Ann Inst Stat Math (2014) 66:193–220
DOI 10.1007/s10463-013-0412-2

The harmonic moment tail index estimator: asymptotic distribution and robustness

Jan Beran · Dieter Schell · Milan Stehlík

Received: 15 August 2012 / Revised: 29 March 2013 / Published online: 18 June 2013

Abstract Asymptotic properties of the harmonic moment tail index Estimator are derived for distributions with regularly varying tails. The estimator shows good robustness properties and stands out for its simplicity. A tuning parameter allows for regulating the trade-off between robustness and efficiency. Small sample properties are illustrated by a simulation study.

Keywords Tail index estimation · Regularly varying tail · Hill estimator · Robustness · Asymptotic distribution

1 Introduction

Many distributions in applied fields such as hydrology, insurance or finance belong to the maximum domain of attraction of the Fréchet distribution which is characterized by heavy-tailed distributions (see e.g. Embrechts et al. 1997; Beirlant et al. 2004; Reiss and Thomas 2005; Resnick 2007). Recall that a (cumulative) distribution F is called heavy-tailed, if $1 - F$ is regularly varying with index $-1/\gamma$, i.e.

Electronic supplementary material The online version of this article (doi:10.1007/s10463-013-0412-2) contains supplementary material, which is available to authorized users.

J. Beran (✉) · D. Schell
Department of Mathematics and Statistics, University of Konstanz,
Universitätsstr. 10, 78457 Konstanz, Germany
e-mail: Jan.Beran@uni-konstanz.de

M. Stehlík
Institute of Applied Statistics, Johannes Keppler University,
Altenbergerstr. 69, 4040 Linz, Austria

$$\lim_{t\to\infty}\frac{1-F(tx)}{1-F(t)}=x^{-1/\gamma}. \tag{1}$$

An equivalent characterization can be given in terms of $U(t) := F^{\leftarrow}(1-t^{-1})$ (with $F^{\leftarrow}(x) := \inf\{y : F(y) \geq x\}$) by

$$\lim_{t\to\infty}\frac{U(tx)}{U(t)}=x^{\gamma}\ (x>1). \tag{2}$$

Suppose we have an i.i.d. sample $X_1, \ldots, X_n$ with common distribution F satisfying (1) and denote by $X_{1,n} \leq \cdots \leq X_{n,n}$ the corresponding order statistics. The best known method for estimating the tail index γ is Hill's estimator (Hill 1975)

$$\gamma_{n,k}^{(H)}=k^{-1}\sum_{i=1}^{k}\log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)\quad (k=1,\ldots,n-1).$$

Consistency of $\gamma_{n,k}^{(H)}$ for i.i.d. observations was first derived by Mason (1982) for $k \to \infty$, $n \to \infty$ and $k/n \to 0$. Under varying conditions on k and the second-order behavior of F, asymptotic normality of $\gamma_{n,k}^{(H)}$ was derived by many authors (see e.g. Hall 1982; Davis and Resnick 1984; Csörgő and Mason 1985; Csörgő et al. 1985; Haeusler and Teugels 1985; Csörgő and Viharos 1997; de Haan and Resnick 1998; de Haan and Peng 1998). An important specification in this context is the rate at which $\bar{F}(tx)/\bar{F}(t)$ converges towards $x^{-1/\gamma}$ (or equivalently, how fast $U(tx)/U(t)$ converges towards x^{γ}). A typical second-order assumption is (de Haan and Peng 1998; de Haan and Ferreira 2006)

$$\lim_{t\to\infty}\frac{1}{A(t)}\left(\frac{U(tx)}{U(t)}-x^{\gamma}\right)=x^{\gamma}\frac{x^{\rho}-1}{\rho}, \tag{3}$$

where $\rho \leq 0$ is a second-order parameter and $|A|$ is some regularly varying function with index ρ. Under this assumption, de Haan and Peng (1998) derived the asymptotic expansion

$$\gamma_{n,k}^{(H)}\stackrel{d}{=}\gamma+\frac{\gamma}{\sqrt{k}}Z_k+\frac{A(n/k)}{1-\rho}(1+o_p(1)),$$

where $Z_k = \sqrt{k}\left(k^{-1}\sum_{i=1}^{k} E_i - 1\right)$ and E_i is a sequence of i.i.d. standard exponential random variables. Hence, choosing k such that $\lim_{n\to\infty}\sqrt{k}A(n/k) = \lambda \neq 0$ leads to asymptotic normality of $\sqrt{k}\left(\gamma_{n,k}^{(H)} - \gamma\right)$ with mean $\lambda/(1-\rho)$ and variance γ^2.

The aim of the current paper is to provide a similar expansion for harmonic moment tail index estimators. This class includes Hill's estimator as a special case. However, it also includes methods that have robustness properties not shared by Hill's estimator. All estimators in this class are explicit and the degree of robustness can be controlled by a tuning parameter.

Robust tail index estimation has received some attention in the recent literature. For instance, Brazauskas and Serfling (2000) studied the lack of robustness of classical tail index estimators in a simple one-parameter Pareto model $F(x) = 1 - (\sigma/x)^{1/\gamma}$, $(x \geq \sigma)$ with $\sigma > 0$ known. Motivated by their findings, Finkelstein et al. (2006) proposed an M-estimator of γ based on the probability integral transform. Their idea is closely related to the harmonic moment tail index estimator (HME) introduced in Fabián and Stehlík (2009) and Henry (2009) which can in principle be used for estimating the tail index of any distribution with a regularly varying tail. Consistency and asymptotic normality are derived in Henry (2009) under the very restrictive assumption that there is an $x_0 > 0$ such that the conditional distribution given $X > x_0$ is exactly Pareto, i.e. the exact equality $P(X > x) = cx^{-1/\gamma}$ holds for $x > x_0$ and some $c, \gamma > 0$. A special case of the HME is discussed in Stehlík et al. (2010), where it is motivated by considering t scores (also see Stehlík et al. 2012 for some consistency results). Robust tail index estimation for the generalized Pareto distribution is also considered in Peng and Welsh (2002) and Juárez and Schucany (2004). Approaches to robust tail index estimation in the general class of regularly varying tails are proposed for instance in Beirlant et al. (2004), Vandewalle (2004), Vandewalle et al. (2004, 2007) and Knight (2012). Beran and Schell (2012) address the issue of robustness with respect to low quantiles using M-estimation. Different authors point out the apparent contradiction between robust statistics and extreme value theory, since classical robustness procedures try to reduce the influence of extreme observations, whereas the methods of extreme value theory mainly focus on exactly these data points (see e.g. Vandewalle et al. 2004). Nevertheless, as pointed out by Vandewalle (2004), small errors in the estimation of the tail index can cause large errors in quantities based on the tail index. Therefore robust procedures are needed to avoid undue sensitivity of the estimate to a small portion of "outlying" observations.

The article is organized as follows. The class of harmonic moment tail index estimators is defined in Sect. 2. Asymptotic properties are obtained in Sect. 3. Some robustness properties are discussed in Sect. 4. The results are illustrated by a small simulation study in Sect. 5. We conclude with some final comments in Sect. 6. Proofs are given in the Appendix.

