beta_n}(X^a; \psi_n)$. This is a Weissman-type estimator of $R_{g,\delta_n}(X^a)$ (see Weissman, 1978, for the estimation of extreme quantiles). Weissman's estimator is actually recovered for $a = 1$, $t_n = 1$ and $g(s) = 0$ if $s < 1$, and the extrapolated PL estimator of El Methni and Stupfler (2017) is obtained for $t_n = 1$.

The third and final main result examines the asymptotic distribution of this class of extrapolated estimators.

Theorem 3. *Assume that U satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$, with $\rho < 0$. Assume further that (ψ_n) is a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n, t_n)$, with $0 < \beta_n < t_n \le 1$, $\beta_n \to 1$, $n(1 - \beta_n) \to \infty$ and $(1 - t_n)/(1 - \beta_n) \to 0$; let finally a sequence $\delta_n \to 1$ be such that $(1 - \delta_n)/(1 - \beta_n) \to 0$ and $\log[(1 - \beta_n)/(1 - \delta_n)]/\sqrt{n(1 - \beta_n)} \to 0$. Pick now distortion functions $g_1, \ldots, g_d$ and $a_1, \ldots, a_d > 0$, and assume that for some $\eta > 0$, we have*

$$\forall j \in \{1, \ldots, d\}, \ \int_0^1 s^{-a_j\gamma - 1/2 - \eta} dg_j(s) < \infty \quad \text{and} \quad \sqrt{n(1 - t_n)} \left(\frac{1 - t_n}{1 - \beta_n}\right)^{\varepsilon} \to 0,$$

for some $\varepsilon \in (0, \min(1/2, \eta))$. If furthermore

$$\sqrt{n(1 - \beta_n)} A((1 - \beta_n)^{-1}) \to \lambda \in \mathbb{R} \quad \text{and} \quad \sqrt{n(1 - \beta_n)}(\widehat{\gamma}_n - \gamma) \xrightarrow{d} \xi,$$

then:

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

Again, in the case $t_n = 1$, we recover the asymptotic normality result of El Methni and Stupfler (2017) for the class of extrapolated PL estimators. Our robust extreme risk measure estimators have therefore got the same asymptotic distribution as the original PL estimator, under the same technical conditions. It can thus be concluded that considering trimmed/winsorised estimators results in a generalisation of the existing theory of estimators of extreme Wang DRMs. The next section shall show that this also results in improved finite-sample performance when the underlying distribution has a very heavy tail.

4 Simulation study

The finite-sample performance of our estimators is illustrated on the following simulation study, where a pair of heavy-tailed distributions and three distortion functions g are considered. 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 $\rho < 0$).

These distributions have tail index γ; meanwhile, their respective second-order parameters are -1 and ρ, see e.g. Beirlant et al. (2004). We can therefore get an idea of the influence of the parameters γ and ρ on the finite-sample behaviour of an estimator using these two distributions. In the case of the Burr distribution, we shall take $\rho \in \{-2, -2/3\}$.

The following distortion functions are considered:

- the Tail-Value-at-Risk (TVaR) 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 $\max(X_1, \dots, X_c)$ for independent copies $X_1, \dots, X_c$ 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 $s \downarrow 0$. When $c := 1/\alpha$ is a positive integer, the related DRM is the expectation of a random variable Y whose distribution is such that X has the same distribution as $\min(Y_1, \dots, Y_c)$ for independent copies $Y_1, \dots, Y_c$ of Y. See also Cherny and Madan (2009).

Each risk measure is estimated, at an extreme level δ_n, using the extrapolated estimator $\widetilde{R}^W_{g,\delta_n}(X; \psi_n)$. The following choices of ψ_n are considered:

- $\psi_n(s) = 1 - (1 - \beta_n)s$, corresponding to the PL estimator;
- $\psi_n(s) = t_n - (t_n - \beta_n)s$, corresponding to the corrected Trim-PL estimator, which we denote by CTrim-PL;
- $\psi_n(s) = \min(t_n, 1 - (1 - \beta_n)s)$, corresponding to the corrected Wins-PL estimator, which we denote by CWins-PL.

Because any of the studied estimators uses a preliminary estimation at level β_n where (β_n) is some intermediate sequence, we first discuss the choice of this level. As noted numerous times in the extreme value literature, this is a crucial step: a value of β_n too close to 1 increases the variance of the estimator dramatically, while a value of β_n too far from 1 results in biased estimates. An overview of possible techniques is given in Section 5.4 of Gomes and Guillou (2015). Here a data-driven criterion, based on the search for a stable part of the plot of a tail index estimator and similar to that of El Methni and Stupfler (2017), is used; see also Stupfler (2013), Gardes and Stupfler (2014) and Stupfler (2016) for other implementations. We work with a bias-reduced version $\widehat{\gamma}_{\beta_n}$ of the Hill estimator (Hill, 1975) suggested by Caeiro et al. (2005) (see also Gomes et al., 2016), which shall also be ultimately used to estimate the parameter γ:

$$\widehat{\gamma}_{\beta_n} = H_n(\lceil n(1-\beta_n)\rceil)\left(1 - \frac{\widehat{B}}{1-\widehat{\rho}}\left(\frac{n}{\lceil n(1-\beta_n)\rceil}\right)^{-\widehat{\rho}}\right),$$

$$\text{with} \quad H_n(k) = \frac{1}{k}\sum_{i=1}^{k} \log\left(X_{n-i+1,n}\right) - \log\left(X_{n-k,n}\right).$$

Here $\widehat{B}$ is an estimator of the parameter B such that the left-continuous inverse U of $1/(1-F)$ satisfies

$$U(z) = Cz^{\gamma}\left(1 + \frac{\gamma}{\rho} B z^{\rho} + \mathrm{o}(z^{\rho})\right) \quad \text{as} \quad z \to \infty,$$

and $\widehat{\rho}$ is an estimator of the second-order parameter ρ. In particular, the version of $\widehat{\gamma}_{\beta_n}$ used in the present simulation study is the one implemented in the function `mop` of the R package `evt0` and discussed in Gomes et al. (2016). The idea is now to detect the last stability region in the plot $\beta \mapsto \widehat{\gamma}_{\beta}$. Specifically:

- choose $\beta_0 > 0$ and a window parameter $h_1 > 1/n$;
- for $\beta_0 < \beta < 1 - h_1$, let $I(\beta) = [\beta, \beta + h_1]$ and compute the standard deviation $\sigma(\beta)$ of the set of estimates $\{\widehat{\gamma}_b, b \in I(\beta)\}$;
- if $\beta \mapsto \sigma(\beta)$ is monotonic, let β_{lm} be β_0 if it is increasing and $1 - h_1$ if it is decreasing;
- otherwise, denote by β_{lm} the last value of β such that $\sigma(\beta)$ is locally minimal and its value is less than the average value of the function $\beta \mapsto \sigma(\beta)$;
- choose β^* such that $\widehat{\gamma}_{\beta^*}$ is the median of $\{\widehat{\gamma}_b, b \in I(\beta_{lm})\}$. In particular, our estimate of γ is $\widehat{\gamma}_{\beta^*}$.

Here this choice procedure is conducted with $\beta_0 = 0.5$ and $h_1 = 0.1$. Once the parameter β_n has been chosen as β^*, we can compute the extrapolated PL estimator $\widetilde{R}^W_{g,\delta_n}(X|\beta^*)$ described in El Methni and Stupfler (2017).

In order to compute the extrapolated CTrim-PL and CWins-PL estimators, the truncation/winsorisation level t_n should also be chosen, and this is done by a stability region argument as well, which this time revolves around these extreme risk measure estimators. Here, the dependence of ψ upon β and t is emphasised by denoting it by $\psi(\beta, t)$. The suggested choice procedure for t, given a function ψ and the tuning parameter β^*, is the following:

- choose $t_0 > 0$ and a window parameter $h_2 > 1/n$;
- for $t_0 < t < 1 - h_2$, let $J(t) = [t, t + h_2]$ and compute the standard deviation $\Sigma(t)$ of the set of estimates $\{\widetilde{R}^W_{g,\delta_n}(X; \psi_n(\beta^*, \theta)), \theta \in J(t)\}$;
- if $t \mapsto \Sigma(t)$ is monotonic, let t_{lm} be t_0 if it is increasing and $1 - h_2$ if it is decreasing;

- otherwise, denote by t_{lm} the last value of t such that $\Sigma(t)$ is locally minimal and its value is less than the average value of the function $t \mapsto \Sigma(t)$;
- choose t^* such that $\widetilde{R}^W_{g,\delta_n}(X; \psi_n(\beta^*, t^*))$ is the median of

$$\{\widetilde{R}^W_{g,\delta_n}(X; \psi_n(\beta^*, \theta)), \theta \in J(t_{lm})\}.$$

In the present simulation study, this choice procedure is conducted with $t_0 = 0.95$ and $h_2 = 0.01$.

The idea is now to compare the performance of the PL estimator of El Methni and Stupfler (2017) to that of the CTrim-PL and CWins-PL estimators, first in the case of moderately heavy tails, when the PL estimator is known to have reasonable theoretical and finite-sample properties, and then in the case of very heavy tails, in order to illustrate the advantages of using the proposed technique. It will in particular be shown that, compared to the PL estimator which uses all the data above a high threshold, the corrected trimmed or winsorised estimators resist fairly well to the presence of heavier tails and atypically high observations. It will also be of interest to compare the performance of the CTrim-PL and CWins-PL estimators, and in particular to assess whether one of them is preferable to the other in terms of bias: recall that before correction, the trimmed estimator should be expected to have worse finite-sample performance that the winsorised estimator.

4.1 Case 1: Moderately heavy tails

We first consider the case of moderately heavy tails. More precisely, the parameter γ is chosen in order to ensure that Theorem 2 (ii) applies to the intermediate versions of all three estimators, and is therefore such that the extrapolated estimators satisfy Theorem 3. This range of values of γ is considered in El Methni and Stupfler (2017), and it is shown there that the extrapolated PL estimator performs reasonably well when γ is moderate. We will, however, consider a range of values of γ containing the highest values of γ for which Theorem 3 applies, in order to assess the behaviour of all three estimators on the full range of moderately heavy tails and particularly in the most difficult situations in this range. The following examples are considered:

- the TVaR and DP(1/3) risk measures. In this case, Theorem 3 applies in the range $\gamma \in (0, 1/2)$. We therefore make γ vary in the interval $[0.25, 0.49]$ for both our tested distributions.
- the PH(1/2) risk measure. Here, Theorem 3 applies in the range $\gamma \in (0, 1/4)$. We therefore choose to have γ vary in the interval $[0.1, 0.24]$.

In each case, the computations are carried out on $N = 5000$ independent samples of $n = 1000$ independent copies of X; a similar simulation study, whose results are deferred to Appendix C in the supplementary material document, examines the case of the lower sample size $n = 100$. Relative biases and relative mean squared errors (MSEs) are recorded:

$$\text{Bias}\left(\widetilde{R}^{W}_{g,\delta}\right) = \frac{1}{N}\sum_{j=1}^{N} \frac{\widetilde{R}^{W}_{g,\delta}(X;\psi_j^*)}{R_{g,\delta}(X)} - 1,$$

$$\text{and} \qquad \text{MSE}\left(\widetilde{R}^{W}_{g,\delta}\right) = \frac{1}{N}\sum_{j=1}^{N} \left(\frac{\widetilde{R}^{W}_{g,\delta}(X;\psi_j^*)}{R_{g,\delta}(X)} - 1\right)^2,$$

at $\delta = 0.999 = 1 - n^{-1}$ (here ψ_j^* is the chosen function ψ for the $j-$th sample and for a given estimator), so as to be able to assess both bias and variability of all the compared techniques.

Results are reported in Figures 1 and 2. Results for the extreme DP risk measure were qualitatively very similar to those obtained for the extreme TVaR and are therefore not reported here. As regards the estimation of the extreme TVaR, it appears on these examples that the proposed CTrim-PL and CWins-PL estimators perform slightly worse in terms of bias than the original PL estimator. This is not surprising: the correction method, based on an approximation of the upper tails of the underlying distribution by a purely Pareto tail, cannot be expected to recover all the information the (highly variable) top order statistics carry about the extremes of the sample. By contrast, our estimators perform essentially comparably to or better than the standard empirical extreme Wang DRM estimator in terms of MSE; for values of γ close to but less than 1/2, the improvement is close to up to 40%, in the case of the Fréchet distribution. Surprisingly, for $|\rho| \geq 1$, the CTrim-PL and CWins-PL estimators seem to provide a much improved technique for the estimation of the extreme PH risk measure, both in terms of bias and MSE, especially when γ is large. Let us also mention that in all cases, results deteriorate when γ increases: this is likely a consequence of the fact that, by Theorem 3, the asymptotic distribution of our estimator is essentially that of $\widehat{\gamma}_{\beta_n} - \gamma$, which is a Gaussian distribution with variance proportional to γ^2 (see Theorem 3.2 in Caeiro et al., 2005). Similarly the results, be it with respect to bias or MSE, also improve when $|\rho|$ increases, which is not surprising either since the larger is $|\rho|$, the smaller is the bias in the estimation and, more generally, the closer is the tail of the underlying distribution to a purely Pareto tail. This is especially critical for the CTrim-PL and CWins-PL estimators, in which the correction step is based on an approximation of the right tail of the underlying distribution by the right tail of (a multiple of) a Pareto distribution. Combined with the stronger emphasis the PH risk measure puts on higher quantiles of the underlying distribution, this explains the deterioration, in terms of finite-sample performance, that the CTrim-PL and CWins-PL estimators suffer from in the case of the Burr distribution with $\rho = -2/3$ relatively to the PL estimator and compared to the other cases considered here.

It should finally be underlined that, on these examples and for smaller values of γ (e.g. in the case of TVaR estimation, $\gamma = 1/4$, corresponding essentially to the existence of a finite fourth moment) the extrapolated PL, CTrim-PL and CWins-PL estimators have virtually indistinguishable finite-sample performance. There seems, therefore, to be no loss in efficiency when using the proposed estimators for small values of γ, while they display an appreciably lower variability for an arguably small potential price in terms of bias when γ is larger and $|\rho|$ is not too small. It also appears that on these examples, for moderately heavy tails, the CTrim-PL and CWins-PL estimators have very similar finite-sample behaviours, so there is on average no clear advantage in using one of these methods over the other.

