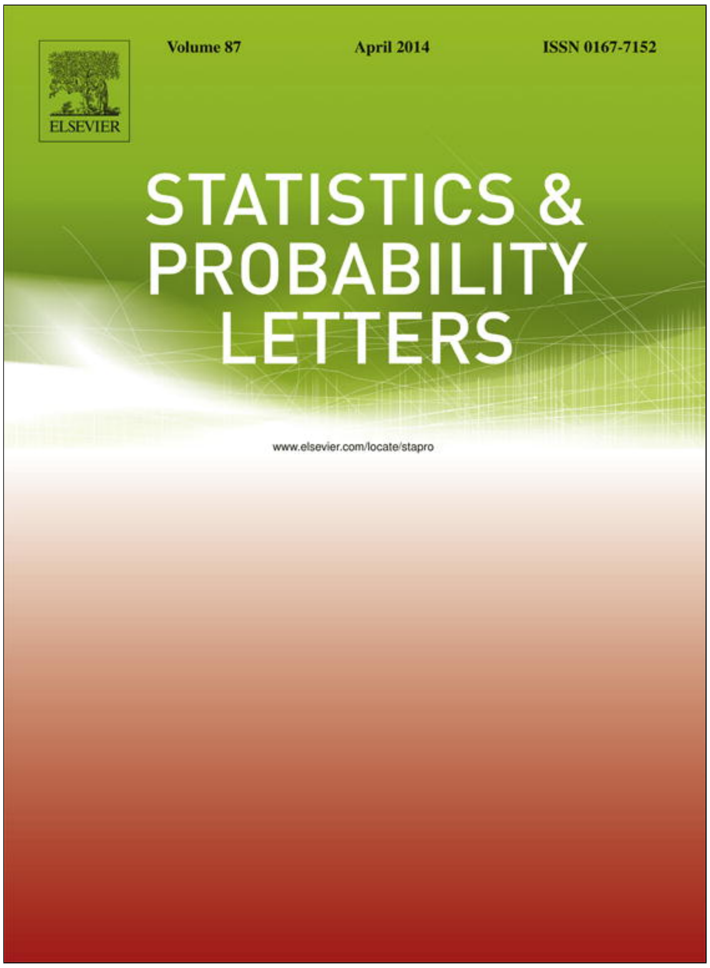
Volume 87
April 2014
ISSN 0167-7152
ELSEVIER
STATISTICS &
PROBABILITY
LETTERS
www.elsevier.com/locate/stapro

Statistics and Probability Letters 87 (2014) 108–114

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Robust and asymptotically unbiased estimation of extreme quantiles for heavy tailed distributions

Yuri Goegebeur [a,*], Armelle Guillou [b], Andréhette Verster [c]

[a] Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

[b] Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

[c] Department of Mathematical Statistics and Actuarial Science, University of the Free State, 205 Nelson Mandela drive, Park West, 9300 Bloemfontein, South Africa

ARTICLE INFO

Article history:
Received 8 November 2013
Received in revised form 9 January 2014
Accepted 14 January 2014
Available online 27 January 2014

Keywords:
Pareto-type distribution
Extreme quantile
Bias correction
Robustness

ABSTRACT

A robust and asymptotically unbiased extreme quantile estimator is derived from a second order Pareto-type model and its asymptotic properties are studied under suitable regularity conditions. The finite sample properties of the proposed estimator are investigated with a small simulation experiment.

1. Introduction

In extreme value statistics, the estimation of extreme quantiles of a distribution function is a central topic. Indeed, many important applications in climatology, finance, actuarial science, hydrology and geology, to name but a few, require extrapolations outside the data range, and extreme value theory provides the only realistic framework for such an exercise. In the present paper we shall address this estimation problem, with special focus on asymptotic unbiasedness and robustness against outliers.

We consider the framework of Pareto-type distributions satisfying a second order condition. In particular, we assume the following (see Beirlant et al., 2009). Let RV_β denote the class of the regularly varying functions at infinity with index β, i.e. Lebesgue measurable ultimately positive functions z satisfying $\lim_{t\to\infty} z(tx)/z(t) = x^\beta$ for all $x > 0$.

Condition $(\mathcal{R})$**.** Let $\gamma > 0$ and $\tau < 0$ be constants. The distribution function F is such that $x^{1/\gamma}\bar{F}(x) \to C \in (0, \infty)$ as $x \to \infty$ and the function δ defined via

$$\bar{F}(x) = Cx^{-1/\gamma}(1 + \gamma^{-1}\delta(x)),$$

is ultimately nonzero, of constant sign and $|\delta| \in RV_\tau$.

* Corresponding author. Tel.: +45 28342140.
E-mail addresses: yuri.goegebeur@imada.sdu.dk (Y. Goegebeur), armelle.guillou@math.unistra.fr (A. Guillou), VersterA@ufs.ac.za (A. Verster).