2 Motivation and definition of the harmonic moment tail index estimator

To motivate the class of HME we start with a Pareto distributed variable X. Thus, $X \geq 1$ and $F(x) = 1 - x^{-1/\gamma}$ for some $\gamma > 0$. For an arbitrary threshold $x_0 > 1$ let $Y = x_0^{-1} X \cdot 1\{X > x_0\}$. Then, conditionally on $X > x_0$, Y is again Pareto distributed with parameter γ, i.e.

$$P(Y \leq y \mid X > x_0) = F_Y(y | X > x_0) = 1 - y^{-1/\gamma}.$$

Moreover, conditionally on $X > x_0$, we have

$$Y^{-1/\gamma} = 1 - F(Y) \overset{d}{=} U,$$

where U is uniformly distributed on [0, 1]. Note that $E[U^p] = (p+1)^{-1}$. Given an i.i.d. sequence $Y_1, \ldots, Y_k$ of relative exceedances over a threshold x_0, the strong law of large numbers implies that

$$k^{-1}\sum_{i=1}^{k} Y_i^{1-\beta} = k^{-1}\sum_{i=1}^{k}\left(Y_i^{-1/\gamma}\right)^{\gamma(\beta-1)} \overset{d}{=} \frac{1}{k}\sum_{i=1}^{k} U_i^{\gamma(\beta-1)}$$
$$\overset{a.s.}{\to} E[U^{\gamma(1-\beta)}] = 1/(\gamma(\beta-1)+1),$$

provided that $\gamma(\beta-1)+1 \neq 0$. Hence,

$$\frac{1}{\beta-1}\left(\frac{1}{k^{-1}\sum_{i=1}^{k} Y_i^{1-\beta}} - 1\right) \overset{a.s.}{\to} \gamma.$$

Suppose now that the distribution of X is not necessarily Pareto but any heavy-tailed distribution defined by (1). The idea is now the same as for Hill's estimator, namely to replace x_0 by an order statistic $X_{n-k,n}$. Thus, we obtain the following definition.

Definition 1 (*Harmonic moment tail index estimators*) Let $X_1, \ldots, X_n$ be i.i.d random variables with distribution F of the form (1). The HME of γ is defined by

$$H_{n,k}^{(\beta)} := \frac{1}{\beta-1}\left\{\left[k^{-1}\sum_{i=1}^{k}\left(\frac{X_{n-k,n}}{X_{n-i+1,n}}\right)^{\beta-1}\right]^{-1} - 1\right\},$$

where $1 \leq k \leq n-1$ and $\beta > 0$ is a tuning parameter. For $\beta = 1$, $H_{n,k}^{(\beta)}$ is interpreted as a limit for $\beta \to 1$, i.e.

$$H_{n,k}^{(1)} := \lim_{\beta\to 1} H_{n,k}^{(\beta)} = k^{-1}\sum_{i=1}^{k}\log\left(X_{n-i+1,n}/X_{n-k,n}\right)$$

is Hill's estimator $\gamma_{n,k}^{(H)}$.

As it turns out, the tuning parameter β allows for regulating the trade-off between efficiency and robustness. For $\beta > 1$ the effect of large contaminations is bounded, since the HME benefits from the properties of the harmonic mean. However, a larger value of β also implies an increased variance. For $\beta < 1$ the harmonic moment tail index estimator also has a higher variance than Hill's estimator. However, in some situations, it possesses a smaller asymptotic bias such that the AMSE is smaller. The class of harmonic moment tail index estimators is introduced in Henry (2009), using the tuning parameter $\theta = 1/(\beta-1)$. Asymptotic results are obtained however only under the trivial and very restrictive assumption of an exact Pareto tail beyond a fixed finite threshold u. In a related paper, Stehlík et al. (2010) propose a t score moment estimator which is a harmonic moment tail index estimator with $\beta = 2$, but only simulation results are reported.

3 Asymptotic properties of the harmonic moment tail index estimator

To prove consistency, we use a similar approach as in Resnick (2007). Let $E_x = (x, \infty]$ and denote by $\mathcal{E}$ the Borel σ-field on $E = E_0 = (0, \infty]$. The so-called tail measure $\nu_\gamma : \mathcal{E} \to \mathbb{R}_+$ is defined by $\nu_\gamma(E_x) := x^{-1/\gamma}$. Moreover, let $M_+(E)$ be the space of nonnegative Radon measures on E endowed with the vague topology. Vague convergence in $M_+(E)$, which is consistent with the metric generating the vague topology, is defined as follows.

Definition 2 Given a sequence $\{\mu_n, n \geq 0\}$ with $\mu_i \in M_+(E)$, μ_n is said to converge vaguely to μ_0 ($\mu_n \overset{v}{\to} \mu_0$) if

$$\mu_n(f) := \int_E f(x)\mu_n(\mathrm{d}x) \to \mu_0(f) := \int_E f(x)\mu_0(\mathrm{d}x)$$

for all $f \in C_K^+(E) := \{f : E \to \mathbb{R}_+ : f \text{ is continuous with compact support}\}$ as $n \to \infty$.

Note that endowed with the vague metric $M_+(E)$ is a complete, separable, metric space. According to Resnick (2007, Theorem 4.1) a regularly varying tail, i.e. $X \sim F$, where $1 - F \in RV_{-1/\gamma}$, implies

$$\frac{n}{k} P\left(\frac{X}{U(n/k)} \in \cdot\right) \overset{v}{\to} \nu_\gamma(\cdot) \tag{4}$$

in $M_+(0, \infty]$ as $n \to \infty$ and $k(n) \to \infty$ with $k/n \to 0$. Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables and define for $k = k(n) \leq n$ the tail empirical measure by

$$\nu_{n,k}(\cdot) := \frac{1}{k} \sum_{i=1}^{n} 1\{(X_i/U(n/k)) \in \ \cdot \ \}.$$

We emphasize that $\nu_{n,k}$ depends on k. The tail empirical measure $\nu_{n,k}$ is a natural estimator of $n/kP\,(X/U(n/k) \in \cdot)$, since for any fixed $t > 0$,

$$\frac{n}{k} P\left(\frac{X}{U(n/k)} \in (t, \infty]\right) = \frac{n}{k}\left(1 - F\left(tU\left(\frac{n}{k}\right)\right)\right).$$

Estimating F by the empirical distribution function, denoted by F_n, results in

$$\frac{n}{k}\left(1 - F_n\left(tU\left(\frac{n}{k}\right)\right)\right) = \frac{1}{k}\sum_{i=1}^{n} 1\{X_i/U(n/k) \in (t, \infty]\} = \nu_{n,k}(t, \infty].$$

To be more precise (see Resnick 2007, Theorem 4.1), one can show

$$\nu_{n,k} \Rightarrow \nu_\gamma,$$

in $M_+(0,\infty]$ provided that $n \to \infty$, $k(n) \to \infty$ and $k(n)/n \to 0$, where $\Rightarrow$ stands for weak convergence in $M_+(0,\infty]$ (endowed with the vague metric). Based on this, we are ready to prove the consistency of the HME $H_{n,k}^{(\beta)}$.