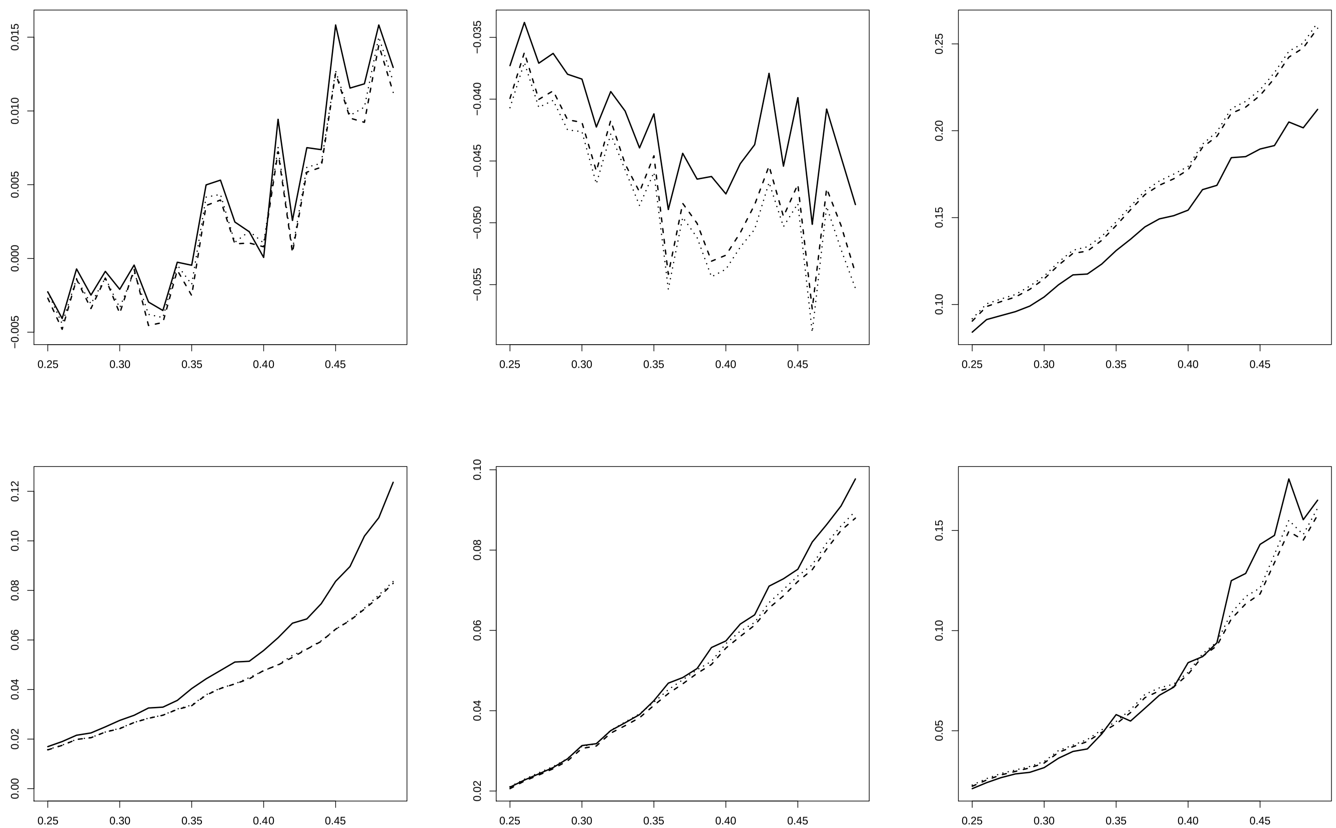

Figure 1: Extreme TVaR estimation with $\delta = 0.999$, for $\gamma \in [0.25, 0.49]$ and on $N = 5000$ replications of a sample of size $n = 1000$. Top panels: relative bias, bottom panels: relative MSE. Left: case of the Fréchet distribution, middle: case of the Burr distribution with $\rho = -2$, right: case of the Burr distribution with $\rho = -2/3$. Full line: PL estimator, dotted line: CTrim-PL estimator, dashed line: CWins-PL estimator.

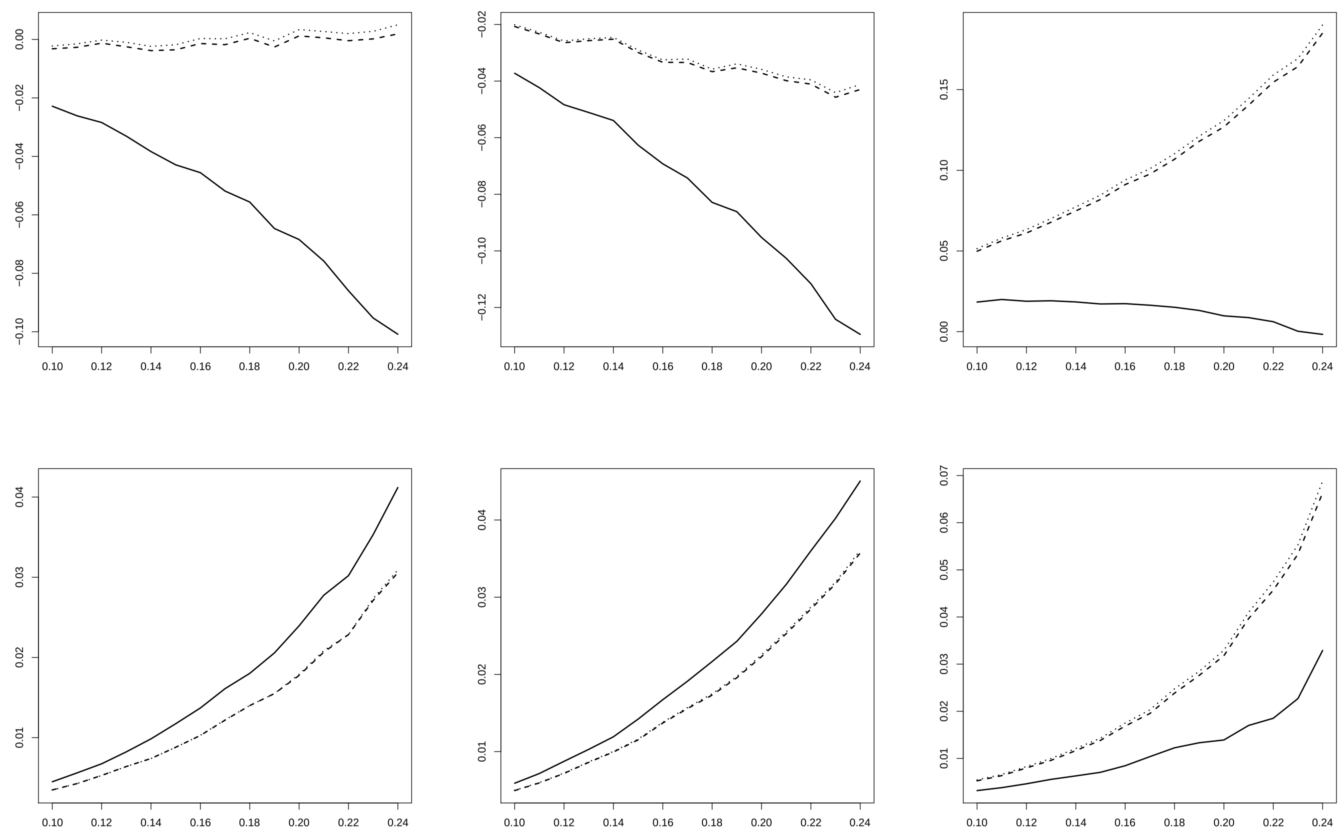

Figure 2: Extreme PH(1/2) risk measure estimation with $\delta = 0.999$, for $\gamma \in [0.1, 0.24]$ and on $N = 5000$ replications of a sample of size $n = 1000$. Top panels: relative bias, bottom panels: relative MSE. Left: case of the Fréchet distribution, middle: case of the Burr distribution with $\rho = -2$, right: case of the Burr distribution with $\rho = -2/3$. Full line: PL estimator, dotted line: CTrim-PL estimator, dashed line: CWins-PL estimator.

4.2 Case 2: Very heavy tails

We now consider the case of heavier tails, when Theorem 3 fails to hold. Studying such cases will make it possible to understand the behaviour of the tested estimators on more challenging situations. Specifically, the following cases are examined:

- the TVaR and DP(1/3) risk measures, with γ varying in the interval $[0.5, 0.75]$ for both our tested distributions.
- the PH(1/2) risk measure, with γ belonging to the interval $[0.25, 0.35]$.

In those cases, sample relative MSEs can still be computed but will not converge anymore, because γ is so large that the relative MSE of our estimators at an intermediate level is infinite:

$$\mathbb{E}\left(\frac{\widetilde{R}_{g,\beta_n}(X;\psi_n)}{R_{g,\beta_n}(X)} - 1\right)^2 = +\infty.$$

In order to assess both bias and variability here, we therefore look at two different situations:

(i) In the first one, $N = 5000$ independent samples of $n = 1000$ independent copies of X are generated and relative biases are recorded for all three estimators.

(ii) In the second one, $N = 5000$ independent samples of $n = 1000$ independent copies of X given that the sample maximum $X_{n,n}$ exceeds the large value $2R_{g,\delta}(X)$, with $\delta = 0.999$, are generated. Again, relative biases are recorded for all three estimators.

The idea here is to first use (i) to assess to which extent the correction factor for the trimmed/winsorised estimators manages to eliminate the bias introduced by the trimming/winsorising scheme, and then to evaluate the advantages, in terms of variability, of using the proposed techniques in challenging cases using (ii). It should be mentioned that although the cases examined in (ii) are in some sense atypical, they are not at all infrequent: for instance, in the case of the Tail-Value-at-Risk for the Fréchet distribution with parameter $\gamma = 1/2$ and $\delta = 0.999$, then $R_{g,\delta}(X) \approx 63.25$ and

$$\mathbb{P}(X_{n,n} > 2R_{g,\delta}(X)) = 1 - [\mathbb{P}(X \leq 2R_{g,\delta}(X))]^n \approx 0.0606,$$

with $n = 1000$. In other words, 6% of samples of size 1000 feature the difficulty considered here, which we believe makes it well worth studying.

As in the previous section, the extreme level of interest is $\delta = 0.999 = 1 - n^{-1}$. Again, a similar simulation study, whose results are deferred to Appendix C in the supplementary material document, considers the case $n = 100$.

Results are reported in Figures 3 and 4, the top panels representing biases recorded in non-conditioned cases and the bottom panels representing biases obtained in the difficult conditioned cases. As in the previous study on moderate tails, results for the extreme DP risk measure were qualitatively very similar to those obtained for the extreme TVaR and are therefore not reported here. When estimating the extreme TVaR in standard cases, the CTrim-PL and CWins-PL estimators perform slightly better, in terms of bias, than the original PL estimator when $|\rho| \geq 1$, although the PL estimator outperforms the CTrim-PL and CWins-PL estimators when $|\rho|$ takes the smaller value $2/3$. By contrast, a real improvement is found using the suggested methods on atypical cases: in this context, the CTrim-PL and CWins-PL estimators more than halve the bias overall in the case of the Fréchet distribution, can reduce it by up to 90% in the case of the Burr distribution with $\rho = -2$, and improve it by up to 50% for very large γ when $\rho = -2/3$. As regards the estimation of the extreme PH risk measure, the surprising conclusion reached when discussing the performance of our estimators with moderately heavy tails is still valid: for $|\rho| \geq 1$, the CTrim-PL and CWins-PL estimators appear to have a much lower bias than the PL estimator in general, all the more so for larger values of γ. There is again a marked improvement in terms of bias in atypical cases, the bias being halved overall in the Fréchet case and in the Burr case with $\rho = -2/3$, the reduction in bias being even more substantial in the Burr case with $\rho = -2$.

As a conclusion, it appears on these heavier-tailed examples that the CTrim-PL and CWins-PL have generally comparable performance to that of the PL estimator in the case of extreme TVaR and DP estimation, while they often provide a significant improvement when estimating the extreme PH risk measure. Moreover, in the most difficult cases with respect to the behaviour of the top order statistics in the sample, the two introduced methods represent overall a great improvement over the PL estimator. It should be pointed out that in these atypical cases, the deterioration of the finite-sample performance of our estimators relatively to that of the PL estimator when $|\rho|$ decreases is much less severe than in the case when the right tail of the underlying distribution is moderately heavy. An explanation is that while the correction factor applied to the Trim-PL or Wins-PL estimator might have a disappointing behaviour when $|\rho|$ is small, the action of deleting unreasonably high top values in the sample and then correcting at least partially for the resulting bias is already enough to obtain a much-improved technique. We would therefore argue that the CTrim-PL and CWins-PL estimators have indeed a good potential for practical use in such a setup, and this was the main goal of our work. Finally, on these examples and similarly to the moderate tails case, the CTrim-PL and CWins-PL estimators exhibit similar finite-sample behaviours, so that there is no obvious reason to choose one over the other in general. Which one of these estimators should actually be chosen has to be decided case by case, and an instance of such a choice is presented in a real data example below.

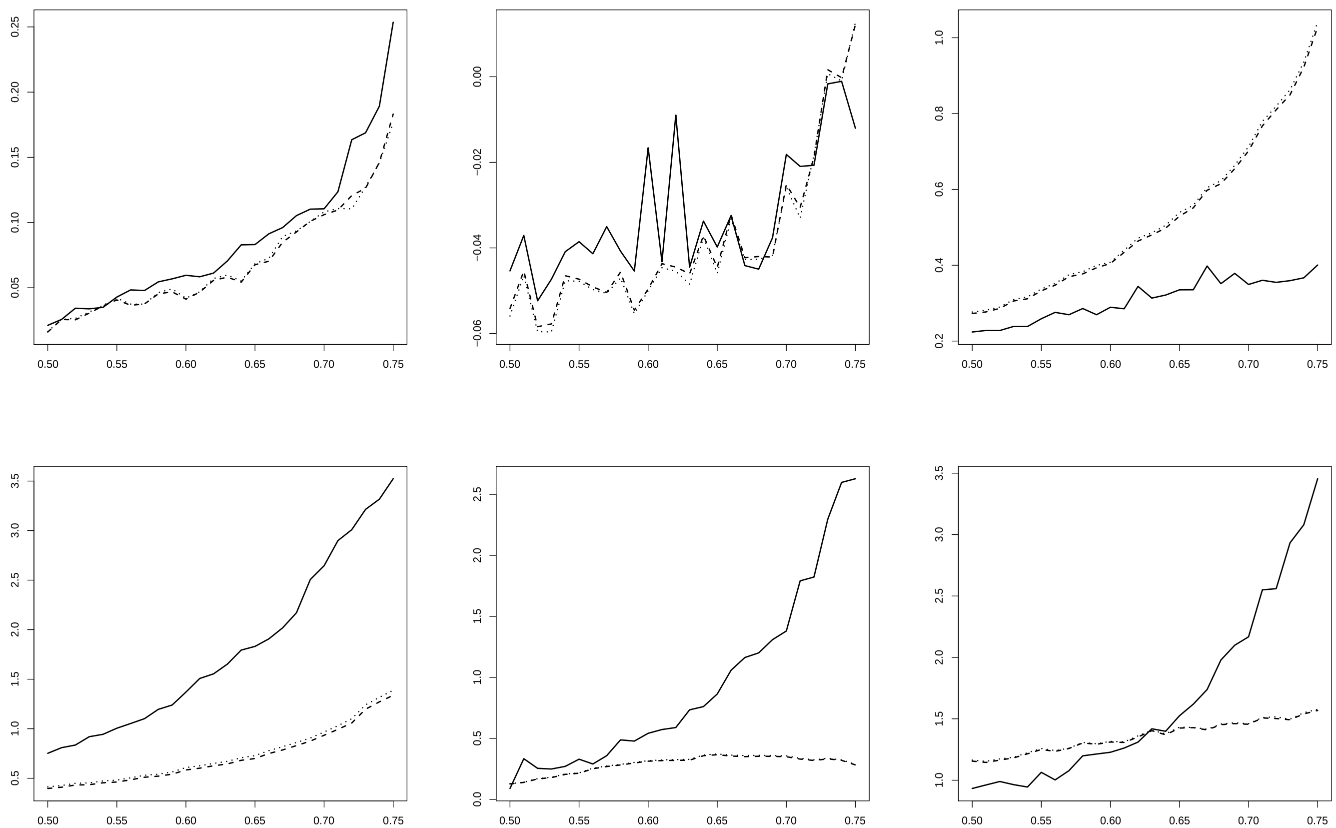

Figure 3: Extreme TVaR estimation with $\delta = 0.999$, for $\gamma \in [0.5, 0.75]$ and on $N = 5000$ replications of a sample of size $n = 1000$. Top panels: relative bias in non-conditioned cases, bottom panels: relative bias in conditioned cases. Left: case of the Fréchet distribution, middle: case of the Burr distribution with $\rho = -2$, right: case of the Burr distribution with $\rho = -2/3$. Full line: PL estimator, dotted line: CTrim-PL estimator, dashed line: CWins-PL estimator.