0167-7152/$ – see front matter
http://dx.doi.org/10.1016/j.spl.2014.01.010

Clearly, condition $(\mathcal{R})$ implies that the tail quantile function U, defined as $U(y) := \inf\{x : F(x) \geq 1 - 1/y\}$, $y > 1$, satisfies $y^{-\gamma}U(y) \to C^{\gamma}$ as $y \to \infty$ and the function a implicitly defined by

$$U(y) = C^{\gamma}y^{\gamma}(1 + a(y)) \tag{1}$$

satisfies $a(y) = \delta(C^{\gamma}y^{\gamma})(1 + o(1))$ as $y \to \infty$, so $|a| \in RV_{\rho}$, with $\rho = \gamma\tau$.

The second order condition $(\mathcal{R})$ can be used to derive the so-called extended Pareto distribution, EPD (Beirlant et al., 2004, 2009), with distribution function given by

$$G(y) = \begin{cases} 1 - [y(1 + \delta - \delta y^{\tau})]^{-1/\gamma}, & y > 1, \\ 0, & y \leq 1, \end{cases} \tag{2}$$

where $\gamma > 0$, $\tau < 0$, and $\delta > \max\{-1, 1/\tau\}$. As shown in Proposition 2.3 of Beirlant et al. (2009), for distribution functions satisfying $(\mathcal{R})$, the distribution function of the relative excess $Y := X/u$ given that $X > u$ can be approximated by (2) with $\delta = \delta(u)$ up to an error that is uniformly $o(\delta(u))$ for $u \to \infty$. In Dierckx et al. (2013), a robust and asymptotically unbiased estimator for γ was introduced by fitting the EPD to a sample of relative excesses by the minimum density power divergence (MDPD) criterion (Basu et al., 1998). In particular, let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with a distribution function satisfying $(\mathcal{R})$, and denote by $X_{1,n} \leq \cdots \leq X_{n,n}$ the corresponding order statistics. The parameters γ and δ of the EPD are then estimated with the minimum density power divergence criterion applied to the relative excesses over the random threshold $u = X_{n-k,n}$, namely $Y_j := X_{n-k+j,n}/X_{n-k,n}$, $j = 1, \ldots, k$, i.e. one minimises the empirical divergence

$$\widehat{\Delta}_{\alpha}(\gamma, \delta) := \int_1^{\infty} g^{1+\alpha}(y)dy - \left(1 + \frac{1}{\alpha}\right)\frac{1}{k}\sum_{j=1}^{k} g^{\alpha}(Y_j),$$

in case $\alpha > 0$, and

$$\widehat{\Delta}_0(\gamma, \delta) := -\frac{1}{k}\sum_{j=1}^{k} \log g(Y_j),$$

in case $\alpha = 0$, where g denotes the density function of G. The parameter ρ is estimated externally, e.g. by using one of the estimators proposed in Fraga Alves et al. (2003) or Goegebeur et al. (2010). Other robust estimators for γ were proposed by e.g. Peng and Welsh (2001); Juárez and Schucany (2004); Vandewalle et al. (2007); Kim and Lee (2008).

In the present paper we will consider robust and asymptotically unbiased extreme quantile estimation under model $(\mathcal{R})$, using the MDPD estimator of Dierckx et al. (2013). Beirlant et al. (2009) studied the asymptotically unbiased estimation of small tail probabilities based on the EPD, fitted by the maximum likelihood method. In Gomes and Pestana (2007) an asymptotically unbiased extreme quantile estimator was introduced for heavy-tailed distributions. These approaches are however not robust against outliers. To the best of our knowledge, robust and asymptotically unbiased extreme quantile estimation has not been considered before.

The remainder of our paper is organised as follows. In the next section we will introduce the robust and asymptotically unbiased estimator for extreme quantiles and study its asymptotic properties under suitable regularity conditions. The finite sample behaviour of the proposed estimator and some alternatives from the literature is illustrated with a small simulation experiment in Section 3.

2. Main result

From the second order condition $(\mathcal{R})$ and using the EPD as approximation to the distribution of X/u_n given $X > u_n$ we can for $\overline{F}(u_n) \to 0$ and $p_n \to 0$ such that $p_n/\overline{F}(u_n) \to c \in [0, \infty)$ as $n \to \infty$ introduce

$$U_0\left(\frac{1}{p_n}\right) := u_n\left(\frac{p_n}{\overline{F}(u_n)}\right)^{-\gamma}\left(1 - \delta(u_n)\left(1 - \left(\frac{p_n}{\overline{F}(u_n)}\right)^{-\rho}\right)\right) \tag{3}$$

as approximation for $U(1/p_n)$.

Lemma 1. *Assume* $(\mathcal{R})$. *If* $\overline{F}(u_n) \to 0$ *and* $p_n \to 0$ *such that* $p_n/\overline{F}(u_n) \to c \in [0, \infty)$ *as* $n \to \infty$ *we have that* $U_0(1/p_n)/U(1/p_n) \to 1$.