Theorem 1 (Consistency of the HME) *If $k(n)/n \to 0$, then $\nu_{n,k} \Rightarrow \nu_\gamma$ implies the consistency of $H_{n,k}^{(\beta)}$, i.e.*

$$H_{n,k}^{(\beta)} \xrightarrow{P} \gamma.$$

Intuitively, the proof is based on the following arguments. Since $\nu_{n,k}$ depends on U, and therefore on the unknown distribution function F, we define an estimator, say $\hat{\nu}_{n,k}$, of $\nu_{n,k}$ and show its convergence to ν_γ in probability. Rewriting $H_{k,n}^{(\beta)}$ as a functional of $\hat{\nu}_{n,k}$ and using Theorem 4.2 in Billingsley (1968) in conjunction with the continuous mapping theorem and a Slutsky argument leads to the desired consistency result. Details are provided in the Appendix.

To derive the asymptotic normality of the HME, define

$$A_0(t) := \begin{cases} \rho[1 - \lim_{s\to\infty} s^{-\gamma}U(s)/(t^{-\gamma}U(t))], & \rho < 0, \\ 1 - \int_0^t s^{-\gamma}U(s)ds/(t^{1-\gamma}U(t)), & \rho = 0. \end{cases} \tag{5}$$

This function is used to bound deviations in (3) uniformly in x (see Appendix). Based on Lemma 1 and (14) (see Appendix), we can state the following theorem.

Theorem 2 (CLT for the HME) *Assume that $X_1, \ldots, X_n$ are i.i.d. random variables with common distribution F, satisfying (3). Then for any intermediate sequence $k(n) \to \infty$, with $k/n \to 0$, satisfying*

$$\lim_{n\to\infty} \sqrt{k}A_0(n/k) = \lambda, \tag{6}$$

we have, for $\beta > 1 - 1/(2\gamma)$

$$\sqrt{k}\left(H_{n,k}^{(\beta)} - \gamma\right) \xrightarrow{d} N\left(\lambda\mu_\beta, \sigma_\beta^2\right), \tag{7}$$

where

$$\mu_\beta := \mu_\beta(\gamma,\rho) = \frac{1+\gamma(\beta-1)}{1-\rho+\gamma(\beta-1)}, \quad \sigma_\beta^2 := \sigma_\beta^2(\gamma) = \frac{\gamma^2(1+\gamma(\beta-1))^2}{1+2\gamma(\beta-1)}.$$

Remark 1 Note that, due to $A_0(t) \sim A(t)$, (6) is equivalent to

$$\lim_{n\to\infty} \sqrt{k}A(n/k) = \lambda \quad \text{for} \quad \rho < 0. \tag{8}$$

Remark 2 Under the trivial assumption that for some fixed $x_0 > 0$, we have the exact equality $1 - F(x) = P(X > x) = cx^{-1/\gamma}$ for all $x > x_0$ and some $c, \gamma > 0$, we obtain asymptotically an $N(0, \sigma_\beta^2)$-distribution, i.e. no asymptotic bias. This simple case has been considered already in Henry (2009). Moreover, for $\beta = 1$, Theorem 2 coincides with previous results for Hill's estimator $H_{n,k}^{(1)}$ (see for instance Ferreira and de Vries (2004), Theorem 3.2.5).

Corollary 1 *Under the conditions of Theorem 2, the asymptotic mean-squared error of* $H_{n,k}^{(\beta)}$ *is given by*

$$\mathrm{AMSE}(\beta) = k^{-1}(\lambda^2 \mu_\beta^2 + \sigma_\beta^2).$$

Moreover,

(a) *if* $\rho = 0$, *then* $\mu_\beta \equiv 1$ *and* σ_β *is minimal for* $\beta = 1$*;*
(b) *if* $\rho < 0$, *then*

$$\mathrm{AMSE}(1) < \mathrm{AMSE}(\beta) \quad (\beta > 1)$$

and there exist some $\beta^* \in (1 - 1/(2\gamma), 1)$ *such that*

$$\mathrm{AMSE}(1) > \mathrm{AMSE}(\beta^*);$$

(c) *if* $\rho \to -\infty$, *then* $\mu_\beta = 0$ *and*

$$\mathrm{eff}(\beta, 1) = \frac{\mathrm{AMSE}(1)}{\mathrm{AMSE}(\beta)} < 1 \quad (\beta \neq 1, \quad \beta > 1 - 1/(2\gamma)).$$

Theorem 2 and Corollary 1 can be used to compare different HME for a fixed intermediate sequence $k(n)$. As in Theorem 2 in de Haan and Peng (1998), a comparison of the mean-squared error at optimal levels $k_0^{(\beta)}$ can be based on the following result:

Theorem 3 *Assume that* (3) *holds with* $\rho < 0$ *and denote by* $k_0^{(\beta)}$ *an intermediate sequence minimizing the asymptotic second moment of* $H_{n,k}^{(\beta)} - \gamma$, *which is given by* $A^2(n/k)\mu_\beta^2 + k^{-1}\sigma_\beta^2$.

If $\mu_\beta \neq 0$, *then*

$$\sqrt{k_0^{(\beta)}} \left(H_{n,k_0^{(\beta)}}^{(\beta)} - \gamma \right) \xrightarrow{d} N\left(\frac{\mathrm{sign}(A)}{\sqrt{-2\rho}} \sigma_\beta, \sigma_\beta^2 \right). \tag{9}$$

Moreover

$$k_0^{(\beta)}(n) := k_0^{(\beta)} \sim \frac{n}{s^{\leftarrow}\left(n^{-1}\, \tau_\beta\right)},$$

where $s^{\leftarrow}$ *is the inverse of* $s \in RV_{2\rho-1}$ *given by*

$$A(t)^2 \sim \int_t^\infty s(u)du \quad \text{and} \quad \tau_\beta := \tau_\beta(\gamma, \rho) = \frac{\sigma_\beta^2}{\gamma^2 \mu_\beta^2} = \frac{(1 - \rho + \gamma(\beta - 1))^2}{1 + 2\gamma(\beta - 1)}.$$

Corollary 2 *Assume that (3) holds with $\rho < 0$. Then we have*

$$\mathrm{eff}_0(\beta, 1) := \frac{\mathrm{AMSE}\left(H^{(\beta)}_{n,k_0^{(\beta)}}\right)}{\mathrm{AMSE}\left(H^{(1)}_{n,k_0^{(1)}}\right)} = \left((1-\rho)^{-2}\tau_\beta\right)^{\frac{1}{2\rho-1}} \frac{(1+\gamma(\beta-1))^2}{1+2\gamma(\beta-1)}. \quad (10)$$

Moreover, for any $(\gamma, \rho) \in (0, \infty) \times (-\infty, 0)$, there exists a $\beta^* \in (1 - 1/(2\gamma), 1)$ such that

$$\mathrm{eff}_0(\beta^*, 1) < 1.$$

Figure 1a illustrates the efficiency of $H^{(\beta)}_{n,k_0^{(\beta)}}$ with respect to $H^{(1)}_{n,k_0^{(1)}}$ for the Pareto distribution, as a function of γ. In particular, it shows a different behaviour for $\beta < 1$ and $\beta > 1$. Figure 1b shows the optimal value of β as a function of γ for different values of ρ, i.e. $\beta_{\mathrm{opt}}(\gamma, \rho) = \arg\min_\beta \mathrm{eff}_0(\beta, 1)$. The corresponding efficiency curves can be found in Fig. 1c. Note that the efficiency gain does not exceed 5 %.