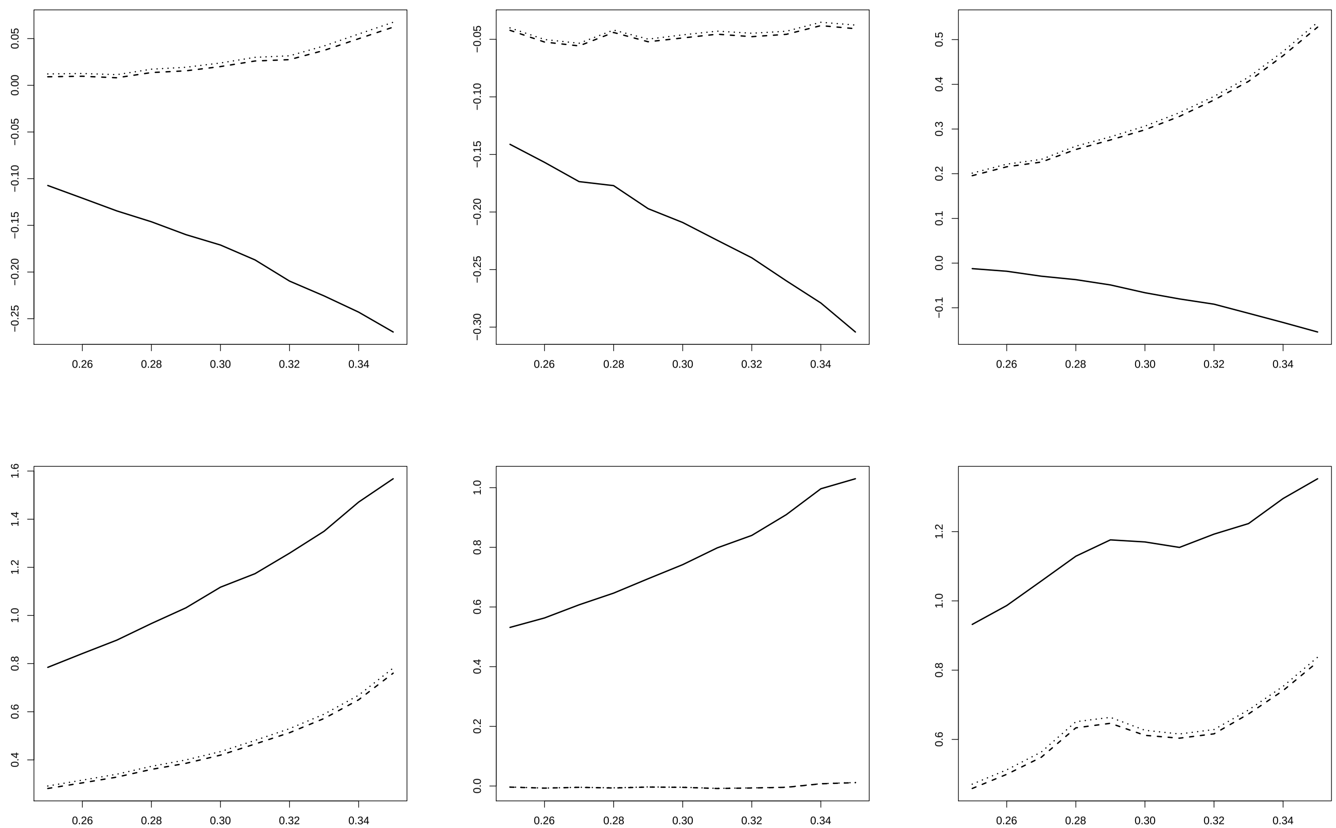

Figure 4: Extreme PH(1/2) risk measure estimation with $\delta = 0.999$, for $\gamma \in [0.25, 0.35]$ and on $N = 5000$ replications of a sample of size $n = 1000$. Top panels: relative bias in non-conditioned cases, bottom panels: relative bias in conditioned cases. Left: case of the Fréchet distribution, middle: case of the Burr distribution with $\rho = -2$, right: case of the Burr distribution with $\rho = -2/3$. Full line: PL estimator, dotted line: CTrim-PL estimator, dashed line: CWins-PL estimator.

5 Real data application

We consider data on $n = 1098$ commercial fire losses recorded between 1 January 1995 and 31 December 1996 by the FFSA (an acronym for the *Fédération Française des Sociétés d'Assurance*), available from the R package `CASdatasets` by prompting `data(frecomfire)`. The data, originally recorded in French francs, is converted into euros, and denoted by $(X_1, \ldots, X_n)$. The analysis of this kind of data set from an extreme value point of view is useful to insurers, especially in view of the Solvency II directive: in order to be able to compute their capital requirements so as to survive the upcoming calendar year with a probability not less than 0.995, insurance companies have to take into account extremely high losses. It is also crucial for insurance companies to estimate the capital requirement as accurately as possible: an underestimation of this quantity can threaten the company's survival, while an overestimation may, among others, lead to the insurer asking for higher premiums on policies, thus reducing the company's competitiveness on the market.

The first step is to estimate the tail index γ. To this end, the procedure outlined in Section 4 is used: the sample fraction chosen to compute the tail index is then $1 - \beta^* \approx 0.120$, for an estimate $\widehat{\gamma}_{\beta^*} \approx 0.697$. This suggests a very heavy tail, in the sense that $\widehat{\gamma}_{\beta^*} > 1/2$ and therefore the underlying distribution seems to have an infinite variance. In particular, we know from the simulation study that this may adversely affect the PL estimator of the extreme TVaR and of the extreme DP risk measure, which justifies comparing the PL estimates to those obtained using our CTrim-PL and CWins-PL estimators. Note that the extreme PH(1/2) risk measure cannot be estimated here since this would require the estimate of γ to be less than 1/2.

We then compute, at the extreme level $\delta = 0.999 \approx 1 - n^{-1}$, the PL, CTrim-PL and CWins-PL estimators of the extreme TVaR and DP(1/3) risk measures, using the procedure of Section 4. Results are summarised in Table 1. It is not clear, from these results, which estimator should be chosen, especially since it was seen in the simulation study that the CTrim-PL and CWins-PL estimators have essentially identical statistical properties.

Our goal is now to offer some insight into this choice, using the mean excess plot of the $n(1-\beta^*) = 132$ data points used in the present analysis. The rationale behind the use of the mean excess plot, i.e. the plot of the function

$$u \mapsto \frac{\sum_{i=1}^{n} (X_i - u)\mathbb{I}_{\{X_i > u\}}}{\sum_{i=1}^{n} \mathbb{I}_{\{X_i > u\}}},$$

is that its empirical counterpart $u \mapsto \mathbb{E}(X - u | X > u)$ is linearly increasing when $0 < \gamma < 1$ and X has a Generalised Pareto distribution (see Davison and Smith, 1990). Therefore, since X can be, above a high level u, approximated by a Generalised Pareto distribution (see e.g. equation (3.1.2) p.65 in de Haan and Ferreira, 2006), the extremes of the data set should be indicated by a roughly linear part at the right of the mean excess plot. This plot can be tricky to use though: apart from the choice of the lower threshold u above which the mean excess function is computed (which is here chosen to be $X_{n\beta^*,n}$), it has been observed that the mean excess function has very often a non-linear behaviour at the right end of the mean excess plot (see Ghosh and Resnick, 2010). This is again because the top order statistics in the sample suffer from a very high variability, and as a consequence the mean excess function is, in its right end, averaging over just a few high-variance values. In other words, the intermediate, roughly linear part of the plot indicates which ones among the top data points can be trusted from the points of view of both bias and variability, and the unstable part at the right end of the mean excess plot represents those highly variable values that may be cut from the analysis using the CTrim-PL and CWins-PL estimators.

We then plot on Figure 5 copies of the mean excess plot above the value $u = X_{n\beta^*,n}$ where the values cut from the analysis by the CTrim-PL and CWins-PL estimators are highlighted. The least squares line related to the data points kept for the analysis is also represented. It can be seen on these plots that there is indeed an unstable part at the right end of the plot, which suggests to use either the CTrim-PL or CWins-PL estimator in order to gain some stability. The linear adjustment for the selected data points is also reasonable in all cases. It is arguable though that the CTrim-PL estimator is too conservative in the sense that the number of data points it discards is high: in the DP case in particular, the estimator trims 37 top order statistics, which is 29% of the available data above the selected threshold $X_{n\beta^*,n}$. The CWins-PL estimator discards much less data points (less than half of what the CTrim-PL estimator discards, see also Table 1), and therefore does not have to compensate for the loss of information this entails as much as the CTrim-PL has to, while the linear adjustment of the least squares line is still perfectly acceptable. It can be argued then that the CWins-PL estimator is preferable here, both for extreme TVaR and DP risk measure estimation. The estimates it yields are appreciably lower (roughly 10% less) than the standard PL estimates, and this makes us think that the extreme TVaR and DP risk measure are actually overestimated by the PL estimator. The conclusion is that, using the CWins-PL estimator, the average loss in the worst 0.1% of cases is estimated to be 208.5 million euros, and the average value of the maximal loss recorded after three extreme fires (i.e. each belonging to the worst 0.1% of fires) to be 404.5 million euros.

	Estimator	Order t^*	Number of top order statistics cut	Estimate (in euros)
TVaR estimation	PL	N/A	N/A	225,122,925
	CTrim-PL	0.9736	29	219,814,856
	CWins-PL	0.9909	10	208,538,799
DP(1/3) estimation	PL	N/A	N/A	459,285,394
	CTrim-PL	0.9663	37	452,920,888
	CWins-PL	0.9863	15	404,498,511

Table 1: French commercial fire losses data set: estimating some risk measures in the case $\delta = 0.999$.

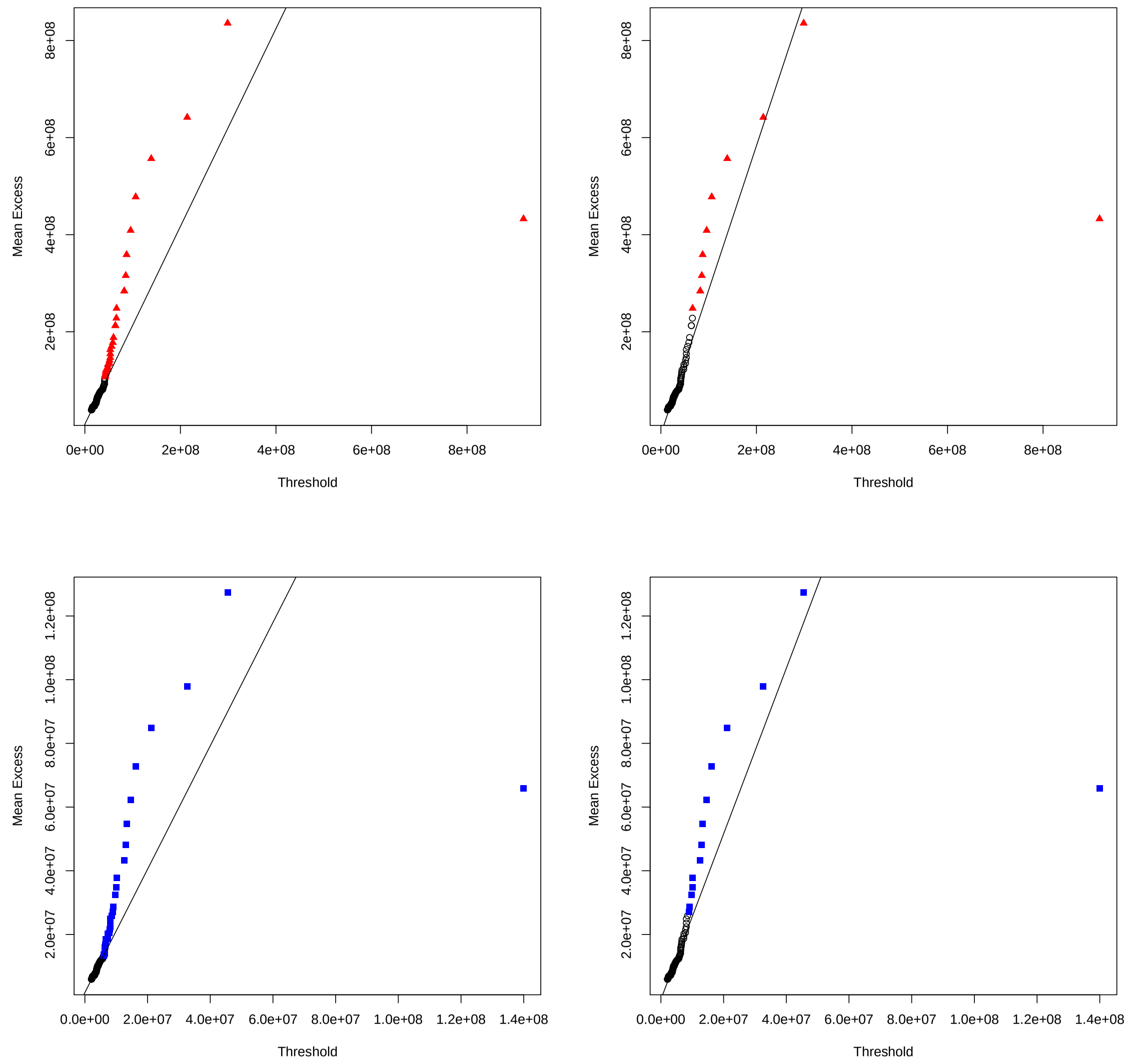

Figure 5: Mean excess plots for the French commercial fire losses data set. Top: the mean excess plot of the top $n(1-\beta^*) = 132$ data points, where data points excluded from the TVaR estimation when using the CTrim-PL and CWins-PL estimators are highlighted using red triangles; left: CTrim-PL estimator, right: CWins-PL estimator. Bottom: the mean excess plot of the top $n(1-\beta^*) = 132$ data points, where data points excluded from the DP(1/3) estimation when using the CTrim-PL and CWins-PL estimators are highlighted using blue squares; left: CTrim-PL estimator, right: CWins-PL estimator. In all four cases the straight line is the least squares line for the set of black data points.

6 Discussion and forthcoming studies

In this paper we studied, empirically and theoretically, corrected versions of the trimmed/winsorised empirical plug-in estimators of extreme Wang DRMs. The value of such estimators lies in the fact that on average, they can be expected to be less variable than the full empirical plug-in estimator thanks to the trimming/winsorising scheme. The proposed method is also implemented on a real actuarial data set, and it is shown how to choose between the corrected trimmed estimator and its corrected winsorised analogue.

Let us reiterate here that the present correction step is a simple one, based on an asymptotic equivalent of the ratio of a trimmed/winsorised extreme Wang DRM and of its full counterpart. Especially, this correction method should not, in our view, be seen as related to bias-correction techniques based on asymptotic results developed in second-order extreme value frameworks, which have been the subject of much interest in extreme value theory in the past twenty years. An essential difference between the two approaches is that the proposed correction step does not take into account second-order information: namely, it uses an estimator of the tail index γ but no estimator of the second-order parameter ρ.

This is why the CTrim-PL and CWins-PL methods should not be expected to show an improved finite-sample performance compared to that of the basic empirical plug-in estimator when ρ is close to 0. Actually, because the expression of the correction factor is based on the asymptotic approximation of the underlying distribution by a multiple of a Pareto distribution, which is known to be poor for ρ close to 0, the CTrim-PL and CWins-PL methods should only be expected to work well when $|\rho|$ *is not too small.* This is confirmed in the simulation study, by noting that in typical cases the CTrim-PL and CWins-PL estimators do on average suffer from a deterioration in performance, relatively to the PL estimator, when $|\rho|$ decreases towards 0. It should be repeated though that simulation results give a strong indication that the CTrim-PL and CWins-PL estimators very often bring an important improvement, including in cases when $|\rho|$ is small, upon the PL estimator in challenging cases when the top values in the sample are extremely high, and that was the main goal of this paper.

It would be very interesting to design another correction factor taking into account second-order information, in order to close the gap between the finite-sample performance of the proposed technique and that of the full PL estimator in standard cases with low $|\rho|$, and retain or even improve its finite-sample performance further in difficult cases. Two reasons why this is a difficult problem are that:

- estimators of ρ typically have a rate of convergence $\sqrt{k}A(n/k)$, with the notation of condition $\mathcal{C}_2(\gamma, \rho, A)$, which, in conjunction with the bias conditions they have to satisfy, makes their rate of convergence lower than that of typical tail index estimators, see *e.g.* p.298 in Gomes et al. (2009) and p.2638 in Goegebeur et al. (2010). This suggests that estimators of the second-order parameter are in general quite volatile;

- tail index estimators tend to have a poor finite-sample behaviour for low $|\rho|$, be it because of their bias if they are not bias-corrected, or of their increased asymptotic variance if they are bias-corrected.

Multiplying the Trim-PL or Wins-PL estimators by a correction factor adapted to low values of $|\rho|$ might therefore entail multiplying by a highly variable quantity and ultimately wipe out part of or all that was gained in terms of variability from using the trimming/winsorising scheme. The problem of constructing a second-order-adapted correction factor is therefore a challenging one and is definitely part of future research on this topic.