The proof of this lemma is straightforward and therefore it is for brevity omitted from the paper. Now, let $X_1, \ldots, X_n$ be i.i.d. random variables with a distribution function satisfying $(\mathcal{R})$, and denote by $X_{1,n} \leq \cdots \leq X_{n,n}$ the corresponding order statistics. Taking $u_n = X_{n-k,n}$, replacing F by the empirical distribution function in (3), and using the fact that $e^{-x} \sim 1 - x$ for $x \to 0$, we can introduce the following extreme quantile estimator

$$\widehat{U}\left(\frac{1}{p_n}\right) := X_{n-k,n}\left(\frac{np_n}{k}\right)^{-\widehat{\gamma}_n} \exp\left(-\widehat{\delta}_n\left(1-\left(\frac{np_n}{k}\right)^{-\widehat{\rho}_n}\right)\right), \tag{4}$$

where $(\widehat{\gamma}_n, \widehat{\delta}_n)$ is the MDPD estimator for (γ, δ) and $\widehat{\rho}_n$ is a consistent estimator sequence for ρ.

In order to study the asymptotic behaviour of $\widehat{U}(1/p_n)$, properly normalised, we need some preliminary results. Firstly, we need the limiting distribution of the MDPD estimator for $(\widehat{\gamma}_n, \widehat{\delta}_n)$. This was already derived in Dierckx et al. (2013), but we repeat the result for completeness here. Let the arrow $\rightsquigarrow$ denote the convergence in distribution as $n \to \infty$, and let $\stackrel{\mathbb{P}}{\to}$ denote the convergence in probability as $n \to \infty$. From now on we denote by γ_0 and ρ_0 the true values of the parameters γ and ρ, respectively, and $\delta_n := \delta(X_{n-k,n})$.

Theorem 1. *Let $X_1, \ldots, X_n$ be a sample of i.i.d. random variables from a distribution function satisfying $(\mathcal{R})$. Then if $k, n \to \infty$ with $k/n \to 0$ and $\sqrt{k}a(n/k) \to \lambda \in \mathbb{R}$, we have that*

$$\sqrt{k}\begin{bmatrix}\hat{\gamma}_n - \gamma_0 \\ \hat{\delta}_n - \delta_n\end{bmatrix} \rightsquigarrow (\Gamma, \Delta)$$

with (Γ, Δ) a bivariate normal random vector,

$$(\Gamma, \Delta) \sim N_2\left(\mathbf{0}, \mathbb{C}^{-1}(\rho_0)\mathbb{B}(\rho_0)\boldsymbol{\Sigma}(\rho_0)\mathbb{B}'(\rho_0)\mathbb{C}^{-1}(\rho_0)\right),$$

where $\boldsymbol{\Sigma}(\rho_0)$ is a symmetric (3×3) matrix with elements

$$\sigma_{11}(\rho_0) := \frac{\alpha^2(1+\gamma_0)^2}{[1+\alpha(1+\gamma_0)]^2[1+2\alpha(1+\gamma_0)]},$$

$$\sigma_{21}(\rho_0) := \frac{\alpha(1+\gamma_0)[\alpha(1+\gamma_0)-\rho_0]}{[1+\alpha(1+\gamma_0)][1-\rho_0+\alpha(1+\gamma_0)][1-\rho_0+2\alpha(1+\gamma_0)]},$$

$$\sigma_{22}(\rho_0) := \frac{[\alpha(1+\gamma_0)-\rho_0]^2}{[1-\rho_0+\alpha(1+\gamma_0)]^2[1-2\rho_0+2\alpha(1+\gamma_0)]},$$

$$\sigma_{31}(\rho_0) := \gamma_0\left(\frac{1}{[1+2\alpha(1+\gamma_0)]^2} - \frac{1}{[1+\alpha(1+\gamma_0)]^3}\right),$$

$$\sigma_{32}(\rho_0) := \gamma_0\left(\frac{1}{[1-\rho_0+2\alpha(1+\gamma_0)]^2} - \frac{1}{[1+\alpha(1+\gamma_0)]^2[1-\rho_0+\alpha(1+\gamma_0)]}\right),$$

$$\sigma_{33}(\rho_0) := \gamma_0^2\left(\frac{2}{[1+2\alpha(1+\gamma_0)]^3} - \frac{1}{[1+\alpha(1+\gamma_0)]^4}\right),$$