In the special case, where $A(t) = Ct^\rho$, the following explicit formula for the AMSE can be obtained.

Corollary 3 *Assume $A(t) = Ct^\rho$, then $k_0^{(\beta)}$ from Theorem 3 is given by*

$$k_0^{(\beta)} := k_0^{(\beta)}(n) = \left[\left(-\frac{1}{2}\rho^{-1}C^{-2}\gamma^2\tau_\beta\right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}\right],$$

where $[x]$ means the integer part of x. Moreover, (9) holds with $\mathrm{sign}(A) = \mathrm{sign}(C)$. *Therefore,*

$$\mathrm{AMSE}\left(H^{(\beta)}_{n,k_0^{(\beta)}}\right) = \sigma_\beta^2\left(1 - \frac{1}{2}\rho^{-1}\right)\cdot\left(-2\rho\, C^2\gamma^{-2}\tau_\beta^{-1}\right)^{1/(1-2\rho)} n^{2\rho/(1-2\rho)}.$$

Example 1 For the Fréchet distribution, we have $A(t) = \frac{1}{2}\gamma t^{-1}$. This leads to

$$\mathrm{AMSE}\left(H^{(\beta)}_{n,k_0^{(\beta)}}\right) = \frac{3}{2}\sigma_\beta^2\left[2^{-1}\tau_\beta^{-1}(\gamma, -1)\right]^{1/3} n^{-2/3}.$$

Example 2 For $Y = |X|$ with X standard Cauchy distributed, we have

$$1 - F_Y(x) = \frac{2}{\pi}x^{-1} - \frac{2}{3\pi}x^{-3} + o(x^{-3})\ (x \to \infty).$$

Therefore, $A(t) = (\pi^2/6)t^{-2}$ and

$$\mathrm{AMSE}\left(H^{(\beta)}_{n,k_0^{(\beta)}}\right) = \frac{5}{4}\sigma_\beta^2\left[(\pi/3)^2\tau_\beta^{-1}(1, -2)\right]^{1/5} n^{-4/5}.$$

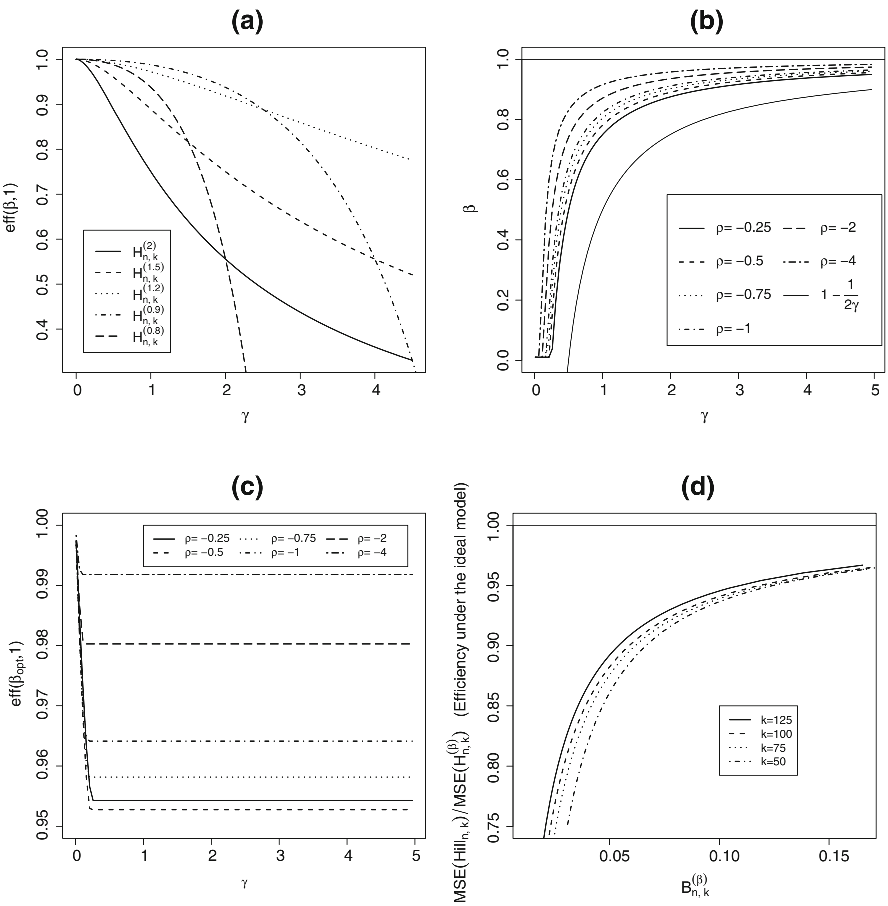

Fig. 1 **a** Asymptotic efficiency eff$(\beta, 1)$ for $\beta = 0.8, 0.9, 1.2, 1.5, 2$, **b** β_{opt} as a function of γ for different ρ-levels, **c** eff$_0(\beta_{\text{opt}}, 1)$ as a function of γ for different ρ-levels. **d** Simulated eff$(\beta, 1)$ for the Fréchet distribution with $\gamma = 0.5$ as a function of $B_{n,k}^{(\beta)}$ with $n = 200$ and $k = 50, 75, 100, 125$. The results are based on 1000 samples

Example 3 For the Burr(η, τ, λ)-distribution (type XII) defined as
$F(x) = 1 - (\eta/(\eta + x^\tau))^\lambda$ $(x > 0)$ with $\eta, \tau, \lambda > 0$, we obtain $\gamma = (\lambda\tau)^{-1}$ and $A(t) = (\lambda\tau)^{-1}t^{-1/\lambda}$. This leads to

$$\text{AMSE}\left(H_{n,k_0^{(\beta)}}^{(\beta)}\right) = \sigma_\beta^2\left(1 + \frac{1}{2}\lambda\right)\left(2\lambda^{-1}\tau_\beta^{-1}((\lambda\tau)^{-1}, -\lambda^{-1})\right)^{\lambda/(\lambda+2)} n^{-2/(2+\lambda)}.$$

Remark 3 (Comments on scale and quantile estimation) The estimation of the tail index using the generalized Pareto Distribution (GPD) is motivated by

$$\lim_{u \to x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\gamma,\sigma(u)}(x)| = 0,$$

where x_F is the right endpoint of the distribution, $F_u(x) = P(X - u \leq x \mid X > u)$, and $G_{\gamma,\sigma}(x) = 1 - \left(1 + \gamma x \sigma^{-1}\right)^{-1/\gamma}$ is the GPD (see Balkema and de Haan 1974; Pickands 1975). Here, γ and σ can be interpreted as shape and scale parameters, respectively. It turns out however that, for the purpose of estimating high quantiles, there is no need for estimating σ separately. The reason is that the general first-order condition for the tail quantile function U implies that

$$U(tx) \approx U(t) + a(t)\frac{x^{\gamma} - 1}{\gamma}.$$

High quantiles $U(1/p) = F^{\leftarrow}(1 - p)$ may therefore be estimated by

$$\hat{U}(1/p) = \hat{U}(n/k) + \hat{a}(n/k)\frac{(k/np)^{\hat{\gamma}} - 1}{\hat{\gamma}}$$