Acknowledgements

The authors acknowledge the Associate Editor and two anonymous referees for their helpful comments which led to an improved presentation of the results in this article.

References

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

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

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

Belles-Sampera, J., Guillén, M., Santolino, M., 2014. GlueVaR risk measures in capital allocation applications. Insurance: Mathematics and Economics 58, 132–137.

Caeiro, F., Gomes, M., Pestana, D., 2005. Direct reduction of bias of the classical Hill estimator. REVSTAT 3 (2), 113–136.

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

Csörgő, M., Csörgő, S., Horváth, L., Mason, D., 1986a. Normal and stable convergence of integral functions of the empirical distribution function. Annals of Probability 14 (1), 86–118.

Csörgő, S., Horváth, L., Mason, D., 1986b. What portion of the sample makes a partial sum asymptotically stable or normal? Probability Theory and Related Fields 72 (1), 1–16.

Davison, A., Smith, R., 1990. Models for exceedances over high thresholds. Journal of the Royal Statistical Society: Series B 52 (3), 393–442.

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

Deme, E., Girard, S., Guillou, A., 2013. Reduced-bias estimator of the Proportional Hazard Premium for heavy-tailed distributions. Insurance: Mathematics and Economics 52 (3), 550–559.

Deme, E., Girard, S., Guillou, A., 2015. Reduced-biased estimators of the Conditional Tail Expectation for heavy-tailed distributions. In: Hallin, M., Mason, D., Pfeifer, D., Steinebach, J. (Eds.), Mathematical Statistics and Limit Theorems. Springer, pp. 105–123.

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

Dowd, K., Blake, D., 2006. After VaR: the theory, estimation, and insurance applications of quantile-based risk measures. Journal of Risk and Insurance 73 (2), 193–228.

Drees, H., 1998. On smooth statistical tail functionals. Scandinavian Journal of Statistics 25 (1), 187–210.

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. Annals of Applied Probability 14 (3), 1179–1201.

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

El Methni, J., Stupfler, G., 2017. Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. Statistica Sinica (to appear).
URL `http://dx.doi.org/10.5705/ss.202015.0460`

Etemadi, N., 1981. An elementary proof of the strong law of large numbers. Zeitschrift für Wahrscheinlichkeitstheorie and verwandte Gebiete 55 (1), 119–122.

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. Stochastic Processes and their Applications 120 (8), 1492–1517.

Goegebeur, Y., Beirlant, J., de Wet, T., 2010. Kernel estimators for the second order parameter in extreme value statistics. Journal of Statistical Planning and Inference 140 (9), 2632–2652.

Gomes, M., Brilhante, M., Pestana, D., 2016. New reduced-bias estimators of a positive extreme value index. Communications in Statistics – Simulation and Computation 45 (3), 833–862.

Gomes, M., Guillou, A., 2015. Extreme value theory and statistics of univariate extremes: a review. International Statistical Review 83 (2), 263–292.

Gomes, M., Pestana, D., Caeiro, F., 2009. A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. Statistics and Probability Letters 79 (3), 295–303.

Hill, B., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3 (5), 1163–1174.

Hosking, J., Wallis, J., Wood, E., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27 (3), 251–261.

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

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.

Necir, A., Meraghni, D., 2009. Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts. Insurance: Mathematics and Economics 45 (1), 49–58.

Necir, A., Rassoul, A., Zitikis, R., 2010. Estimating the conditional tail expectation in the case of heavy-tailed losses. Journal of Probability and Statistics 2010 (Article ID 596839), 1–17.

Peng, L., 1998. Asymptotically unbiased estimators for the extreme-value index. Statistics and Probability Letters 38 (2), 107–115.

Pickands, J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3 (1), 119–131.

Rootzén, H., Tajvidi, N., 1997. Extreme value statistics and wind storm losses: a case study. Scandinavian Actuarial Journal 1, 70–94.

Smith, R., 1987. Estimating tails of probability distributions. The Annals of Statistics 15 (3), 1174–1207.

Stigler, S., 1973. Simon Newcomb, Percy Daniell, and the history of robust estimation 1885-1920. Journal of the American Statistical Association 68 (344), 872–879.

Stupfler, G., 2013. A moment estimator for the conditional extreme-value index. Electronic Journal of Statistics 7, 2298–2343.

Stupfler, G., 2016. Estimating the conditional extreme-value index under random right-censoring. Journal of Multivariate Analysis 144, 1–24.

Vandewalle, B., Beirlant, J., 2006. On univariate extreme value statistics and the estimation of reinsurance premiums. Insurance: Mathematics and Economics 38 (3), 441–459.

Wang, S., 1995. Insurance pricing and increased limits ratemaking by proportional hazards transforms. Insurance: Mathematics and Economics 17 (1), 43–54.

Wang, S., 1996. Premium calculation by transforming the layer premium density. ASTIN Bulletin 26 (1), 71–92.

Wang, S., 2004. CAT bond pricing using probability transforms. In: Geneva Papers, special issue on insurance and the state of the art in CAT bond pricing. Vol. 278, pp. 19–29.

Wang, S., Young, V., Panjer, H., 1997. Axiomatic characterization of insurance prices. Insurance: Mathematics and Economics 21 (2), 173–183.

Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. Journal of the American Statistical Association 73 (364), 812–815.

Yang, F., 2015. First- and second-order asymptotics for the tail distortion risk measure of extreme risks. Communication in Statistics – Theory and Methods 44 (3), 520–532.

Supplementary material to: Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions

This supplementary material collects the proofs of our auxiliary and main results, in Appendix A and B respectively, and presents additional simulation results in Appendix C.

In all what follows, let F_a be the cdf of X^a and denote by $U_a(x) := [U(x)]^a$ the left-continuous inverse of $1/(1-F_a)$. It is crucial for our purpose to note that by Lemma 1 in El Methni and Stupfler (2017), if U satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$, then U_a satisfies condition $\mathcal{C}_2(a\gamma,\rho,aA)$. As a consequence, Theorem 2.3.9 in de Haan and Ferreira (2006) states that one may find a Borel measurable function B_a, asymptotically equivalent to aA and having constant sign, such that for any η, $\varepsilon > 0$, there is $t_0 > 0$ such that for $t, tx \geq t_0$:

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

Appendix A: Preliminary results and their proofs

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

Lemma 1. *Let g be a distortion function. Assume that f is a Borel measurable regularly varying function with index $b \in \mathbb{R}$, that w is a continuous and bounded function on $(0,1]$ and that (φ_n) is a sequence of Borel measurable positive functions such that*

$$\forall s \in [0,1], \ \ s \leq \varphi_n(s) \leq 1 \quad \text{and} \quad \varphi_n(s) \to s \quad \text{as} \quad n \to \infty.$$

If for some $\eta > 0$:

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

then we have the following, provided $k = k(n) \to \infty$ and $k/n \to 0$: for any $\delta \in \mathbb{R}$ such that $\delta < \eta$,

$$\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)} [\varphi_n(s)]^{-\delta} w(\varphi_n(s)) dg(s) \to \int_0^1 s^{-b-\delta} w(s) dg(s). \tag{i}$$

Moreover,

$$\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)} \log(\varphi_n(s)) w(\varphi_n(s)) dg(s) \to \int_0^1 s^{-b} \log(s) w(s) dg(s). \tag{ii}$$

Proof of Lemma 1. We start by proving (i). Pick $\delta < \eta$ and define $\varepsilon := (\eta - \delta)/2 > 0$, so that $\delta + \varepsilon < \eta$. 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):

$$r_n := \sup_{0<s\le 1} [\varphi_n(s)]^{b+\varepsilon} \left| \frac{f(n/k\varphi_n(s))}{f(n/k)} - [\varphi_n(s)]^{-b} \right| \le \sup_{0<s\le 1} s^{b+\varepsilon} \left| \frac{f(n/ks)}{f(n/k)} - s^{-b} \right|$$
$$= \sup_{t\ge 1} \left| \frac{f_1(tn/k)}{f_1(n/k)} - t^{-\varepsilon} \right| \to 0.$$

Thus, if C is an upper bound for $|w|$ on $(0,1]$:

$$\left| \int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)} [\varphi_n(s)]^{-\delta} w(\varphi_n(s)) dg(s) - \int_0^1 [\varphi_n(s)]^{-b-\delta} w(\varphi_n(s)) dg(s) \right|$$
$$\le r_n \int_0^1 [\varphi_n(s)]^{-b-\delta-\varepsilon} |w(\varphi_n(s))| dg(s) \le r_n \int_0^1 C \max(s^{-b-\delta-\varepsilon}, 1) dg(s) \to 0. \tag{2}$$

Besides

$$\forall s \in (0,1],\ [\varphi_n(s)]^{-b-\delta} |w(\varphi_n(s))| \le C \max(s^{-b-\delta}, 1)$$

and the right-hand side defines an integrable function with respect to the measure $dg(\cdot)$, so that by the dominated convergence theorem:

$$\left| \int_0^1 [\varphi_n(s)]^{-b-\delta} w(\varphi_n(s)) dg(s) - \int_0^1 s^{-b-\delta} w(s) dg(s) \right| \to 0. \tag{3}$$

Combining (2) and (3) completes the proof of (i). To show (ii), the idea is similar: pick $\varepsilon \in (0, \eta/2)$ and write that, with the notation of the proof of part (i):

$$\left| \int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)} \log(\varphi_n(s)) w(\varphi_n(s)) dg(s) - \int_0^1 [\varphi_n(s)]^{-b} \log(\varphi_n(s)) dg(s) \right|$$
$$\le r_n \int_0^1 [\varphi_n(s)]^{-b-\varepsilon} |\log(\varphi_n(s))| |w(\varphi_n(s))| dg(s)$$
$$\le r_n \int_0^1 K [\varphi_n(s)]^{-b-2\varepsilon} dg(s) \le r_n \int_0^1 K \max(s^{-b-2\varepsilon}, 1) dg(s) \to 0. \tag{4}$$

Here K is a positive constant, and we used the fact that the function $t \mapsto t^{\varepsilon} \log t$ is bounded on $(0,1]$. Using this very same fact again we get:

$$\forall s \in (0,1],\ [\varphi_n(s)]^{-b} |\log(\varphi_n(s))| |w(\varphi_n(s))| \le K \max(s^{-b-\varepsilon}, 1)$$

and the right-hand side defines an integrable function with respect to the measure $dg(\cdot)$, so that by the dominated convergence theorem:

$$\left| \int_0^1 [\varphi_n(s)]^{-b} \log(\varphi_n(s)) w(\varphi_n(s)) dg(s) - \int_0^1 s^{-b} \log(s) w(s) dg(s) \right| \to 0. \tag{5}$$

Combine (4) and (5) to complete the proof of (ii). ■

The second lemma collects an inequality we shall use in the proof of Lemma 3 below.

Lemma 2. *Pick $\alpha > 0$ and x, $y > 0$ such that $x \leq y$. Then there is a positive constant C_α such that*

$$0 \leq y^\alpha - x^\alpha \leq C_\alpha y^{\alpha-1}(y-x).$$

Proof of Lemma 2. When $\alpha \geq 1$, the result is a consequence of the mean value theorem:

$$y^\alpha - x^\alpha \leq \sup_{t \in [x,y]} \{\alpha t^{\alpha-1}\}(y-x) = \alpha y^{\alpha-1}(y-x).$$

In the case $\alpha < 1$, then for any positive integer m we have by multiplying by conjugates:

$$y^\alpha - x^\alpha = \left(y^{2^m\alpha} - x^{2^m\alpha}\right) \Big/ \prod_{p=0}^{m-1} \left[x^{2^p\alpha} + y^{2^p\alpha}\right] \leq \left(y^{2^m\alpha} - x^{2^m\alpha}\right) \Big/ \prod_{p=0}^{m-1} y^{2^p\alpha}$$

so that

$$y^\alpha - x^\alpha \leq y^{-(2^m-1)\alpha} \left(y^{2^m\alpha} - x^{2^m\alpha}\right).$$

Let then m be so large that $2^m\alpha \geq 1$ and use once again the mean value theorem to get

$$y^\alpha - x^\alpha \leq 2^m\alpha y^{-(2^m-1)\alpha} y^{2^m\alpha-1}(y-x) = 2^m\alpha y^{\alpha-1}(y-x)$$

which is the result. ■

In all what follows, we let $\beta_n \to 1$ and $t_n \to 1$ be such that $n(1-\beta_n) \to \infty$ and $(1-t_n)/(1-\beta_n) \to 0$. Let (ψ_n) be a sequence of functions such that for all n, $\psi_n \in \mathcal{F}(\beta_n, t_n)$ and (φ_n) be the sequence of functions defined as

$$\varphi_n(s) = \frac{1-\psi_n(s)}{1-\beta_n}.$$

It is straightforward to show, using

$$\forall s \in [0,1],\ \psi_n(s) \geq \psi_n(1) = \beta_n \ \text{ and } \ 0 \leq 1-(1-\beta_n)s - \psi_n(s) \leq 1-t_n,$$

that

$$\forall s \in [0,1],\ s \leq \varphi_n(s) \leq \min\left(1, s + \frac{1-t_n}{1-\beta_n}\right) \ \text{ and thus } \ \varphi_n(s) \to s \ \text{ as } \ n \to \infty. \tag{6}$$

The third lemma essentially shows that the full extreme Wang DRM $R_{g,\beta_n}(X^a)$, obtained for $t_n = 1$, is equivalent to its modified version

$$R_{g,\beta_n}(X^a; \psi_n) := \int_0^1 [q \circ \psi_n(s)]^a dg(s)$$

and gives a bound for the remainder.

Lemma 3. *Let g be a distortion function on $[0,1]$, $a > 0$ and (ψ_n) be a sequence of functions such that $\psi_n \in \mathcal{F}(\beta_n, t_n)$ for all n.*

(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:

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

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

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

for some $\eta > 0$, we have that for any $\varepsilon \in (0, \min(1/2, \eta))$:

$$\begin{aligned}\frac{R_{g,\beta_n}(X^a;\psi_n)}{U_a([1-\beta_n]^{-1})} &= \int_0^1 s^{-a\gamma} dg(s) + \frac{a\lambda}{\sqrt{n(1-\beta_n)}} \int_0^1 \frac{s^{-\rho}-1}{\rho} s^{-a\gamma} dg(s) \\ &+ \mathrm{O}\left(\left[\frac{1-t_n}{1-\beta_n}\right]^{1/2+\varepsilon}\right) + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).\end{aligned}$$

Proof of Lemma 3. Let $k = n(1-\beta_n)$ so that $k \to \infty$, $k/n \to 0$ and φ_n can be rewritten as

$$\varphi_n(s) = \frac{n}{k}(1-\psi_n(s)).$$

The first statement is proven by using (6) and applying Lemma 1 (i):

$$\frac{R_{g,\beta_n}(X^a;\psi_n)}{U_a([1-\beta_n]^{-1})} = \int_0^1 \frac{U_a(n/k\varphi_n(s))}{U_a(n/k)} dg(s) = \int_0^1 s^{-a\gamma} dg(s)(1+\mathrm{o}(1)). \tag{7}$$

To show the second statement, use (1) and (7) together to get:

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

Lemma 1 with $f = w = 1$ entails

$$\frac{R_{g,\beta_n}(X^a;\psi_n)}{U_a([1-\beta_n]^{-1})} = \int_0^1 [\varphi_n(s)]^{-a\gamma} dg(s) + \frac{a\lambda}{\sqrt{n(1-\beta_n)}} \int_0^1 \frac{s^{-\rho}-1}{\rho} s^{-a\gamma} dg(s) + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right). \tag{8}$$

We focus on the first integral on the right-hand side of this equality. Write

$$\left|\int_0^1 [\varphi_n(s)]^{-a\gamma} dg(s) - \int_0^1 s^{-a\gamma} dg(s)\right| \le \int_0^1 \left([\varphi_n(s)]^{a\gamma} - s^{a\gamma}\right) [\varphi_n(s)]^{-a\gamma} s^{-a\gamma} dg(s).$$