$\mathbb{C}(\rho_0)$ is a symmetric (2×2) matrix with elements

$$c_{11}(\rho_0) := \gamma_0^{-\alpha-2}\,\frac{1+\alpha^2(1+\gamma_0)^2}{[1+\alpha(1+\gamma_0)]^3},$$

$$c_{12}(\rho_0) := \gamma_0^{-\alpha-2}\,\frac{\rho_0(1-\rho_0)[1+\alpha(1+\gamma_0)+\alpha^2(1+\gamma_0)^2]+\alpha^3\rho_0(1+\gamma_0)^3}{[1+\alpha(1+\gamma_0)]^2[1-\rho_0+\alpha(1+\gamma_0)]^2},$$

$$c_{22}(\rho_0) := \gamma_0^{-\alpha-2}\,\frac{(1-\rho_0)\rho_0^2+\alpha\rho_0^2(1+\gamma_0)[\alpha(1+\gamma_0)-\rho_0]}{[1+\alpha(1+\gamma_0)][1-\rho_0+\alpha(1+\gamma_0)][1-2\rho_0+\alpha(1+\gamma_0)]},$$

and

$$\mathbb{B}(\rho_0) := \gamma_0^{-\alpha-2}\begin{bmatrix}\gamma_0 & 0 & -1 \\ \gamma_0 & -\gamma_0(1-\rho_0) & 0\end{bmatrix}.$$

Secondly, we need the limiting distribution of the intermediate order statistic $X_{n-k,n}$ under $(\mathcal{R})$, properly normalised.

Lemma 2. *Let $X_1, \ldots, X_n$ be a sample of i.i.d. random variables from a distribution function satisfying $(\mathcal{R})$. For $k, n \to \infty$ such that $k = o(n)$ and $\sqrt{k}a(n/k) \to \lambda \in \mathbb{R}$ we have that*

$$\sqrt{k}\left(\frac{X_{n-k,n}}{U(n/k)} - 1\right) \rightsquigarrow \mathbb{X}$$

where $\mathbb{X}$ is a normal random variable, $\mathbb{X} \sim N(0, \gamma_0^2)$.