where k is an intermediate sequence, $p = p_n = o(1)$, and $\hat{a}(n/k)$ is a scale estimate. For $\gamma > 0, a(t) = \gamma U(t)$ is a valid choice for the auxiliary function. A consistent scale estimator is then obtained for instance by $\hat{a}(n/k) = \hat{\gamma}\hat{U}(n/k)$ (cf. Ferreira and de Vries 2004). Given an asymptotically normal estimator $\hat{\gamma}$ of the tail index, this directly leads to consistent estimation of high quantiles $x_{p_n} := U(1/p_n)$ (where $np_n = o(k)$), without the need for a separate scale estimator. For instance, we may set

$$\hat{x}_{p_n} := X_{n-k,n}\left(\frac{k}{np_n}\right)^{\hat{\gamma}}$$

(Weissman 1978). Under the assumptions of Theorem 2 ($\rho < 0$), $np_n = o(k)$, $\log(np_n) = o(\sqrt{k})$, and

$$\sqrt{k}\left(\hat{\gamma} - \gamma\right) \xrightarrow{d} \zeta,$$

where ζ is a Gaussian random variable whose variance does not depend on ρ, the (high) quantile estimator $\hat{x}_{p_n}$ is asymptotically normal in the sense that

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1\right) \xrightarrow{d} \zeta$$

(see Theorem 4.3.8 and Corollary 4.3.9 in Ferreira and de Vries 2004). The only requirement on the tail index estimator $\hat{\gamma}$ is that, under the second-order condition, the limit ζ is normal with a variance that does not depend on ρ. This is true for Hill's estimator as well as for the HME discussed here. In particular, defining $\hat{x}_{p_n}$ by

$$\hat{x}_{p_n}^{(\beta)} := X_{n-k,n}\left(\frac{k}{np_n}\right)^{H_{n,k}^{(\beta)}}$$

we obtain

$$\frac{\sqrt{k}}{\log(d_n)}\left(\frac{\hat{x}_{p_n}^{(\beta)}}{x_{p_n}}-1\right)\xrightarrow{d} N(\lambda\mu_\beta,\sigma_\beta^2).$$

Similarly, estimates of tail probabilities (return levels) can be obtained applying Theorem 4.4.7 in Ferreira and de Vries (2004).

Remark 4 The (high) quantile estimator $\hat{x}_{p_n}$ includes $\hat{\gamma}$ in the exponent of $(kn^{-1}p_n^{-1})$. This means that even a small bias in $\hat{\gamma}$ may cause large errors (see also Beirlant et al. 2004). Thus, in most applications high quantile estimates based on a robust method such as the HME with $\beta > 1$ are likely to perform better than nonrobust estimates, because usually the Pareto tail is reached only asymptotically. The same comments apply to return levels of extreme events.

Remark 5 Although many tail index estimators exist in the literature, we focus on a comparison with Hill's method. There are several reasons for doing so. First of all, Hill's estimator is usually considered a benchmark, because it has the smallest asymptotic variance. The second reason is that the asymptotic normality of the HME was established under the same second-order conditions as for the Hill estimator. Due to having the same normalizing function A and the same second-order parameter, the relative asymptotic efficiency (defined by the ratio of asymptotic mean-squared errors) can easily be derived. As pointed out by de Haan and Ferreira (2004, p. 116–118), a completely general comparison of tail index estimators is difficult due to differing second-order conditions. The essential reason is that the asymptotic bias depends on the auxiliary function and is therefore not directly comparable. Moreover, even for estimators whose asymptotic normality has been established under the same second-order condition, it does not seem to be possible to find a uniformly best tail index estimator (see e.g. de Haan and Peng 1998).

Remark 6 (Comments on confidence intervals) Confidence intervals based on Theorems 2 and 3 can be obtained in an analogous manner as for Hill's estimator (see e.g. Ferreira and de Vries 2004; Cheng and Peng 2001; Lu and Peng 2002; Qi 2008; Worms and Worms 2011, and references therein). Essentially, two different approaches can be distinguished: (a) the sequence $k = k(n)$ is such that the asymptotic mean-squared error is minimized or (b) $k(n)$ is such that the variance dominates the mean-squared error asymptotically. In case (a), a bias correction is needed so that the second-order parameter ρ as well as the function A (or equivalently the parameter λ) has to be estimated. In case (b), no asymptotic bias occurs and the usual bootstrap techniques can be adopted.

4 Robustness of the harmonic moment tail index estimator

After considering the efficiency of $H_{n,k}^{(\beta)}$ in the previous section, we investigate its robustness properties. The definition of an influence function (Hampel 1968; Hampel

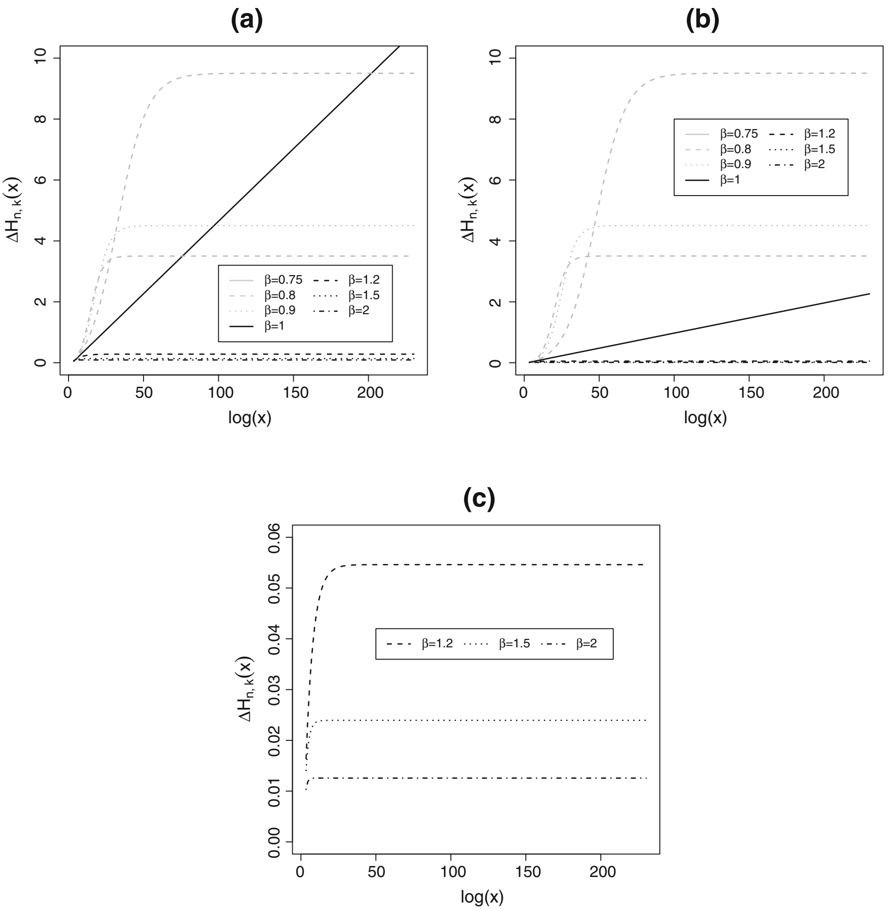

Fig. 2 Simulated mean of $\Delta H_{n,k}^{(\beta)}(x)$ (on log-scale) for different values of β and sample size $n = 1000$. The results are based on 1000 samples from a Pareto distribution with $\gamma = 0.5$. The number of order statistics included in the estimation was $k = 20$ (**a**) and $k = 100$ (**b**, **c**)