By Lemma 2, there is a positive constant C such that

$$\forall s \in (0,1),\ [\varphi_n(s)]^{a\gamma} - s^{a\gamma} \le C[\varphi_n(s)]^{a\gamma-1}(\varphi_n(s) - s).$$

Consequently,

$$\left|\int_0^1 [\varphi_n(s)]^{-a\gamma} dg(s) - \int_0^1 s^{-a\gamma} dg(s)\right| \leq C \int_0^1 \frac{\varphi_n(s) - s}{\varphi_n(s)} s^{-a\gamma} dg(s).$$

Finally, pick $\varepsilon \in (0, \min(1/2, \eta))$ and notice that since

$$0 \leq \frac{\varphi_n(s) - s}{\varphi_n(s)} \leq 1 \quad \text{and} \quad 0 \leq \varphi_n(s) - s \leq \frac{1 - t_n}{1 - \beta_n},$$

we have:

$$\forall s \in [0,1], \ \frac{\varphi_n(s) - s}{\varphi_n(s)} \leq \left[\frac{\varphi_n(s) - s}{\varphi_n(s)}\right]^{1/2+\varepsilon} \leq \left(\frac{1 - t_n}{1 - \beta_n}\right)^{1/2+\varepsilon} [\varphi_n(s)]^{-1/2-\varepsilon}.$$

Therefore

$$\begin{aligned}\left|\int_0^1 [\varphi_n(s)]^{-a\gamma} dg(s) - \int_0^1 s^{-a\gamma} dg(s)\right| \ &\leq \ C \left(\frac{1 - t_n}{1 - \beta_n}\right)^{1/2+\varepsilon} \int_0^1 s^{-a\gamma} [\varphi_n(s)]^{-1/2-\varepsilon} dg(s) \\ &= \ \mathrm{O}\left(\left[\frac{1 - t_n}{1 - \beta_n}\right]^{1/2+\varepsilon}\right) \qquad (9)\end{aligned}$$

by the dominated convergence theorem. Combining (8) and (9) completes the proof. ■

The fourth lemma is the key element for the proof of our main result. It shows first that the study of the consistency of any empirical counterpart

$$\widehat{R}_{g,\beta_n}(X^a; \psi_n) = \int_0^1 [\widehat{q}_n \circ \psi_n(s)]^a dg(s) = \int_0^1 X^a_{n-\lfloor n(1-\psi_n(s))\rfloor, n} dg(s)$$

of $R_{g,\beta_n}(X^a; \psi_n)$ reduces to that of a proper trimmed/winsorised estimator, and it then examines the asymptotic behaviour of some weighted integrals of the empirical modified tail quantile process $s \mapsto X_{n-\lfloor n(1-\psi_n(s))\rfloor, n}$. For this result, define $k_1 = nt_n$ and $k = n(1 - \beta_n)$.

Lemma 4. *Assume that condition $\mathcal{C}_2(\gamma, \rho, A)$ is satisfied. Let (ψ_n) be a sequence of functions such that $\psi_n \in \mathcal{F}(\beta_n, t_n)$ for all n and*

$$\varphi_n(s) = \frac{1 - \psi_n(s)}{1 - \beta_n} = \frac{n}{k}(1 - \psi_n(s)).$$

(i) Let $a > 0$ and g be a distortion function. Assume that for some $\eta > 0$:

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

Then there is a sequence of functions (Ψ_n) with $\Psi_n \in \mathcal{F}(\min(t_n, 1 - 1/n), \beta_n)$ for all n and:

$$\frac{\widehat{R}_{g,\beta_n}(X^a; \psi_n)}{U_a(n/k)} = \frac{\widehat{R}_{g,\beta_n}(X^a; \Psi_n)}{U_a(n/k)} + \mathrm{o}_{\mathbb{P}}(1).$$

(ii) Let further f be a Borel measurable regularly varying function with index $b \leq a\gamma$. Pick $\delta \in (0, a\gamma - b + \eta)$, and set

$$I_n(f, g, a, \delta) := \int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)} \left(\frac{X^a_{n-\lfloor k\varphi_n(s)\rfloor, n}}{U_a(n/k)} - [\varphi_n(s)]^{-a\gamma}\right) [\varphi_n(s)]^{a\gamma-\delta} dg(s).$$

If $n - k_1 \geq 1$ and $\sqrt{k}A(n/k) = \mathrm{O}(1)$ then $I_n(f, g, a, \delta) \xrightarrow{\mathbb{P}} 0$.

(iii) Let now $a_1, \ldots, a_d > 0$, $f_1, \ldots, f_d$ 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 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 $\delta_j \in (0, a_j\gamma - b_j + \eta)$. If $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$ then

$$(\sqrt{k} I_n(f_j, g_j, a_j, \delta_j))_{1 \le j \le d} \xrightarrow{d} \mathcal{N}(\lambda C, \Sigma)$$

with C being the column vector having $j-$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 having $(i,j)-$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 4. We start by the proof of (i). If $t_n \le 1 - 1/n$, one can take $\Psi_n = \psi_n$ and there is nothing to show. If $t_n > 1 - 1/n$, define $\Psi_n(s) = \min(\psi_n(s), 1 - 1/n)$ and set

$$\Delta_n = \frac{\widehat{R}_{g,\beta_n}(X^a; \psi_n) - \widehat{R}_{g,\beta_n}(X^a; \Psi_n)}{U_a(n/k)} = \frac{1}{U_a(n/k)} \int_0^1 (X^a_{n - \lfloor n(1-\psi_n(s)) \rfloor, n} - X^a_{n - \lfloor n(1-\Psi_n(s)) \rfloor, n}) dg(s).$$

For any n, $\Psi_n(s)$ is clearly Borel measurable, nonincreasing and takes its values in $[0,1]$. Moreover

$$\begin{aligned} & \forall s \in [0,1], \ \psi_n(s) \le 1 - (1-\beta_n)s \\ \Rightarrow \ & \forall s \in [0,1], \ \Psi_n(s) = \min(\psi_n(s), 1 - 1/n) \le 1 - (1-\beta_n)s. \end{aligned} \tag{10}$$

Define now $t'_n = 1 - 1/n$: then $t'_n \le t_n$ which entails

$$\forall s \in [0,1], \ 1 - (1-\beta_n)s - (1 - t'_n) \le 1 - (1-\beta_n)s - (1-t_n) \le \psi_n(s).$$

Moreover:

$$\forall s \in [0,1], \ 1 - (1-\beta_n)s - (1 - t'_n) \le t'_n = 1 - 1/n.$$

These last two chains of inequalities show that

$$\forall s \in [0,1], \ 1 - (1-\beta_n)s - (1-t'_n) \le \Psi_n(s). \tag{11}$$

Combining (10) and (11), we get $\Psi_n \in \mathcal{F}(t'_n, \beta_n) = \mathcal{F}(\min(t_n, 1-1/n), \beta_n)$. Let now

$$\forall s \in [0,1], \ \theta_n(s) = \frac{n}{k}(1 - \Psi_n(s)) = \max(\varphi_n(s), 1/k)$$

so that

$$\Delta_n = \frac{1}{U_a(n/k)} \int_0^1 (X^a_{n - \lfloor k\varphi_n(s) \rfloor, n} - X^a_{n - \lfloor k\theta_n(s) \rfloor, n}) dg(s).$$

Because $ks \le k\varphi_n(s)$ we clearly have

$$\begin{aligned}\Delta_n &\le \frac{1}{U_a(n/k)}\int_0^1 (X^a_{n-\lfloor ks\rfloor,n} - X^a_{n-\lfloor k\theta_n(s)\rfloor,n})dg(s)\\ &= \frac{1}{U_a(n/k)}\int_0^1 (X^a_{n-\lfloor k\max(s,1/2k)\rfloor,n} - X^a_{n-\lfloor k\theta_n(s)\rfloor,n})dg(s).\end{aligned}$$

Set finally $\mu_n(s) = \max(s, 1/2k)$; it is then enough to show that

$$\frac{1}{U_a(n/k)}\int_0^1 (X^a_{n-\lfloor k\mu_n(s)\rfloor,n} - X^a_{n-\lfloor k\theta_n(s)\rfloor,n})dg(s) \xrightarrow{\mathbb{P}} 0. \tag{12}$$

The pivotal idea is to apply Theorem 2.4.8 in de Haan and Ferreira (2006): we may find a Borel measurable function B_a which has constant sign and is asymptotically equivalent to aA at infinity such that

$$s^{a\gamma+1/2+\eta}\left|\sqrt{k}\left(\frac{X^a_{n-\lfloor ks\rfloor,n}}{U_a(n/k)} - s^{-a\gamma}\right) - a\gamma s^{-a\gamma-1}W_n(s) - \sqrt{k}B_a(n/k)s^{-a\gamma}\frac{s^{-\rho}-1}{\rho}\right| \xrightarrow{\mathbb{P}} 0 \tag{13}$$

uniformly in $s \in (0,1]$, where W_n is an appropriate sequence of standard Brownian motions, or equivalently

$$\frac{X^a_{n-\lfloor ks\rfloor,n}}{U_a(n/k)} = s^{-a\gamma}\left(1 + \frac{1}{\sqrt{k}}a\gamma s^{-1}W_n(s) + B_a(n/k)\frac{s^{-\rho}-1}{\rho} + \frac{1}{\sqrt{k}}s^{-1/2-\eta}\,\mathrm{o}_{\mathbb{P}}(1)\right)$$

with the $\mathrm{o}_{\mathbb{P}}(1)$ being uniform in $s \in (0,1]$. Replacing $n-\lfloor ks\rfloor$ by first $n-\lfloor k\mu_n(s)\rfloor$ and then $n-\lfloor k\theta_n(s)\rfloor$, this yields

$$\begin{aligned}\frac{X^a_{n-\lfloor k\mu_n(s)\rfloor,n}}{U_a(n/k)} &= [\mu_n(s)]^{-a\gamma}\left(1 + \frac{1}{\sqrt{k}}a\gamma[\mu_n(s)]^{-1}W_n(\mu_n(s)) + B_a(n/k)\frac{[\mu_n(s)]^{-\rho}-1}{\rho}\right)\\ &+ \frac{1}{\sqrt{k}}[\mu_n(s)]^{-a\gamma-1/2-\eta}\,\mathrm{o}_{\mathbb{P}}(1)\end{aligned} \tag{14}$$

and

$$\begin{aligned}\frac{X^a_{n-\lfloor k\theta_n(s)\rfloor,n}}{U_a(n/k)} &= [\theta_n(s)]^{-a\gamma}\left(1 + \frac{1}{\sqrt{k}}a\gamma[\theta_n(s)]^{-1}W_n(\theta_n(s)) + B_a(n/k)\frac{[\theta_n(s)]^{-\rho}-1}{\rho}\right)\\ &+ \frac{1}{\sqrt{k}}[\theta_n(s)]^{-a\gamma-1/2-\eta}\,\mathrm{o}_{\mathbb{P}}(1).\end{aligned} \tag{15}$$

We first work on the remainder terms. We have

$$\begin{aligned}\frac{1}{\sqrt{k}}[\mu_n(s)]^{-a\gamma-1/2-\eta} = \frac{[\max(1/2k,s)]^{-1/2}}{\sqrt{k}}[\max(1/2k,s)]^{-a\gamma-\eta} &\le \frac{[1/2k]^{-1/2}}{\sqrt{k}}s^{-a\gamma-\eta}\\ &= s^{-a\gamma-\eta}\times \mathrm{O}(1)\end{aligned}$$

where the $\mathrm{O}(1)$ is uniform in $s \in [0,1]$. Thus

$$\frac{1}{\sqrt{k}}\int_0^1 [\mu_n(s)]^{-a\gamma-1/2-\eta}dg(s) = \mathrm{O}(1). \tag{16}$$

A similar result holds with $\mu_n(s)$ replaced by $\theta_n(s)$ since

$$\forall s \in [0,1],\ \theta_n(s) = \max(\varphi_n(s), 1/k) \ge \max(s, 1/k) \ge \max(s, 1/2k) = \mu_n(s).$$

Now, for any $s \in (0,1]$ and $\alpha > 0$,

$$|[\mu_n(s)]^{-\alpha} - [\theta_n(s)]^{-\alpha}| = ([\theta_n(s)]^{\alpha} - [\mu_n(s)]^{\alpha})[\mu_n(s)]^{-\alpha}[\theta_n(s)]^{-\alpha}.$$

A consequence of this inequality is, by Lemma 2,

$$\begin{aligned} |[\mu_n(s)]^{-\alpha} - [\theta_n(s)]^{-\alpha}| &\leq C_\alpha [\theta_n(s)]^{-1}(\theta_n(s) - \mu_n(s))[\mu_n(s)]^{-\alpha} \\ &= C_\alpha \left(1 - \frac{\mu_n(s)}{\theta_n(s)}\right)[\mu_n(s)]^{-\alpha} \end{aligned} \tag{17}$$

where C_α is a positive constant. Since $\varphi_n(s) \to s$ pointwise on $[0,1]$ and

$$1 - \frac{\mu_n(s)}{\theta_n(s)} = \frac{\max(\varphi_n(s), 1/k) - \max(s, 1/2k)}{\max(\varphi_n(s), 1/k)} \leq 1, \tag{18}$$

the function on the left-hand side of (17) converges pointwise to 0 on $(0,1]$ and is bounded by a multiple of $s^{-\alpha}$ on this interval. Thus

$$\int_0^1 ([\mu_n(s)]^{-a\gamma-\eta} - [\theta_n(s)]^{-a\gamma-\eta})dg(s) \to 0 \tag{19}$$

by the dominated convergence theorem. We shall now show that

$$\frac{1}{\sqrt{k}}\int_0^1 \left\{[\mu_n(s)]^{-a\gamma-1}W_n(\mu_n(s))dg(s) - [\theta_n(s)]^{-a\gamma-1}W_n(\theta_n(s))\right\}dg(s) \xrightarrow{\mathbb{P}} 0. \tag{20}$$

Because the W_n are all standard Brownian motions, we may replace W_n by a standard Brownian motion W to show this weak convergence. For any $s \in [0,1]$:

$$\begin{aligned} & \left|[\mu_n(s)]^{-a\gamma-1}W(\mu_n(s)) - [\theta_n(s)]^{-a\gamma-1}W(\theta_n(s))\right| \\ \leq\ & \left|[\mu_n(s)]^{-a\gamma-1} - [\theta_n(s)]^{-a\gamma-1}\right| |W(\mu_n(s))| + [\theta_n(s)]^{-a\gamma-1}\left|W(\mu_n(s)) - W(\theta_n(s))\right|. \end{aligned}$$

The first term on the right-hand side is controlled using (17), the inequality $\mu_n(s) \geq 1/2k$ and the fact that, by the law of the iterated logarithm, $s^{-1/2+\eta}W(s)$ is uniformly stochastically bounded on $(0,1]$:

$$\begin{aligned} & \left|[\mu_n(s)]^{-a\gamma-1} - [\theta_n(s)]^{-a\gamma-1}\right| |W(\mu_n(s))| \\ \leq\ & [\mu_n(s)]^{-1/2}\left(1 - \frac{\mu_n(s)}{\theta_n(s)}\right)\left\{[\mu_n(s)]^{-1/2+\eta}|W(\mu_n(s))|\right\}[\mu_n(s)]^{-a\gamma-\eta}\,\mathrm{O}(1) \\ \leq\ & \sqrt{k}\left(1 - \frac{\mu_n(s)}{\theta_n(s)}\right)[\mu_n(s)]^{-a\gamma-\eta}\,\mathrm{O}_{\mathbb{P}}(1) \end{aligned} \tag{21}$$

where the $\mathrm{O}_{\mathbb{P}}(1)$ term is uniform in $s \in (0,1]$. To control the second term, combine the $(1-\eta)/2-$local Hölder continuity of the standard Brownian motion, which translates to Hölder continuity on the compact interval $[0,1]$, with the inequality $\theta_n(s) \geq \mu_n(s) \geq 1/2k$ and inequality (18):

$$\begin{aligned} [\theta_n(s)]^{-a\gamma-1}\left|W(\mu_n(s)) - W(\theta_n(s))\right| &\leq [\theta_n(s)]^{-a\gamma-1}|\mu_n(s) - \theta_n(s)|^{(1-\eta)/2}\,\mathrm{O}_{\mathbb{P}}(1) \\ &\leq k^{(1-\eta)/2}[\theta_n(s)]^{-a\gamma-\eta}\left(1 - \frac{\mu_n(s)}{\theta_n(s)}\right)^{(1-\eta)/2}\mathrm{O}_{\mathbb{P}}(1) \\ &\leq k^{(1-\eta)/2}[\theta_n(s)]^{-a\gamma-\eta}\,\mathrm{O}_{\mathbb{P}}(1) \end{aligned} \tag{22}$$

where again the $\mathrm{O}_{\mathbb{P}}(1)$ term is uniform in $s \in (0,1]$. Combining (21) and (22) entails

$$\begin{aligned}
& \left|[\mu_n(s)]^{-a\gamma-1}W(\mu_n(s)) - [\theta_n(s)]^{-a\gamma-1}W(\theta_n(s))\right| \\
\leq \; & \left[\sqrt{k}\left(1-\frac{\mu_n(s)}{\theta_n(s)}\right)[\mu_n(s)]^{-a\gamma-\eta} + k^{(1-\eta)/2}[\theta_n(s)]^{-a\gamma-\eta}\right]\mathrm{O}_{\mathbb{P}}(1). \qquad (23)
\end{aligned}$$

By (18), the sequence of functions $1-\mu_n/\theta_n$ converges pointwise to 0 on $(0,1)$; use then (18) and (23) together with the dominated convergence theorem to obtain (20). Finally, recall that $B_a(n/k) = \mathrm{O}(1/\sqrt{k})$ so that

$$B_a(n/k)\int_0^1 [\mu_n(s)]^{-a\gamma}\frac{[\mu_n(s)]^{-\rho}-1}{\rho}dg(s) \to 0 \qquad (24)$$

by the dominated convergence theorem again, with an analogue result for μ_n replaced by θ_n. Combining (14), (15), (16), (19), (20) and (24) completes the proof of (12) and thus of the first statement.