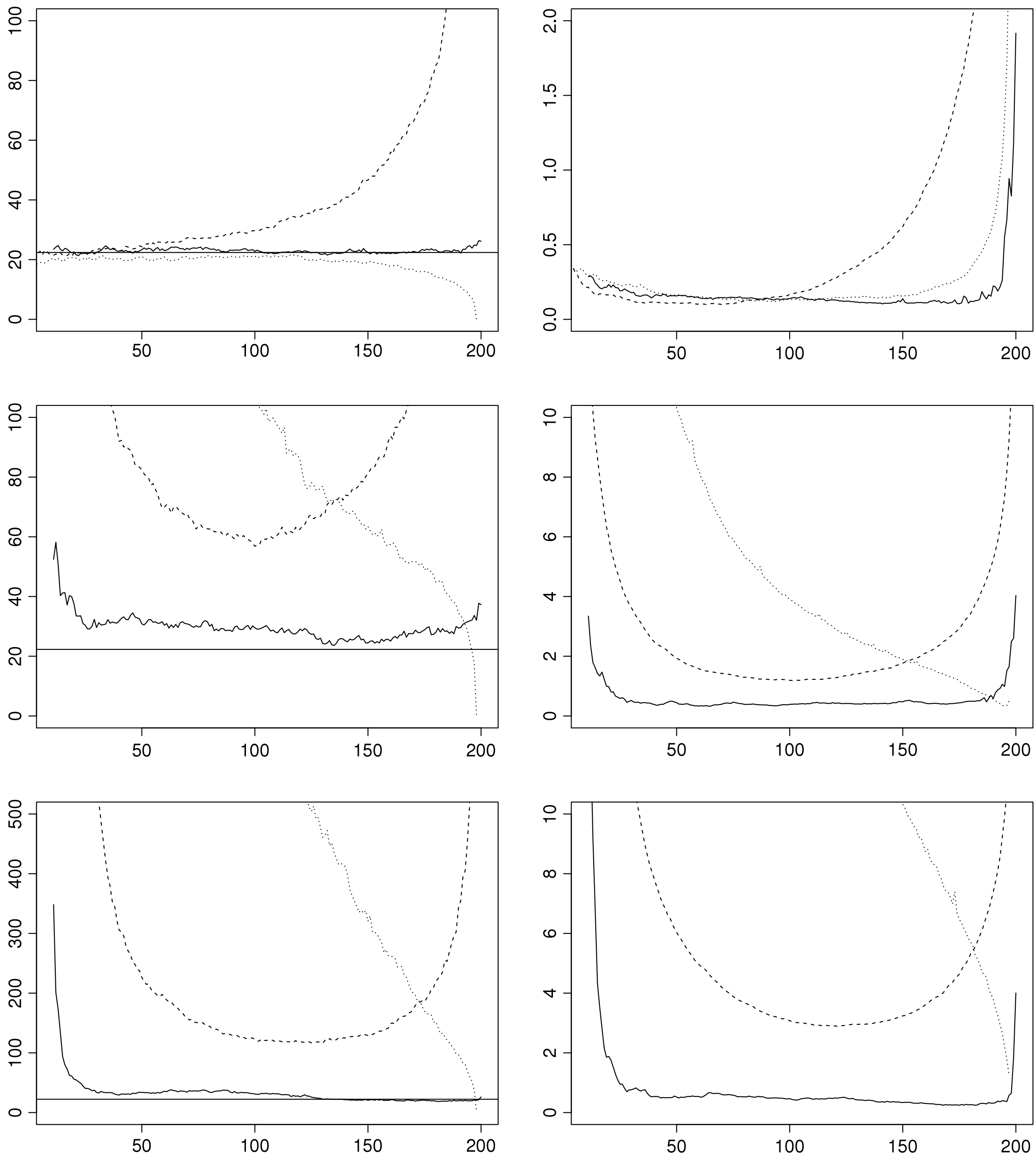

Fig. 1. Fréchet simulation, quantile 1–1/500. Median (left) and MSE (right) of the MDPD based estimator (solid), MLE based estimator (dotted) and Weissman estimator (dashed). No contamination (top), 1% contamination (middle) and 2% contamination (bottom).

In the next theorem we state the limiting distribution of the extreme quantile estimator (4), when properly normalised.

Theorem 2. *Let $X_1, \ldots, X_n$ be a sample of i.i.d. random variables from a distribution function satisfying $(\mathcal{R})$. Then if $k \to \infty$ as $n \to \infty$ with $k/n \to 0$, $\sqrt{k}a(n/k) \to \lambda \in \mathbb{R}$, $np_n/k \to 0$ and $\ln(np_n)/\sqrt{k} \to 0$ we have that*

$$\frac{\sqrt{k}}{\ln \frac{k}{np_n}} \left(\frac{\widehat{U}\left(\frac{1}{p_n}\right)}{U\left(\frac{1}{p_n}\right)} - 1 \right) \rightsquigarrow \Gamma.$$

Theorem 2 indicates that the normalised extreme quantile estimator inherits the asymptotic distribution of the MDPD estimator for γ_0. As shown in Dierckx et al. (2013), the MDPD estimator for γ_0 based on the EPD is robust against outliers and asymptotically unbiased.