et al. 1986) in the context of consistent tail index estimation poses considerable difficulties due to the asymptotically vanishing portion k/n of data used in the functional. We therefore consider instead

$$\Delta H_{n,k}^{(\beta)}(x) := H_{n,k}^{(\beta)}(X_1, \ldots, X_{n-1}, x) - H_{n-1,k-1}^{(\beta)}(X_1, \ldots, X_{n-1})$$

and

$$B_{n,k}^{(\beta)} := \lim_{x \to \infty} \Delta H_{n,k}^{(\beta)}(x).$$

Hence, for fixed n and k, $B_{n,k}^{(\beta)}$ measures the worst effect of one arbitrarily large contamination on the HME. The asymptotic value of $B_{n,k}^{(\beta)}$ is given by

$$B_{n,k}^{(\beta)} \xrightarrow{P} \begin{cases} (1-\beta)^{-1}-\gamma, & \text{for } 1-\gamma^{-1}<\beta<1, \\ \infty, & \text{for } \beta=1, \\ 0, & \text{for } \beta>1. \end{cases}$$

provided that $n, k \to \infty$, but $k/n \to 0$ (see Appendix).

Figure 2 displays $\Delta H_{n,k}^{(\beta)}(x)$, for $n = 1000$ and $k = 20, 100$ using the Pareto distribution with $\gamma = 0.5$. The outlier x ranges from 5 to 10^{100} times the expected maximum of a sample. Figure 2a illustrates the lack of robustness of the Hill estimator (see $B_{n,k}^{(1)}$). In contrast, $B_{n,k}^{(\beta)}$ ($\beta \neq 1$) is bounded asymptotically.

Figure 3a shows (for $\rho = -0.25, -0.5, -0.75, -1, -2, -4$) values of $\beta = \beta(\gamma)$ as a function of γ where $\text{eff}(\beta, 1) = 0.9$. Figure 3b is the same but with $\text{eff}(\beta, 1) = 0.8$. Note that for β above the corresponding curve the efficiency loss is higher. Figure 1d illustrates the trade-off between efficiency and robustness in the class of HME.

Remark 7 In summary, under the ideal assumption of an exact Pareto distribution, Hill's estimator is more efficient than $H_{n,k}^{(\beta)}$ for all $\beta > 1$, but the efficiency of $H_{n,k}^{(\beta)}$ increases as β approaches 1 from above. On the other hand, for Hill's estimator $B_{n,k}^{(1)}$ diverges to infinity, whereas $B_{n,k}^{(\beta)}$ converges to zero for $\beta > 1$, thus indicating much smaller sensitivity to contamination. In practice, one may use for instance a range of $\beta \in [1.2, 1.5]$, since the efficiency loss remains acceptable (less than 20 %) at least for small values of γ, while maintaining a relatively high degree of robustness. It should be noted however that the efficiency loss depends on the unknown parameter γ (see Fig. 1). A possible pragmatic approach may therefore be as follows. Given a desired efficiency value e_0 solve $\text{AEFF}(\gamma, \beta) = e_0$ to obtain $\beta = \beta(\gamma)$ as a function of γ.

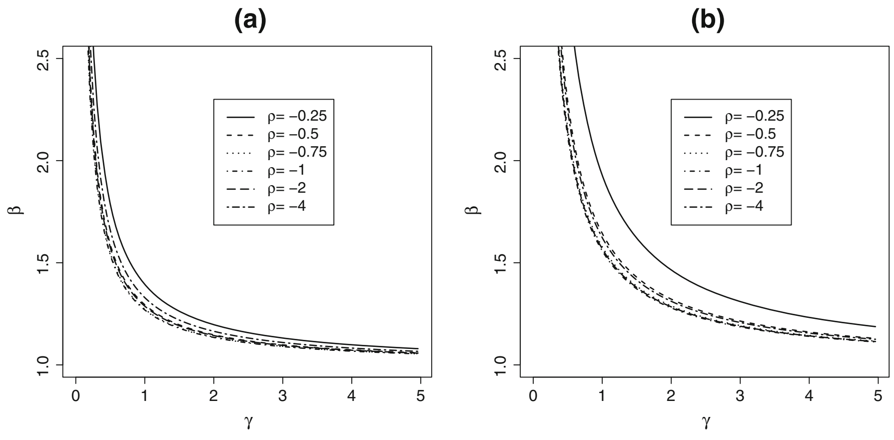

Fig. 3 **a** (for $\rho = -0.25, -0.5, -0.75, -1, -2, -4$) values of $\beta = \beta(\gamma)$ as a function of γ where $\text{eff}(\beta, 1) = 0.9$. **b** same as in a, but with $\text{eff}(\beta, 1) = 0.8$