We proceed with the proof of point (ii). Let $\varepsilon' \in (0,1/2)$ be such that $\delta + 2\varepsilon' < \eta$. From Lemma 1 in El Methni and Stupfler (2017), U_a satisfies condition $\mathcal{C}_2(a\gamma,\rho,aA)$. By (14) with η replaced by ε' and $\mu_n(s)$ replaced by $\varphi_n(s)$,

$$\begin{aligned}
I_n(f,g,a,\delta) &= \zeta_n + \xi_n + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)}[\varphi_n(s)]^{-1/2-\delta-\varepsilon'}dg(s)\right) \\
\text{with } \zeta_n &= a\gamma\frac{1}{\sqrt{k}}\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)}[\varphi_n(s)]^{-1-\delta}W_n(\varphi_n(s))dg(s) \\
\text{and } \xi_n &= B_a(n/k)\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)}\frac{[\varphi_n(s)]^{-\rho}-1}{\rho}[\varphi_n(s)]^{-\delta}dg(s).
\end{aligned}$$

The remainder term is controlled in the following way: notice that $\varphi_n(s) \geq \varphi_n(0) = (n-k_1)/k \geq 1/k$ and thus, by Lemma 1 (i):

$$\begin{aligned}
\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)}[\varphi_n(s)]^{-1/2-\delta-\varepsilon'}dg(s) &\leq k^{1/2-\varepsilon'}\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)}[\varphi_n(s)]^{-\delta-2\varepsilon'}dg(s) \\
&= \mathrm{O}\left(k^{1/2-\varepsilon'}\right) \qquad (25)
\end{aligned}$$

which leads to

$$I_n(f,g,a,\delta) = \zeta_n + \xi_n + \mathrm{o}_{\mathbb{P}}(1). \qquad (26)$$

Recall now that 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 $s \to 0$. Thus

$$\zeta_n = \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)}[\varphi_n(s)]^{-1/2-\delta-\varepsilon'}dg(s)\right) = \mathrm{o}_{\mathbb{P}}(1) \qquad (27)$$

by (25). Finally, we obtain, using the bound $\sqrt{k}A(n/k) = \mathrm{O}(1)$ and the fact that B_a is asymptotically equivalent to aA:

$$\xi_n = \mathrm{O}\left(\frac{1}{\sqrt{k}}\int_0^1 \frac{f(n/k\varphi_n(s))}{f(n/k)}[\varphi_n(s)]^{-\delta-\varepsilon'}dg(s)\right) = \mathrm{o}(1) \qquad (28)$$

by Lemma 1. Combining (26), (27) and (28) completes the proof of the second statement.

The proof of (iii) is actually that of Lemma 3 in El Methni and Stupfler (2017) up to slight changes essentially due to s having to be replaced by $\varphi_n(s)$ throughout, which can be handled by using Lemma 1. We omit it for the sake of brevity. ■

Our final lemma is the key to the proof of Theorem 2(ii).

Lemma 5. *Let g be a distortion function on $[0,1]$, $a>0$, (ψ_n) be a sequence of functions such that $\psi_n \in \mathcal{F}(\beta_n, t_n)$ for all n and*

$$\varphi_n(s) = \frac{1-\psi_n(s)}{1-\beta_n}.$$

(i) If $\widehat{\gamma}_n$ is a consistent estimator of γ and there is $\eta>0$ such that

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

then we have that:

$$\frac{\int_0^1 [\varphi_n(s)]^{-a\widehat{\gamma}_n} dg(s)}{\int_0^1 s^{-a\gamma} dg(s)} \xrightarrow{\mathbb{P}} 1.$$

(ii) If furthermore $\widehat{\gamma}_n$ is a $\sqrt{n(1-\beta_n)}-$consistent estimator of γ, then we have that:

$$\frac{\int_0^1 [\varphi_n(s)]^{-a\widehat{\gamma}_n} dg(s)}{\int_0^1 [\varphi_n(s)]^{-a\gamma} dg(s)} = 1 - a(\widehat{\gamma}_n - \gamma)\frac{\int_0^1 s^{-a\gamma}\log(1/s) dg(s)}{\int_0^1 s^{-a\gamma} dg(s)} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).$$

Proof of Lemma 5. To prove the first assertion, note that $\varphi_n(s) \to s$ pointwise on $[0,1]$ and $[\varphi_n(s)]^{-a\gamma-\eta} \le s^{-a\gamma-\eta}$, so that by the dominated convergence theorem,

$$\int_0^1 [\varphi_n(s)]^{-a\widehat{\gamma}_n} dg(s) = \int_0^1 [\varphi_n(s)]^{-a(\widehat{\gamma}_n-\gamma)} \left[\frac{\varphi_n(s)}{s}\right]^{-a\gamma} s^{-a\gamma} dg(s) \xrightarrow{\mathbb{P}} \int_0^1 s^{-a\gamma} dg(s).$$

To show the second result, set $\kappa(x) = e^x - 1 - x$ and notice that

$$\begin{aligned}
\frac{\int_0^1 [\varphi_n(s)]^{-a\widehat{\gamma}_n} dg(s)}{\int_0^1 [\varphi_n(s)]^{-a\gamma} dg(s)} &= 1 - a(\widehat{\gamma}_n-\gamma)\frac{\int_0^1 [\varphi_n(s)]^{-a\gamma}\log(\varphi_n(s)) dg(s)}{\int_0^1 [\varphi_n(s)]^{-a\gamma} dg(s)} \\
&+ \frac{\int_0^1 [\varphi_n(s)]^{-a\gamma}\kappa(-a(\widehat{\gamma}_n-\gamma)\log(\varphi_n(s))) dg(s)}{\int_0^1 [\varphi_n(s)]^{-a\gamma} dg(s)}.
\end{aligned}$$

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

$$\begin{aligned}
&\left|\int_0^1 [\varphi_n(s)]^{-a\gamma}\kappa(-a(\widehat{\gamma}_n-\gamma)\log(\varphi_n(s))) dg(s)\right| \\
\le\ & \frac{a^2}{2}(\widehat{\gamma}_n-\gamma)^2 \int_0^1 [\varphi_n(s)]^{-a\gamma}\log^2(\varphi_n(s))[\varphi_n(s)]^{-a|\widehat{\gamma}_n-\gamma|} dg(s) \\
\le\ & \frac{a^2}{2}(\widehat{\gamma}_n-\gamma)^2 \int_0^1 s^{-a\gamma}\log^2(1/s) s^{-a|\widehat{\gamma}_n-\gamma|} dg(s)
\end{aligned}$$

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

$$\left|\int_0^1 [\varphi_n(s)]^{-a\gamma}\kappa(-a(\widehat{\gamma}_n-\gamma)\log(\varphi_n(s)))dg(s)\right| = \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)$$

and thus

$$\frac{\int_0^1 [\varphi_n(s)]^{-a\widehat{\gamma}_n}dg(s)}{\int_0^1 [\varphi_n(s)]^{-a\gamma}dg(s)} = 1 - a(\widehat{\gamma}_n-\gamma)\frac{\int_0^1 [\varphi_n(s)]^{-a\gamma}\log(\varphi_n(s))dg(s)}{\int_0^1 [\varphi_n(s)]^{-a\gamma}dg(s)} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).$$

Note finally that $\varphi_n(s) \to s$ pointwise on $[0,1]$ and

$$[\varphi_n(s)]^{-a\gamma} \le s^{-a\gamma-\eta} \quad \text{and} \quad |[\varphi_n(s)]^{-a\gamma}\log(\varphi_n(s))| \le s^{-a\gamma}\log(1/s) \le K s^{-a\gamma-\eta}$$

where K is a positive constant. Both right-hand sides being integrable with respect to the measures dg, we conclude by the dominated convergence theorem that

$$\frac{\int_0^1 [\varphi_n(s)]^{-a\widehat{\gamma}_n}dg(s)}{\int_0^1 [\varphi_n(s)]^{-a\gamma}dg(s)} = 1 - a(\widehat{\gamma}_n-\gamma)\frac{\int_0^1 s^{-a\gamma}\log(1/s)dg(s)}{\int_0^1 s^{-a\gamma}dg(s)} + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).$$

This is the desired result. ■

Appendix B: Proofs of the main results

Proof of Proposition 1. By definition:

$$R_{g,\beta,t}^{\mathrm{Trim}}(h(X)) = \int_0^1 h \circ q(t-(t-\beta)s)dg(s). \tag{29}$$

It follows from the condition on q that the function F is necessarily continuous on an open interval containing $[q(\beta),\infty)$. The cdf of X given $X \in [q(\beta), q(t)]$ is then

$$F_{\beta,t}(x) := \mathbb{P}(X \le x | X \in [q(\beta), q(t)]) = \begin{cases} 0 & \text{if} \quad x < q(\beta), \\ \dfrac{F(x)-\beta}{t-\beta} & \text{if} \quad x \in [q(\beta), q(t)], \\ 1 & \text{if} \quad x > q(t). \end{cases}$$

The related quantile function is defined by

$$\begin{aligned} \forall \alpha \in (0,1),\ q_{\beta,t}(\alpha) := \inf\{x \in \mathbb{R} \,|\, F_{\beta,t}(x) \ge \alpha\} &= \inf\{x \in \mathbb{R} \,|\, F(x) \ge \beta + (t-\beta)\alpha\} \\ &= q(\beta + (t-\beta)\alpha) \end{aligned}$$

and thus

$$R_g(h(X_{\beta,t}^{\mathrm{Trim}})) = \int_0^1 h \circ q_{\beta,t}(1-s)dg(s) = \int_0^1 h \circ q(t-(t-\beta)s)dg(s). \tag{30}$$

Combining (29) and (30) completes the proof. ■

Proof of Proposition 2. We have:

$$R^{\mathrm{Wins}}_{g,\beta,t}(h(X)) = \int_0^1 h \circ q(\min(t, 1-(1-\beta)s)dg(s). \tag{31}$$

The function F is continuous and increasing on an open interval containing $[q(\beta),\infty)$, so that the cdf of $X^{\mathrm{Wins}}_{\beta,t}$ is

$$G_{\beta,t}(x) := \begin{cases} 0 & \text{if} \quad x < q(\beta), \\ \dfrac{F(x)-\beta}{1-\beta} & \text{if} \quad x \in [q(\beta), q(t)), \\ 1 & \text{if} \quad x \geq q(t). \end{cases}$$

The related quantile function is defined by

$$\forall \alpha \in (0,1),\ Q_{\beta,t}(\alpha) := \inf\{x \in \mathbb{R} \,|\, G_{\beta,t}(x) \geq \alpha\}.$$

When $\alpha < (t-\beta)/(1-\beta)$, this is

$$Q_{\beta,t}(\alpha) = \inf\{x \in \mathbb{R} \,|\, F(x) \geq \beta + (1-\beta)\alpha\} = q(\beta + (1-\beta)\alpha),$$

and if $\alpha \geq (t-\beta)/(1-\beta)$ then, since

$$\forall x \in [q(\beta), q(t)),\ G_{\beta,t}(x) = \frac{F(x)-\beta}{1-\beta} < \frac{t-\beta}{1-\beta} \leq \alpha \ \text{ and } \ G_{\beta,t}(q(t)) = 1,$$

we must have $Q_{\beta,t}(\alpha) = q(t)$. It follows that $Q_{\beta,t}(\alpha) = q(\min(t, \beta + (1-\beta)\alpha))$ and therefore

$$R_g(h(X^{\mathrm{Wins}}_{\beta,t})) = \int_0^1 h \circ Q_{\beta,t}(1-s)dg(s) = \int_0^1 h \circ q(\min(t, 1-(1-\beta)s))dg(s). \tag{32}$$

Combining (31) and (32) completes the proof. ■

Proof of Theorem 1. We start by proving (i). We have the equality

$$\widehat{R}_{g,\beta_n}(X^a; \psi_n) = \int_0^1 X^a_{n-\lfloor n(1-\psi_n(s))\rfloor, n}\, dg(s).$$

By Lemmas 3 (i) and 4 (i),

$$\frac{\widehat{R}_{g,\beta_n}(X^a; \psi_n)}{R_{g,\beta_n}(X^a)} = \frac{\widehat{R}_{g,\beta_n}(X^a; \Psi_n)}{R_{g,\beta_n}(X^a; \Psi_n)}(1 + \mathrm{o}_{\mathbb{P}}(1))$$

with $\Psi_n(s) = \min(\psi_n(s), 1-1/n)$, and in particular it is enough to tackle the case when $n(1-t_n) \geq 1$. By Lemma 3 (i) again, it suffices to prove that:

$$\frac{\widehat{R}_{g,\beta_n}(X^a; \psi_n) - R_{g,\beta_n}(X^a; \psi_n)}{U_a([1-\beta_n]^{-1})} \xrightarrow{\mathbb{P}} 0.$$

Define $k = n(1-\beta_n)$, notice that $k/n \to 0$ and write

$$\frac{\widehat{R}_{g,\beta_n}(X^a;\psi_n) - R_{g,\beta_n}(X^a;\psi_n)}{U_a([1-\beta_n]^{-1})} = \zeta_n(a,g) + \xi_n(a,g) \tag{33}$$

$$\text{with } \zeta_n(a,g) = \int_0^1 \frac{U_a(n/k\varphi_n(s))}{U_a(n/k)} \left(\frac{X^a_{n-\lfloor k\varphi_n(s)\rfloor,n}}{U_a(n/k)} - [\varphi_n(s)]^{-a\gamma} \right) [\varphi_n(s)]^{a\gamma} dg(s)$$

$$\text{and } \xi_n(a,g) = \int_0^1 \frac{U_a(n/k\varphi_n(s))}{U_a(n/k)} \frac{X^a_{n-\lfloor k\varphi_n(s)\rfloor,n}}{U_a(n/k)} \left(\frac{U_a(n/k)}{U_a(n/k\varphi_n(s))} - [\varphi_n(s)]^{a\gamma} \right) dg(s)$$

$$\text{where } \varphi_n(s) = \frac{n}{k}(1-\psi_n(s)). \tag{34}$$

By Lemma 4 (ii), $\zeta_n(a,g) = \mathrm{o}_{\mathbb{P}}(1)$. We control $\xi_n(a,g)$ by Proposition B.1.10 in de Haan and Ferreira (2006): since $\varphi_n(s) \in [s,1]$, we have for any $\delta \in (0,\eta)$ and n large enough,

$$\begin{aligned} |\xi_n(a,g)| &\leq \delta \int_0^1 \frac{U_a(n/k\varphi_n(s))}{U_a(n/k)} \left(\frac{X^a_{n-\lfloor k\varphi_n(s)\rfloor,n}}{U_a(n/k)} - [\varphi_n(s)]^{-a\gamma} \right) [\varphi_n(s)]^{a\gamma-\delta} dg(s) \\ &+ \delta \int_0^1 \frac{U_a(n/k\varphi_n(s))}{U_a(n/k)} [\varphi_n(s)]^{-\delta} dg(s). \end{aligned}$$

The first-term on the right-hand side is controlled by Lemma 4 (ii), while the second integral converges to a finite positive constant by Lemma 1 (i). Since δ can be taken arbitrarily small, we get $\xi_n(a,g) = \mathrm{o}_{\mathbb{P}}(1)$ and the proof is complete.