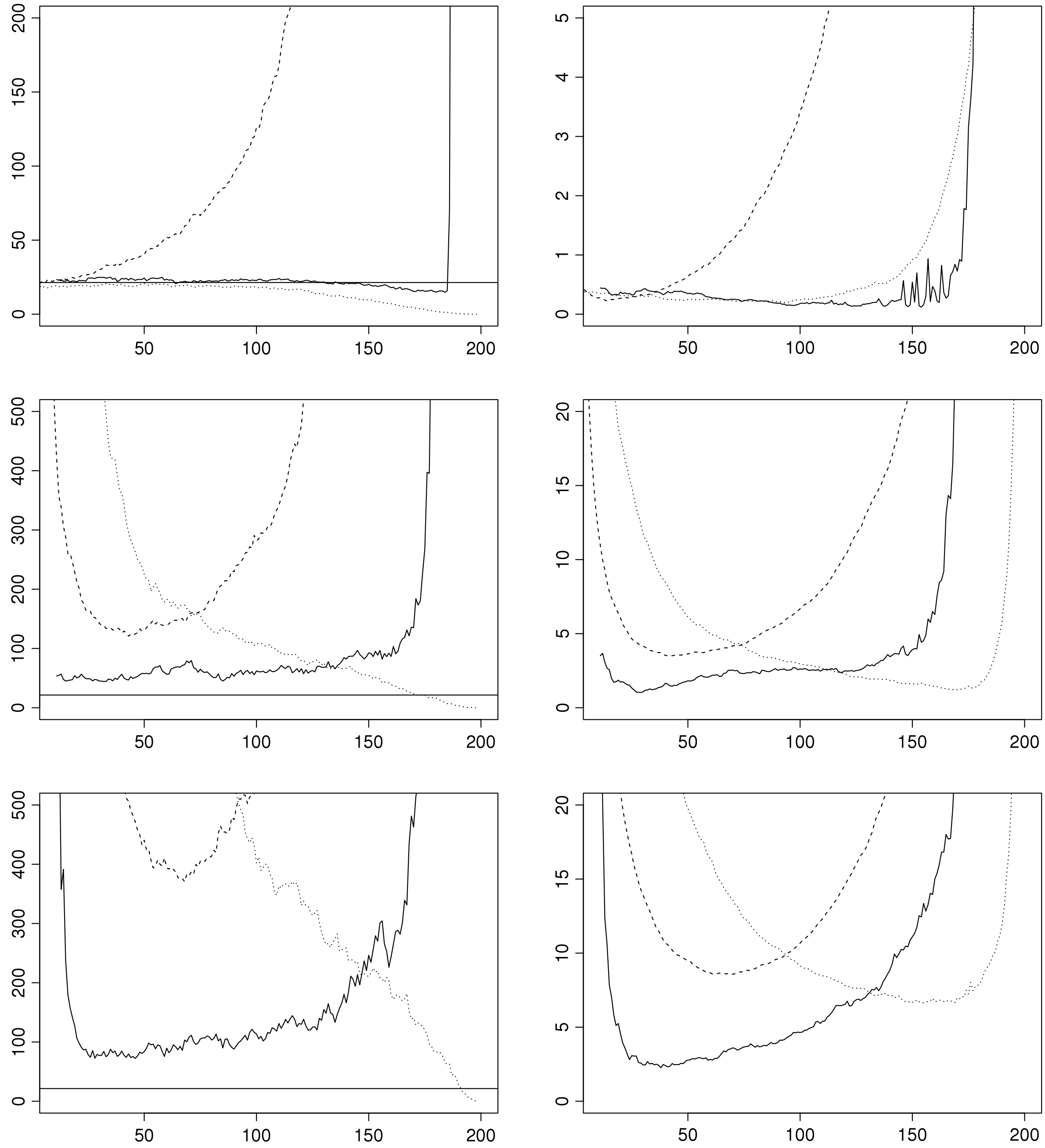

Fig. 2. Burr simulation, quantile 1–1/500. Median (left) and MSE (right) of the MDPD based estimator (solid), MLE based estimator (dotted) and Weissman estimator (dashed). No contamination (top), 1% contamination (middle) and 2% contamination (bottom).

3. Simulation experiment

In this section we investigate the finite sample properties of $\widehat{U}(1/p_n)$ as given in (4) with different parameter estimators, in particular the MDPD estimator $\widehat{\gamma}_n$ and $\widehat{\delta}_n$ with $\alpha > 0$, and the maximum likelihood estimator (corresponding to MDPD with $\alpha = 0$, see also Beirlant et al., 2009). We also consider the Weissman estimator (Weissman, 1978) given by

$$\widehat{U}_W(1/p_n) = X_{n-k,n}\left(\frac{np_n}{k}\right)^{-H_{k,n}},$$

with $H_{k,n}$ being Hill's estimator (Hill, 1975). For the parameter ρ we use the estimator of Fraga Alves et al. (2003).

Figs. 1 and 2 illustrate the results of a small simulation study based on 100 datasets, each of size $n = 200$, simulated from the distributions given below. The same distributions were considered in Dierckx et al. (2013).

- Uncontaminated Fréchet distribution: $F(x) = \exp(-x^{-\beta}), x > 0, \beta > 0$, denoted Fréchet($\beta$). For this study β was chosen as 2.

- Contaminated Fréchet distribution: $F_\epsilon(x) = (1-\epsilon)F(x) + \epsilon\tilde{F}(x)$ where $F(x)$ represents the uncontaminated Fréchet(2) and $\tilde{F}(x) = 1-(x/x_c)^{-\beta}$, $x > x_c$ where β is chosen as 0.5 and $x_c = 2$ times the 99.99% quantile of the uncontaminated Fréchet(2). We take $\epsilon = 0.01$ and $\epsilon = 0.02$.
- Uncontaminated Burr distribution: $F(x) = 1-(\eta/(\eta+x^\tau))^\lambda$, $x > 0$, $\eta, \tau, \lambda > 0$, denoted Burr(η, τ, λ). For this study we have chosen $\eta = 1$, $\tau = 1$ and $\lambda = 2$.
- Contaminated Burr distribution: $F_\epsilon(x) = (1-\epsilon)F(x)+\epsilon\tilde{F}(x)$ where $F(x)$ represents the uncontaminated Burr(1,1,2) and $\tilde{F}(x) = 1-(x/x_c)^{-\beta}$, $x > x_c$ where $\beta = 0.5$ and $x_c = 1.2$ times the 99.99% quantile of the uncontaminated Burr(1,1,2). We take $\epsilon = 0.01$ and $\epsilon = 0.02$.

We report only the results for quantile 1–1/500. The 1–1/1000 quantile was also considered and resulted in similar outcomes.