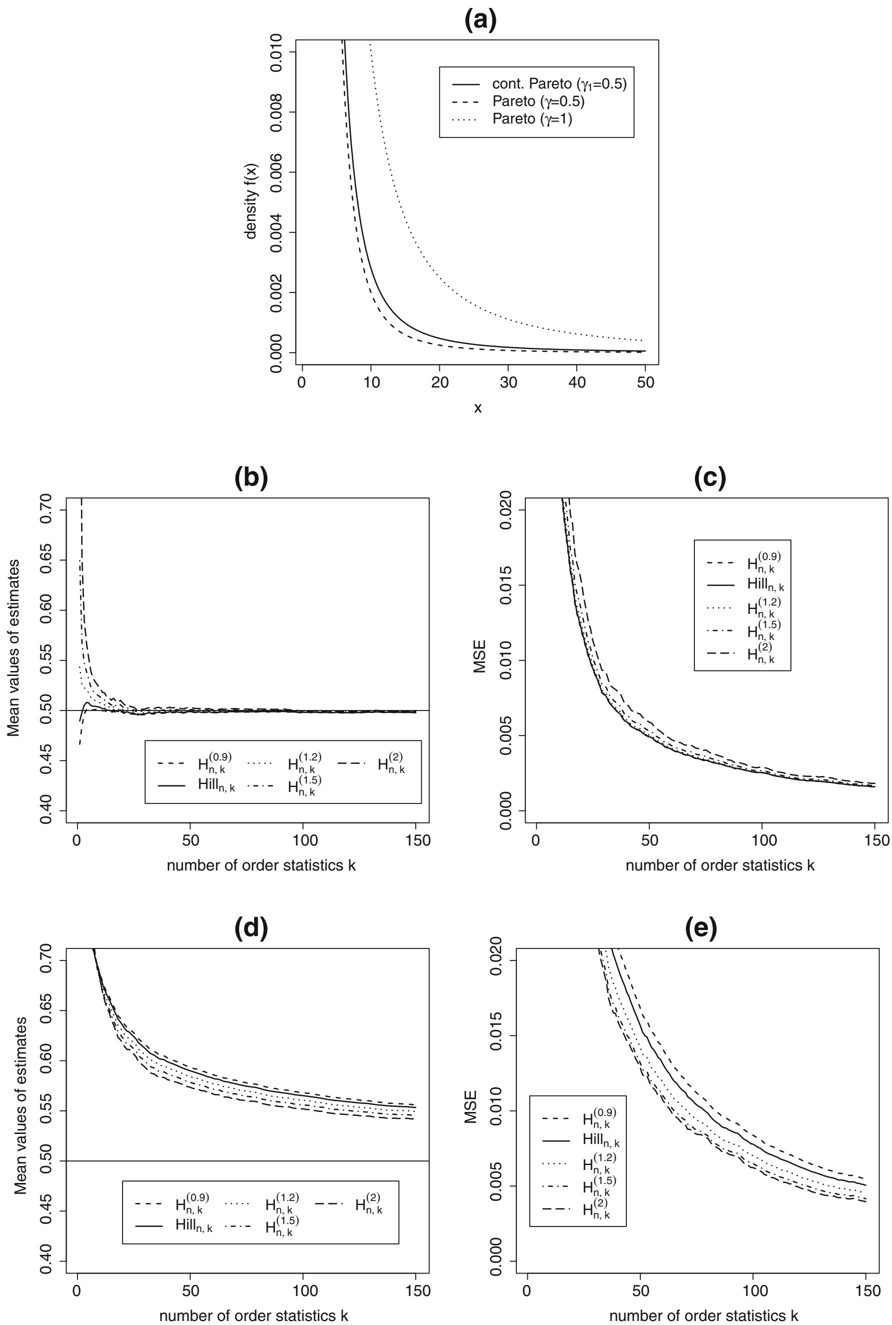

Fig. 4 **a** Densities of $F_{\text{par}}(x; 0.5)$, $F_{\text{par}}(x; 1)$ and $F_{\text{par}}(x; 0.5, 1, 0.1)$; **b** and **c** averaged Hill-plot and MSE based on samples from $F_{\text{par}}(x; 0.5)$ with sample size $n = 200$; **d** and **e** averaged Hill-plot and MSE based on samples from $F_{\text{par}}(x; 0.5)$ with sample size $n = 200$ contaminated by samples from $F_{\text{par}}(x; 1)$ with sample size $\tilde{n} = 20$. The results in **b**, **c**, **d** and **e** are based on 1000 samples

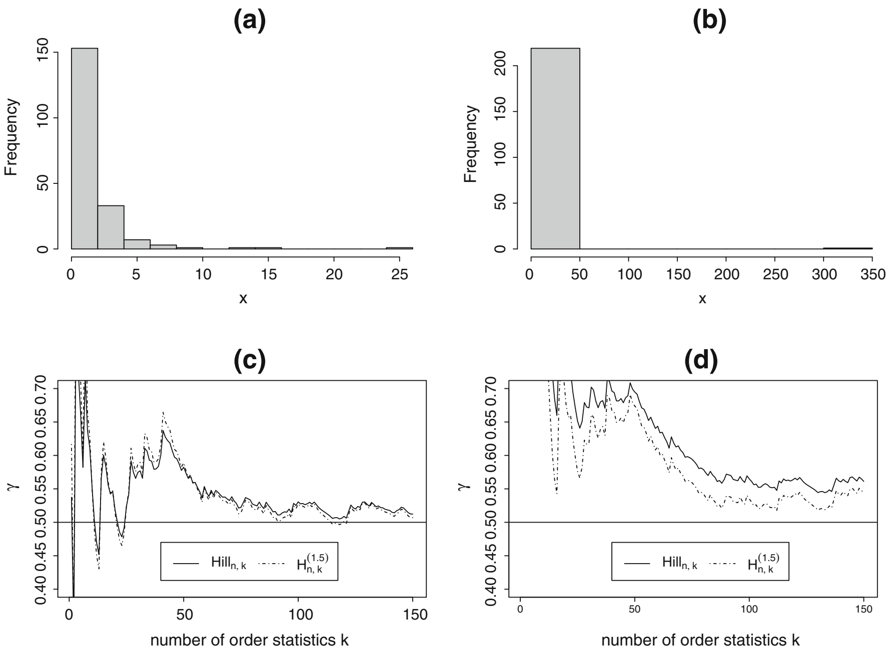

Fig. 5 One sample from a Pareto distribution ($\gamma = 0.5$) with sample size $n = 200$ (*left*) and the same sample contaminated by 20 observations from Pareto with $\gamma = 2$ (*right*). Histograms are shown in **a** and **b**, Hill-plots in **c** and **d**

Compute an initial consistent estimate $\hat{\gamma}_0$. Then choose $\hat{\beta} := \beta(\hat{\gamma}_0)$ and calculate the corresponding estimate $\hat{\gamma} = H^{(\hat{\beta})}_{k,n}$. Note that, as illustrated in Fig. 3, the second-order parameter ρ does not have much influence on the value of $\beta(\gamma)$.

Remark 8 It may be useful to point out that robustness and modelling of extreme values are not necessarily contradictory concepts. Though it may be difficult to identify outliers formally when estimating the tail index, it can generally be recommended in practice to use several different robust and nonrobust estimators. In case of a strong discrepancy between the computed values, one may suspect the presence of outliers with respect to a Pareto tail.

5 Simulation study

We consider the following mixture of Pareto distributions,

$$\begin{aligned} F_{\text{par}}(x; \gamma_1, \gamma_2, \varepsilon) &= (1-\varepsilon) F_{\text{par}}(x; \gamma_1) + \varepsilon F_{\text{par}}(x; \gamma_2) \\ &= 1 - (1-\varepsilon) x^{-1/\gamma_1} - \varepsilon x^{-1/\gamma_2}, \end{aligned} \tag{11}$$

where $\gamma_2, \gamma_1 > 0$, and $0 \leq \varepsilon < 0.5$ is the fraction of contamination. Note that for $\varepsilon = 0$, $H^{(\beta)}_{n,k}$ is asymptotically unbiased. Therefore, for $\varepsilon \neq 0$, the effect of contamination becomes immediately apparent. Moreover, if $\varepsilon > 0$ and $\gamma_1 > \gamma_2$, then