We now turn to the proof of the second statement. By Lemma 3 (ii),

$$\sqrt{n(1-\beta_n)} \left(\frac{R_{g_j,\beta_n}(X^{a_j};\psi_n)}{R_{g_j,\beta_n}(X^{a_j})} - 1 \right) = \mathrm{o}(1) + \mathrm{O}\left(\sqrt{n(1-t_n)} \left[\frac{1-t_n}{1-\beta_n} \right]^{\varepsilon} \right)$$

for any $\varepsilon \in (0, \min(1/2,\eta))$. Using an assumption on the pair of sequences $((\beta_n),(t_n))$, we obtain with a suitable choice of ε that:

$$\sqrt{n(1-\beta_n)} \left(\frac{R_{g_j,\beta_n}(X^{a_j};\psi_n)}{R_{g_j,\beta_n}(X^{a_j})} - 1 \right) = \mathrm{o}(1)$$

and thus it is enough to show the convergence

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

or equivalently, by Lemma 3 (i),

$$\sqrt{n(1-\beta_n)} \left(\frac{\widehat{R}_{g_j,\beta_n}(X^{a_j};\psi_n) - R_{g_j,\beta_n}(X^{a_j};\psi_n)}{U_{a_j}([1-\beta_n]^{-1})} \right)_{1\leq j\leq d} \xrightarrow{d} \mathcal{N}(0,M) \tag{35}$$

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

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

To this end, we use equation (33):

$$\sqrt{n(1-\beta_n)}\frac{\widehat{R}_{g_j,\beta_n}(X^{a_j};\psi_n)-R_{g_j,\beta_n}(X^{a_j};\psi_n)}{U_{a_j}([1-\beta_n]^{-1})}=\sqrt{k}\zeta_n(a_j,g_j)+\sqrt{k}\xi_n(a_j,g_j).$$

By Lemma 4 (iii):

$$\left(\sqrt{k}\zeta_{j,n}(a_j,g_j)\right)_{1\leq j\leq n}\xrightarrow{d}\mathcal{N}\left(\lambda C,M\right) \tag{36}$$

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 (1), there exist 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 $\delta>0$:

$$\forall s\in(0,1],\ \left|\frac{1}{B_{a_j}(n/ks)}\left(\frac{U_{a_j}(n/k)}{U_{a_j}(n/ks)}-s^{a_j\gamma}\right)-s^{a_j\gamma}\frac{s^{\rho}-1}{\rho}\right|\leq\delta s^{a_j\gamma+\rho-\delta} \tag{37}$$

for n sufficiently large. Replacing s by $\varphi_n(s)$ makes us consider the following decomposition of $\xi_{j,n}$:

$$\xi_{j,n}(a_j,g_j)=\xi^{(1)}_{j,n}(a_j,g_j)+\xi^{(2)}_{j,n}(a_j,g_j) \tag{38}$$

with

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

Here the bound on $\xi^{(2)}_{j,n}(a_j,g_j)$ holds for any $\delta\in(0,\eta)$ when n is large enough. Writing

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

we get by Lemma 4 (ii):

$$\xi^{(1)}_{j,n}(a_j,g_j)=\int_0^1\frac{U_{a_j}(n/k\varphi_n(s))}{U_{a_j}(n/k)}B_{a_j}(n/k\varphi_n(s))\frac{[\varphi_n(s)]^{\rho}-1}{\rho}dg_j(s)+\mathrm{o}_{\mathbb{P}}(B_{a_j}(n/k)).$$

Applying Lemma 1 with $f=U_{a_j}|B_{a_j}|$ and using the convergence $\sqrt{k}B_{a_j}(n/k)\to a_j\lambda$, we get

$$\begin{aligned}\sqrt{k}\xi^{(1)}_{j,n}(a_j,g_j)&=\sqrt{k}B_{a_j}(n/k)\int_0^1 s^{-a_j\gamma}\frac{1-s^{-\rho}}{\rho}dg_j(s)+\mathrm{o}_{\mathbb{P}}(1)\\&=-a_j\lambda\int_0^1 s^{-a_j\gamma}\frac{s^{-\rho}-1}{\rho}dg_j(s)+\mathrm{o}_{\mathbb{P}}(1)=-\lambda C_j+\mathrm{o}_{\mathbb{P}}(1)\end{aligned} \tag{39}$$

since B_{a_j} is equivalent to a_jA. Besides, the ideas used to control $\xi^{(1)}_{j,n}(a_j,g_j)$ yield for n large enough:

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

which, since δ can be taken arbitrarily small, entails

$$\left|\sqrt{k}\xi_{j,n}^{(2)}(a_j, g_j)\right| = \mathrm{o}_{\mathbb{P}}(1). \tag{40}$$

Combining (38), (39) and (40) entails

$$\left(\sqrt{k}\xi_{j,n}(a_j, g_j)\right)_{1\le j\le d} \xrightarrow{\mathbb{P}} -\lambda C. \tag{41}$$

Combine finally (33), (36) and (41) to obtain (35): the proof is complete. ∎

Proof of Theorem 2. The first assertion is a direct consequence of Theorem 1(i) and of Lemma 5(i).

To show the second result, write

$$\frac{\widetilde{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}{\widehat{R}_{g_j,\beta_n}(X^{a_j};\psi_n)} = \frac{\int_0^1 s^{-a_j\gamma}dg_j(s)}{\int_0^1 [\varphi_n(s)]^{-a_j\gamma}dg_j(s)} \times \frac{\int_0^1 [\varphi_n(s)]^{-a_j\gamma}dg_j(s)}{\int_0^1 [\varphi_n(s)]^{-a_j\widehat{\gamma}_n}dg_j(s)} \times \frac{\int_0^1 s^{-a_j\widehat{\gamma}_n}dg_j(s)}{\int_0^1 s^{-a_j\gamma}dg_j(s)}.$$

We first work with the first factor on the right-hand side. Writing

$$\begin{aligned}\frac{\int_0^1 s^{-a_j\gamma}dg_j(s)}{\int_0^1 [\varphi_n(s)]^{-a_j\gamma}dg_j(s)} &= \frac{\int_0^1 [(1-\beta_n)s]^{-a_j\gamma}dg_j(s)}{[(1-\beta_n)]^{-a\gamma}} \times \frac{[(1-\beta_n)]^{-a\gamma}}{\int_0^1 [(1-\beta_n)\varphi_n(s)]^{-a_j\gamma}dg_j(s)} \\ &= \frac{R_{g_j,\beta_n}(Y^{a_j})}{[(1-\beta_n)]^{-a_j\gamma}} \times \frac{[(1-\beta_n)]^{-a_j\gamma}}{R_{g_j,\beta_n}(Y^{a_j};\psi_n)}\end{aligned}$$

where Y has a Pareto distribution with tail index γ, and applying Lemma 3(ii) twice, we obtain

$$\frac{\int_0^1 s^{-a_j\gamma}dg_j(s)}{\int_0^1 [\varphi_n(s)]^{-a_j\gamma}dg_j(s)} = 1 + \mathrm{o}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).$$

To control the second and third terms, we apply Lemma 5(ii) to each term successively to get

$$\frac{\int_0^1 [\varphi_n(s)]^{-a_j\gamma}dg_j(s)}{\int_0^1 [\varphi_n(s)]^{-a_j\widehat{\gamma}_n}dg_j(s)} \times \frac{\int_0^1 s^{-a_j\widehat{\gamma}_n}dg_j(s)}{\int_0^1 s^{-a_j\gamma}dg_j(s)} = 1 + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).$$

Combining these two results entails

$$\frac{\widetilde{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}{\widehat{R}_{g_j,\beta_n}(X^{a_j};\psi_n)} = 1 + \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).$$

Apply Theorem 1(ii) to complete the proof. ∎

Proof of Theorem 3. We start by writing:

$$\frac{\widetilde{R}^W_{g_j,\delta_n}(X^{a_j};\psi_n)}{R_{g_j,\delta_n}(X^{a_j})} = \left(\frac{1-\beta_n}{1-\delta_n}\right)^{a_j(\widehat{\gamma}_n-\gamma)} \frac{\widetilde{R}_{g_j,\beta_n}(X^{a_j};\psi_n)}{R_{g_j,\beta_n}(X^{a_j})} \times \frac{R_{g_j,\beta_n}(X^{a_j})}{R_{g_j,\delta_n}(X^{a_j})}\left(\frac{1-\beta_n}{1-\delta_n}\right)^{a_j\gamma} \tag{42}$$

which is the basic step for our proof. Taking logarithms and applying Lemma 4 of El Methni and Stupfler (2017) with $Y = X^{a_j}$, we get

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

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

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

The hypothesis on $\widehat{\gamma}_n$ and a Taylor expansion of the exponential function now make it clear that

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

which completes the proof. ■

Appendix C: Additional simulation results

In this Appendix we examine some simulation results in the case of a sample size $n = 100$. We work using the methodology described in Section 4 of the main paper: in particular, we set $\delta = 0.99 = 1 - n^{-1}$. The only modification is that the window parameter in the choice of the trimming/winsorising order t is now $h_2 = 0.03$. This is needed because we should have $h_2 > 1/n = 0.01$ in this case, see the description of the choice procedure for t on pages 21–22 of the main article.

Appendix C.1: Moderately heavy tails

Results for the case of moderately heavy tails are given in Figures 1 and 2. In the case of extreme TVaR estimation, the conclusions are similar to those reached in the case $n = 1000$: for $|\rho| \geq 1$, the bias and MSE of our CTrim-PL and CWins-PL estimators are comparable to those of the full PL estimator, indicating that the correction factor works as expected. When ρ is closer to 0, the performance of the CTrim-PL and CWins-PL estimators deteriorates relatively to the PL estimator. Note though that in the case $\rho = -2/3$ and $n = 100$, both bias and MSE of the CTrim-PL and CWins-PL estimators are higher than those of the PL estimator, while when n increases to 1000, the MSEs of the CTrim-PL and CWins-PL estimators become lower than the MSE of the PL estimator and the gap in terms of bias becomes much narrower (see Section 4.1 of the main article). In other words, the suggested estimators first catch up with and then perform better than the PL estimator as the sample size grows. This can also be observed when the estimation of the extreme PH risk measure is considered: for a sample size $n = 100$, the proposed estimators are still much better than the PL estimator in terms of bias when $|\rho| \geq 1$, while being poorer in terms of MSE. When the sample size increases, the MSEs of our estimators actually become appreciably lower than the MSE of the PL estimator, and the relative gap in performance between the CTrim-PL and CWins-PL estimators and the PL estimator reduces sharply in a case with a lower $|\rho|$.

These conclusions for the low sample size $n = 100$ could have been expected from the way the proposed estimators are constructed. When the sample size is low, the correction factors used in the CTrim-PL and CWins-PL estimators have to reconstruct a bigger proportion of the information missing about the right tail of the underlying distribution – impacting performance in terms of bias – while at the same time being built on a sample a lot smaller and therefore being much more variable than in the case $n = 1000$ – thus increasing the MSE. To put it differently, the deletion of even a handful of top order statistics in the case $n = 100$ is much more problematic in terms of loss of information, and therefore in terms of statistical efficiency, than the deletion of this same number of order statistics in the case $n = 1000$. This effect is of course amplified as $|\rho|$ decreases, since for small $|\rho|$ only the very top order statistics, which are precisely those targeted by the trimming/winsorising scheme, can bring meaningful information about the unknown right tail of the variable of interest. This is why, for smaller samples and a moderate value of γ, we would advise to assess the suitability of our estimators on a case-by-case basis, all the more so when a challenging Wang DRM such as the PH risk measure is considered. In particular, in a situation where (i) the tail appears to be moderately heavy, (ii) the sample size is low and (iii) there is no specific indication of unreasonably high values in the sample, then one might simply use the basic PL estimator in place of the CTrim-PL or CWins-PL estimator.

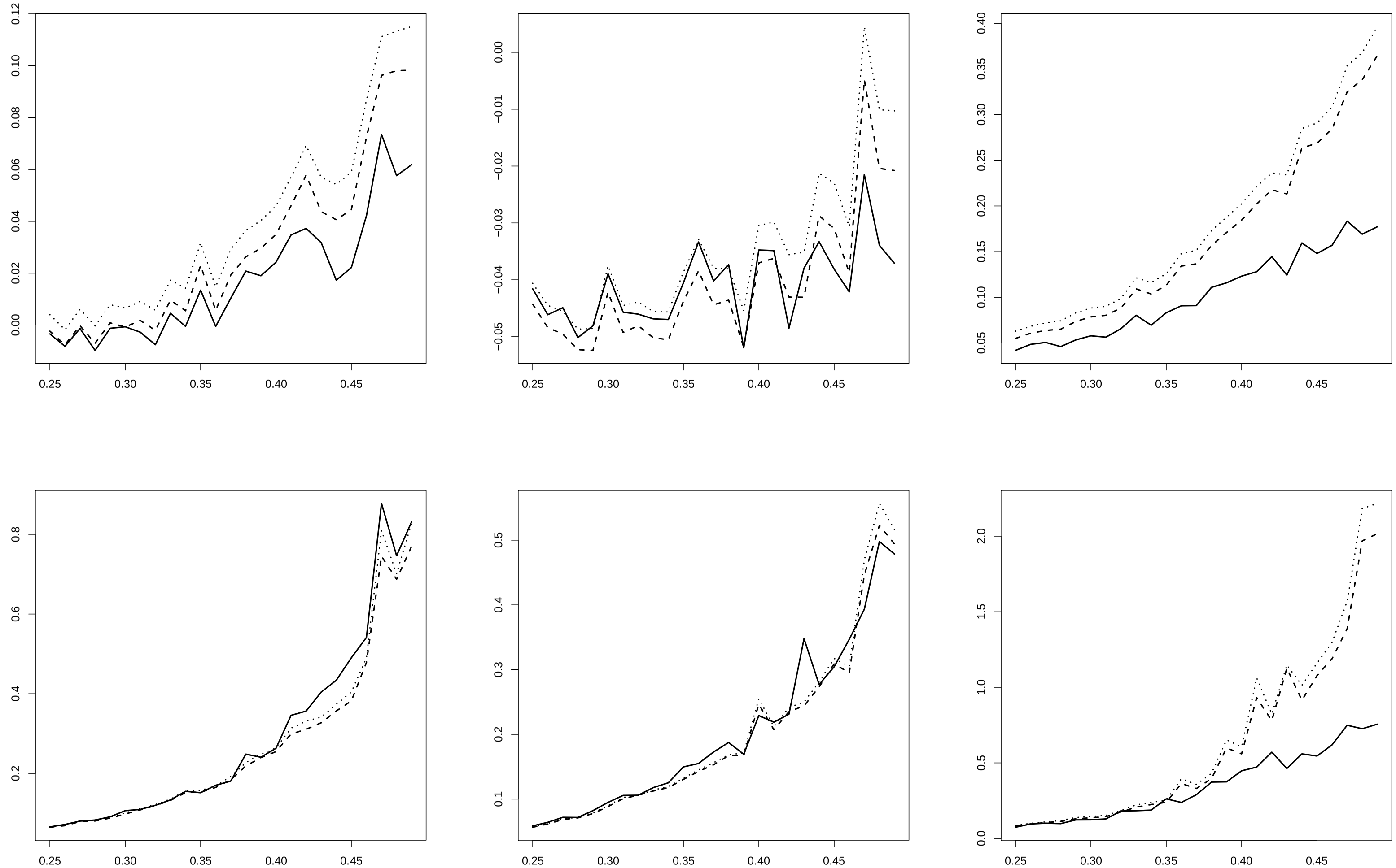

Figure 1: Extreme TVaR estimation with $\delta = 0.99$, for $\gamma \in [0.25, 0.49]$ and on $N = 5000$ replications of a sample of size $n = 100$. Top panels: relative bias, bottom panels: relative MSE. Left: case of the Fréchet distribution, middle: case of the Burr distribution with $\rho = -2$, right: case of the Burr distribution with $\rho = -2/3$. Full line: PL estimator, dotted line: CTrim-PL estimator, dashed line: CWins-PL estimator.