In Figs. 1 and 2, the left panels show the median of the extreme quantile estimators and the right panels the mean squared error (MSE) of $\ln(\widehat{U}_*(1/p_n)/U(1/p_n))$, where $\widehat{U}_*(1/p_n)$ denotes any of the considered estimators of $U(1/p_n)$, as a function of k. The true quantile is indicated by the horizontal reference line on the left panels. For our MDPD based estimators for extreme quantiles we set $\alpha = 0.1$ in uncontaminated cases and $\alpha = 0.5$ in contaminated cases. These values are motivated by a simulation experiment where the MDPD based estimators were implemented for different values of α, and where $\alpha = 0.1$ and 0.5 showed a good overall performance (we refer to the technical report of the paper, Goegebeur et al., 2013). The top panels of Figs. 1 and 2 illustrate the behaviour of MDPD estimator with $\alpha = 0.1$ (solid line), the MLE based estimator (dotted line) and the Weissman estimator (dashed line) in absence of contamination. As expected, the second order estimators for $U(500)$ (MDPD and MLE) show sample paths that are much more stable around the true value of the quantile than those of the Weissman estimator, which only touch the true quantile value at the smaller values of k, especially in case of the Burr distribution where $\rho_0 = -0.5$. Comparing the behaviour of the MDPD and MLE based estimators, we see a slight advantage for the MDPD estimators. Also in terms of MSE the MDPD based estimator is very competitive relative to the benchmarks. The middle and bottom panels of Figs. 1 and 2 show the behaviour of the estimators under 1% and 2% of contamination, respectively. Note that the generated contamination is quite severe in terms of shift of distribution and tail heaviness, especially if one takes the small sample size of $n = 200$ into account (the actual estimation is even based on only the top k observations). Obviously, increasing the fraction of contamination negatively affects all estimators. In terms of bias the robust MDPD based estimator clearly outperforms the non-robust estimators, and shows quite stable sample paths. Also in terms of MSE the MDPD estimator performs best, in particular it shows a low MSE value for a wide range of k values.

Acknowledgements

The authors are very grateful to the two anonymous referees for their constructive comments on the initial version of the paper.

Appendix

Proof of Lemma 2. Using the inverse probability integral transform we have that $X_{n-k,n} \stackrel{D}{=} U(Y_{n-k,n})$, where $Y_{n-k,n}$ denotes the order statistic $n-k$ of a random sample $Y_1, \ldots, Y_n$ from the unit Pareto distribution with distribution function $H(y) = 1-1/y$, $y > 1$. Thus,

$$\begin{aligned}\sqrt{k}\ln\frac{X_{n-k,n}}{U(n/k)} &\stackrel{D}{=} \sqrt{k}\ln\frac{U(Y_{n-k,n})}{U(n/k)}\\ &= \gamma_0\sqrt{k}\ln\left(\frac{k}{n}Y_{n-k,n}\right) + \sqrt{k}\ln\frac{1+a(Y_{n-k,n})}{1+a(n/k)}\\ &=: L_1 + L_2.\end{aligned}$$

For L_1, use the well-known fact that $\sqrt{k}(k/nY_{n-k,n} - 1) \rightsquigarrow Z$ where $Z \sim N(0,1)$ (see for instance Corollary 2.2.2 in de Haan and Ferreira, 2006) and the delta method to obtain that $L_1 \rightsquigarrow \mathbb{X}$ under the conditions of the lemma. For the term L_2, a straightforward application of Taylor's theorem gives

$$L_2 = \sqrt{k}a(n/k)\left(\frac{a(Y_{n-k,n})}{a(n/k)} - 1 + o(1) + o\left(\frac{a(Y_{n-k,n})}{a(n/k)}\right)\right).$$

Since a is regularly varying we have that $a(tx)/a(t) \to x^\rho$ as $t \to \infty$, locally uniformly for $x > 0$. Combining this with the fact that $k/nY_{n-k,n} \to 1$ a.s. and the assumption $\sqrt{k}a(n/k) \to \lambda \in \mathbb{R}$ we have that $L_2 \stackrel{\mathbb{P}}{\to} 0$. Lemma 2 follows then by collecting the terms and another application of the delta method.

Proof of Theorem 2. First we comment on the joint convergence in distribution of the random vector

$$\left(\sqrt{k}(\widehat{\gamma}_n - \gamma_0), \sqrt{k}(\widehat{\delta}_n - \delta_n), \sqrt{k}(X_{n-k,n}/U(n/k) - 1), \widehat{\rho}_n\right).$$

According to the proof of Theorem 2 in Dierckx et al. (2013), we have that

$$\left(\sqrt{k}(\widehat{\gamma}_n-\gamma_0),\sqrt{k}\widehat{\delta}_n\right)\rightsquigarrow(\Gamma,\widetilde{\Delta}),$$

where $(\Gamma,\widetilde{\Delta})\sim N_2((0,\lambda),\mathbb{C}^{-1}(\rho_0)\mathbb{B}(\rho_0)\boldsymbol{\Sigma}(\rho_0)\mathbb{B}'(\rho_0)\mathbb{C}^{-1}(\rho_0))$. From the proof of Lemma 1 and Theorem 2 in Dierckx et al. (2013) we can deduce that $\widehat{\gamma}_n$ and $\widehat{\delta}_n$ are independent of $X_{n-k,n}$, and therefore

$$\left(\sqrt{k}(\widehat{\gamma}_n-\gamma_0),\sqrt{k}\widehat{\delta}_n,\sqrt{k}(X_{n-k,n}/U(n/k)-1)\right)\rightsquigarrow(\Gamma,\widetilde{\Delta},\mathbb{X}),$$

where $(\Gamma,\widetilde{\Delta},\mathbb{X})\sim N_3((0,\lambda,0),\Psi)$, with

$$\Psi:=\begin{bmatrix}\mathbb{C}^{-1}(\rho_0)\mathbb{B}(\rho_0)\boldsymbol{\Sigma}(\rho_0)\mathbb{B}'(\rho_0)\mathbb{C}^{-1}(\rho_0) & \mathbf{0}\\ \mathbf{0} & \gamma_0^2\end{bmatrix}.$$