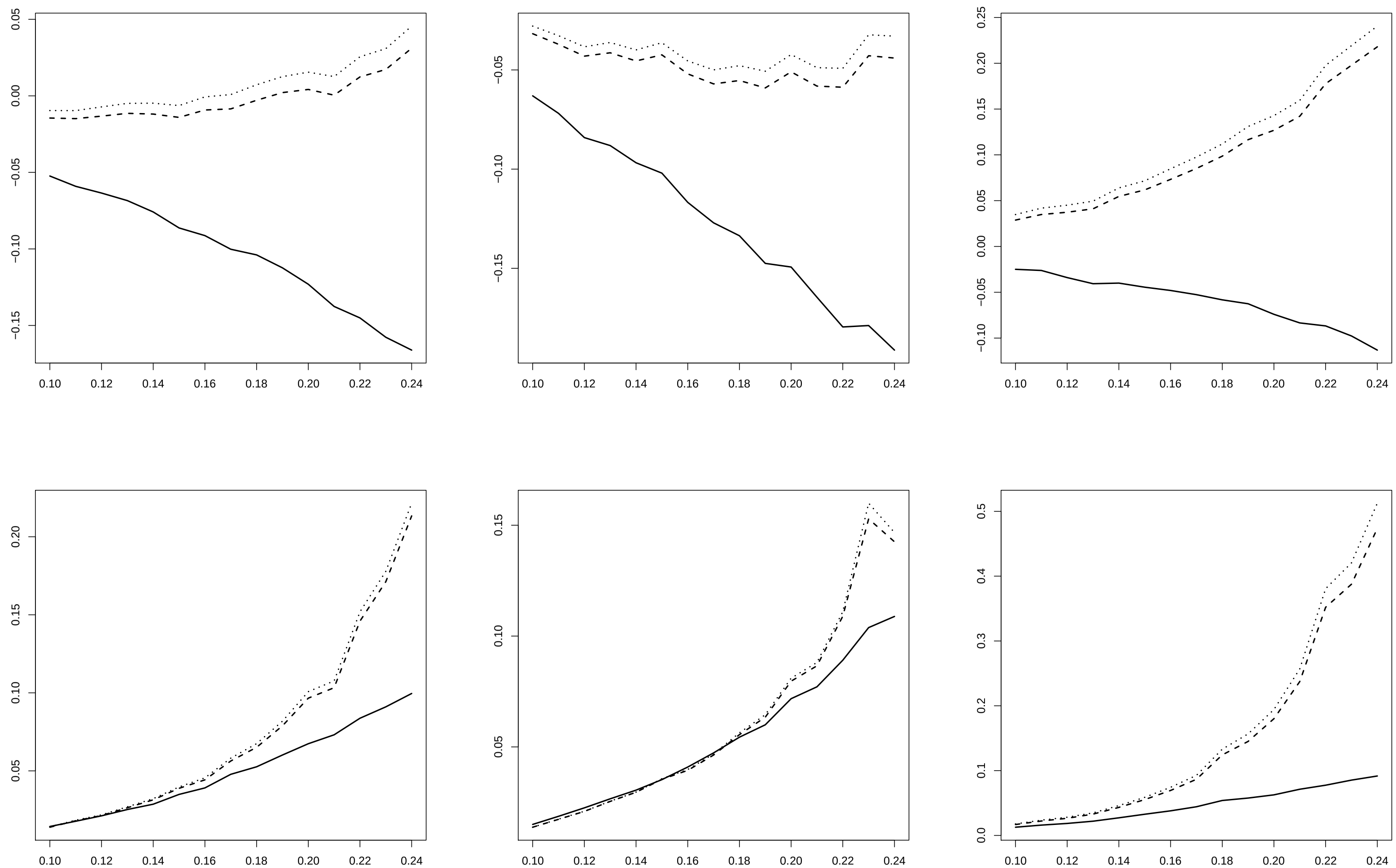

Figure 2: Extreme PH(1/2) estimation with $\delta = 0.99$, for $\gamma \in [0.1, 0.24]$ and on $N = 5000$ replications of a sample of size $n = 100$. Top panels: relative bias, bottom panels: relative MSE. Left: case of the Fréchet distribution, middle: case of the Burr distribution with $\rho = -2$, right: case of the Burr distribution with $\rho = -2/3$. Full line: PL estimator, dotted line: CTrim-PL estimator, dashed line: CWins-PL estimator.

Appendix C.2: Very heavy tails

Results in the challenging case of very heavy tails are given in Figures 3 and 4. The finite-sample performance of the CTrim-PL and CWins-PL estimators when estimating the TVaR in typical cases is somewhat disappointing relatively to the performance of the full PL estimator; in that case, just as in the moderately heavy tails context, the CTrim-PL and CWins-PL estimators catch up with (and then perform better than, for $|\rho| \geq 1$) the PL estimator as the sample size grows, see Section 4.2 in the main paper. By contrast, the conclusion reached in the case of moderately heavy tails for the estimation of the PH risk measure stays true: the CTrim-PL and CWins-PL estimators surprisingly have a lower bias than the PL estimator when $|\rho| \geq 1$. An important improvement can appear on atypical cases: in the example of TVaR estimation, the CTrim-PL and CWins-PL estimators more than halve the bias overall in the case of the Fréchet distribution and can reduce it by more than 75% in the case of the Burr distribution with $\rho = -2$; the improvement is less clear for lower values of $|\rho|$ and for this small sample size, although the CTrim-PL and CWins-PL seem to outperform the PL estimator for large values of γ. In the case of the estimation of the extreme PH risk measure, the CTrim-PL and CWins-PL reduce the bias by roughly a third overall in the Fréchet case and by a half overall in the Burr case with $\rho = -2$. Again, the performance of the CTrim-PL and CWins-PL estimators deteriorates with respect to that of the PL estimator when $|\rho|$ decreases, for the reasons we explained at the end of Appendix C.1. We note in particular that in all the conditioned cases considered, the performance of the CTrim-PL and CWins-PL improves sharply, relatively to the PL estimator, when a larger sample size is considered, see the lower panels of Figures 3 and 4 in the main article.

As a conclusion from this simulation study, it appears that the proposed CTrim-PL and CWins-PL can still bring a significant improvement overall relatively to the PL estimator in the most difficult cases when γ is large and the top order statistics have very high values, even for the low sample size $n = 100$. The results therefore suggest that our estimators have practical value in atypical cases, even for small sample sizes. Let us finally reiterate that, for such a low sample size and as in the previous case of moderately heavy tails, our simulation study indicates that the suitability of our estimators should be decided case by case: in particular, if there is no evidence of extremely high values in the sample, then one might simply decide to use the basic PL estimator instead.

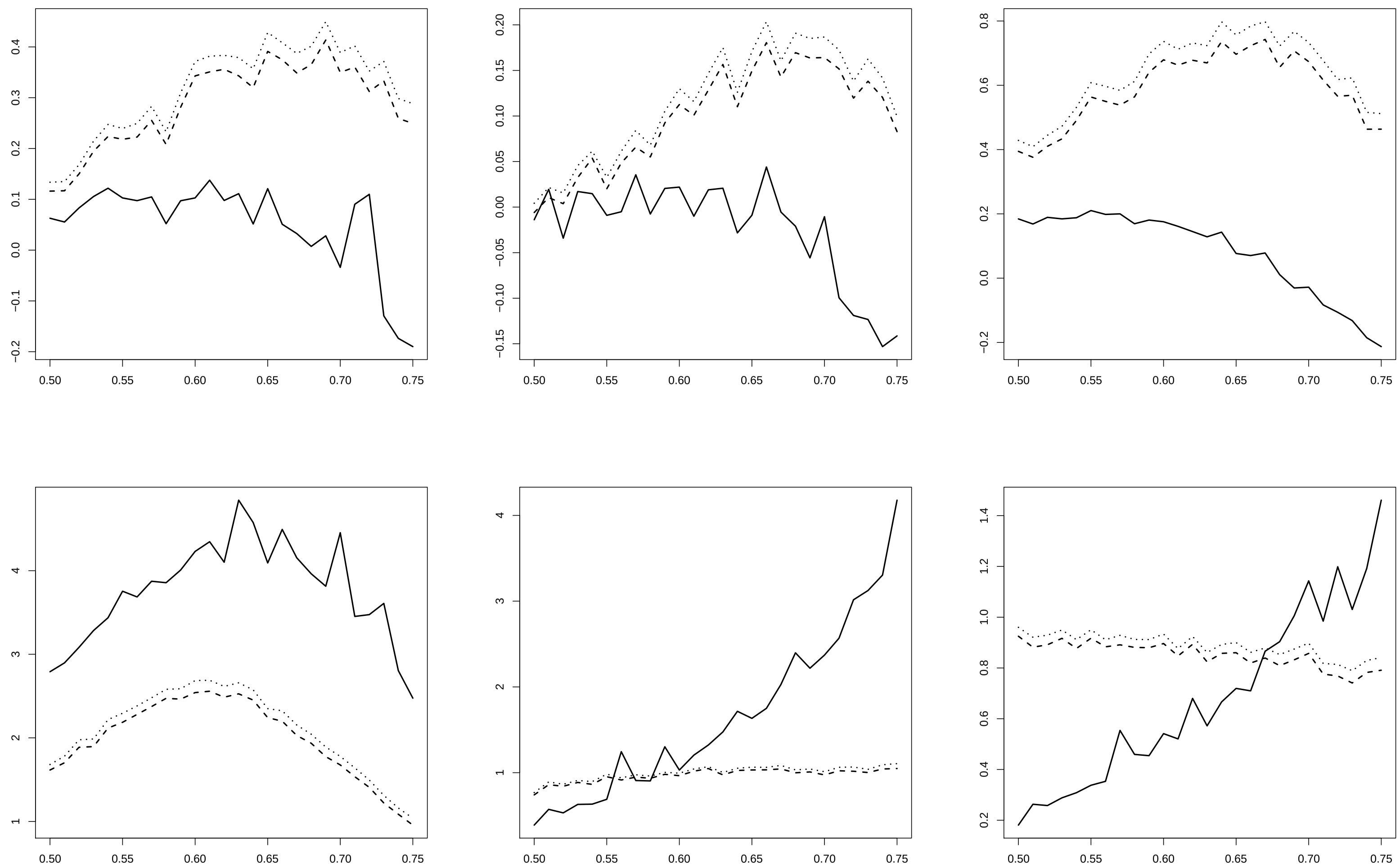

Figure 3: Extreme TVaR estimation with $\delta = 0.99$, for $\gamma \in [0.5, 0.75]$ and on $N = 5000$ replications of a sample of size $n = 100$. Top panels: relative bias in non-conditioned cases, bottom panels: relative bias in conditioned cases. Left: case of the Fréchet distribution, middle: case of the Burr distribution with $\rho = -2$, right: case of the Burr distribution with $\rho = -2/3$. Full line: PL estimator, dotted line: CTrim-PL estimator, dashed line: CWins-PL estimator.

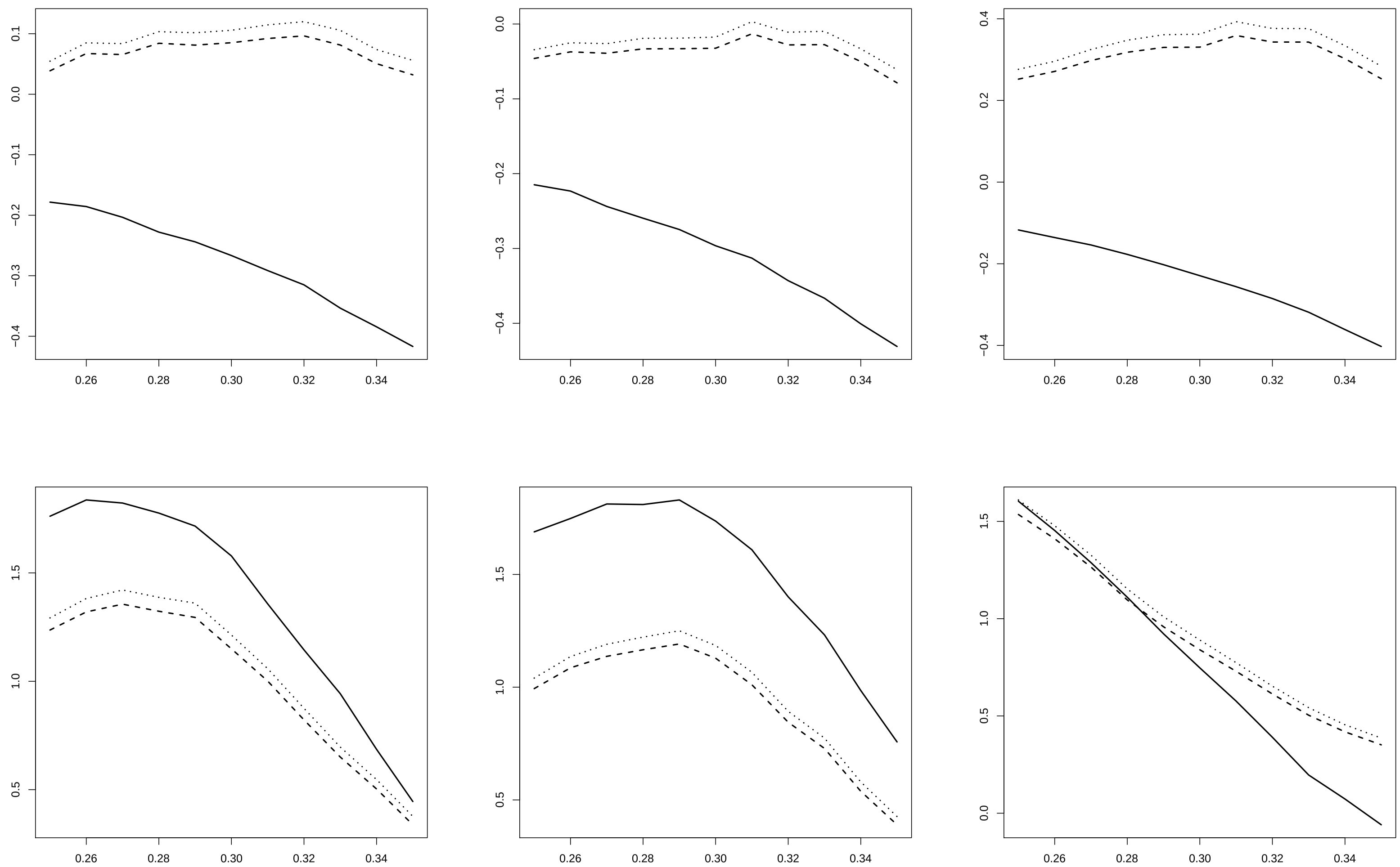

Figure 4: Extreme PH(1/2) risk measure estimation with $\delta = 0.99$, for $\gamma \in [0.25, 0.35]$ and on $N = 5000$ replications of a sample of size $n = 100$. Top panels: relative bias in non-conditioned cases, bottom panels: relative bias in conditioned cases. Left: case of the Fréchet distribution, middle: case of the Burr distribution with $\rho = -2$, right: case of the Burr distribution with $\rho = -2/3$. Full line: PL estimator, dotted line: CTrim-PL estimator, dashed line: CWins-PL estimator.

References

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

El Methni, J., Stupfler, G. (2017). Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. To appear in *Statistica Sinica.* DOI: 10.5705/ss.202015.0460

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

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