Finally, using the fact that $\sqrt{k}\delta_n\stackrel{\mathbb{P}}{\to}\lambda$ and $\widehat{\rho}_n\stackrel{\mathbb{P}}{\to}\rho_0$ we have also that

$$\left(\sqrt{k}(\widehat{\gamma}_n-\gamma_0),\sqrt{k}(\widehat{\delta}_n-\delta_n),\sqrt{k}(X_{n-k,n}/U(n/k)-1),\widehat{\rho}_n\right)\rightsquigarrow(\Gamma,\Delta,\mathbb{X},\rho_0).$$

Now, consider $\ln(\widehat{U}(1/p_n)/U(1/p_n))$. Let $d_n:=k/(np_n)$. Straightforward calculations give

$$\begin{aligned}\ln\frac{\widehat{U}\left(\frac{1}{p_n}\right)}{U\left(\frac{1}{p_n}\right)} &= \ln\frac{X_{n-k,n}}{U\left(\frac{n}{k}\right)}+(\widehat{\gamma}_n-\gamma_0)\ln d_n+\ln\frac{1+a\left(\frac{n}{k}\right)}{1+a\left(\frac{1}{p_n}\right)}-\widehat{\delta}_n\left(1-d_n^{\widehat{\rho}_n}\right)\\ &=: T_1+T_2+T_3-T_4.\end{aligned}\tag{5}$$

Clearly $T_1=O_{\mathbb{P}}(1/\sqrt{k})$ by Lemma 2 and $T_2=O_{\mathbb{P}}\left(\frac{\ln d_n}{\sqrt{k}}\right)$ by Theorem 1. From Taylor's theorem we can write

$$T_3=a(n/k)\left(1-\frac{a\left(\frac{1}{p_n}\right)}{a\left(\frac{n}{k}\right)}+o(1)+o\left(\frac{a\left(\frac{1}{p_n}\right)}{a\left(\frac{n}{k}\right)}\right)\right).$$

By using the regular variation properties of the function a and the fact that $d_n\to\infty$ we have that $a(1/p_n)/a(n/k)\to 0$, and thus under the conditions of the theorem $T_3=O(a(n/k))$. Finally, for T_4 note that $d_n^{\widehat{\rho}_n}=o_{\mathbb{P}}(1)$ and $\widehat{\delta}_n=O_{\mathbb{P}}(1/\sqrt{k})$. Collecting all the terms we see thus that the rate of convergence of $\ln(\widehat{U}(1/p_n)/U(1/p_n))$ is given by $\frac{\ln d_n}{\sqrt{k}}$. Multiplying both sides of (5) by $\sqrt{k}/\ln d_n$, the result of the theorem follows.

References

Basu, A., Harris, I.R., Hjort, N.L., Jones, M.C., 1998. Robust and efficient estimation by minimizing a density power divergence. Biometrika 85, 549–559.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes—Theory and Applications. Wiley Series in Probability and Statistics.

Beirlant, J., Joossens, E., Segers, J., 2009. Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. J. Statist. Plann. Inference 139, 2800–2815.

de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An Introduction. Springer.

Dierckx, G., Goegebeur, Y., Guillou, A., 2013. An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. J. Multivariate Anal. 121, 70–86.

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

Goegebeur, Y., Beirlant, J., de Wet, T., 2010. Kernel estimators for the second order parameter in extreme value statistics. J. Statist. Plann. Inference 140, 2632–2652.

Goegebeur, Y., Guillou, A., Verster, A., 2013. Robust and asymptotically unbiased estimation of extreme quantiles for heavy tailed distributions. Technical report available on http://hal.archives-ouvertes.fr/hal-00880530.

Gomes, M.I., Pestana, D., 2007. A sturdy reduced-bias extreme quantile (VaR) estimator. J. Amer. Statist. Assoc. 102, 280–291.

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

Juárez, S.F., Schucany, W.R., 2004. Robust and efficient estimation for the generalized Pareto distribution. Extremes 7, 237–251.

Kim, M., Lee, S., 2008. Estimation of a tail index based on minimum density power divergence. J. Multivariate Anal. 99, 2453–2471.

Peng, L., Welsh, A.H., 2001. Robust estimation of the generalized Pareto distribution. Extremes 4, 53–65.

Vandewalle, B., Beirlant, J., Christmann, A., Hubert, M., 2007. A robust estimator for the tail index of Pareto-type distributions. Comput. Statist. Data Anal. 51, 6252–6268.

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