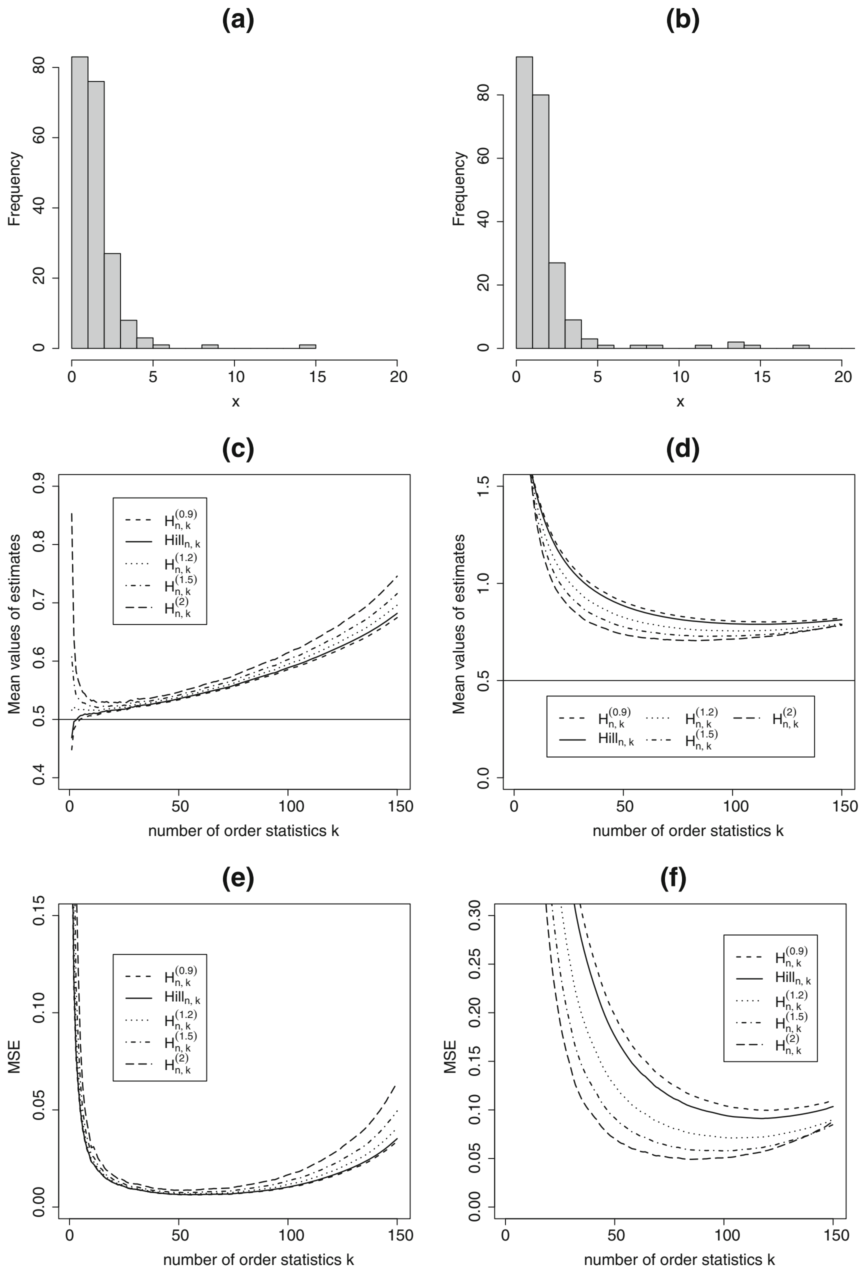

Fig. 6 Simulated results for Fréchet samples ($\gamma = 1$) with sample size $n = 200$ (*left*) and Fréchet samples ($\gamma = 1$) where 20 out of $\tilde{n} = 220$ observations were replaced by Fréchet variables with $\gamma = 2$ (*right*); **a** and **b** histograms of the samples; **c** and **d** averaged Hill-plots based on $N = 1000$ simulated samples; **e** and **f** MSE based on $N = 1000$ simulated samples

$F_{\mathrm{par}}(x; \gamma_1, \gamma_2, \varepsilon)$ satisfies (3) with $\gamma = \gamma_1$ and $\rho = (\gamma_2 - \gamma_1)/\gamma_2 < 0$. On the other hand, if $\gamma_1 < \gamma_2$ (and $\varepsilon > 0$), (11) corresponds to a Pareto distribution contaminated by a longer tailed distribution.

Figure 4 illustrates the results for $\varepsilon > 0$ and $\gamma_1 < \gamma_2$. Figure 5 shows that using $H_{n,k}^{(\beta)}$ with $\beta > 1$ can lead to a substantial bias reduction, even if the contamination (of the Pareto distribution) is hardly visible.

Finally, we consider 1000 Fréchet distributed samples with $\gamma = 0.5$ and sample size $n = 200$. These were contaminated by Fréchet distributed samples with $\gamma = 2$ and sample size $\tilde{n} = 20$. Hence, 10 % of the resulting sample originates from a distribution with much heavier tails, causing observations far on the right. Nevertheless, it is unlikely that this kind of contamination is noticed by looking at the histogram (see Fig. 6a,b). As expected, Hill's estimator as well as $H_{n,k}^{(\beta)}$ with $\beta < 1$ turn out to be more sensitive to this type of contaminations than $H_{n,k}^{(\beta)}$ with $\beta > 1$ (see Fig. 6d,f).

6 Final remarks

In this paper we considered HME, $\hat{\gamma} = H_{n,k}^{(\beta)}$, characterized by a tuning parameter β. The asymptotic distribution of $\hat{\gamma}$ was derived under general conditions. For $\beta > 1$, $\hat{\gamma}$ is robust against large outliers. The HME class thus provides a simple approach to robust tail index estimation, with $\hat{\gamma}$ being an explicit function of order statistics. Moreover, for $\beta < 1$ it is possible to obtain a smaller mean-squared error than for Hill's estimator ($\beta = 1$). Further optimality properties under robustness constraints and more general classes of estimators including more general transformations are some of the questions to be considered in future research.

Appendix

Proof of Theorem 1 See supplementary material. □

Proof of Theorem 2 Two auxiliary lemmas, taken from de Haan and Ferreira (2006), are required for the proof of asymptotic normality. The following Lemma allows to bound the deviations in (3) uniformly in x. □

Lemma 1 (de Haan and Ferreira (2006, p. 48, Theorem 2.3.9)) *Suppose the second-order condition* (3) *holds, then for any* $\varepsilon, \delta > 0$, *there exists* $t_0 = t_0(\varepsilon, \delta) > 1$ *such that for all* $t, tx > t_0$,

$$\left| \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A_0(t)} - x^\gamma \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon x^{\gamma+\rho} \max(x^\delta, x^{-\delta}), \tag{12}$$

with

$$A_0(t) := \begin{cases} \rho[1 - \lim_{s\to\infty} s^{-\gamma} U(s)/(t^{-\gamma} U(t))], & \rho < 0, \\ 1 - \int_0^t s^{-\gamma} U(s) \mathrm{d}s/(t^{1-\gamma} U(t)), & \rho = 0. \end{cases} \tag{13}$$

Theorem 5.1.4 in de Haan and Ferreira (2006) provides an expansion of the empirical tail function $1 - F_n$.

Lemma 2 (de Haan and Ferreira (2006, Theorem 5.1.4)) *let $X_1, X_2, \ldots$ be i.i.d. with distribution F as in (1). Moreover, suppose that the function U satisfies (3) and let $\varepsilon > 0$, $k(n) \to \infty$ such that $\sqrt{k}A_0(n/k)$ is bounded as $n \to \infty$. Then the underlying sample space can be enlarged to include a sequence of Brownian motions W_n such that for all $x_0 > 0$,*

$$\sup_{x \geq x_0} x^{(1/2-\varepsilon)/\gamma} \left| \sqrt{k} \left\{ \frac{n}{k} [1 - F_n(xU(n/k))] - x^{-1/\gamma} \right\} \right. \\ \left. - W_n(x^{-1/\gamma}) - \sqrt{k}A_0(n/k) x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\gamma\rho} \right| \xrightarrow{P} 0, \tag{14}$$

as $n \to \infty$.

The essential approximation will be

$$\sqrt{k}\left(\hat{x}_{H,k}^{(\beta)} - 1/(1 + \gamma(\beta - 1))\right) = \mathrm{I} + \mathrm{II} + \mathrm{III}, \tag{15}$$

where

$$\hat{x}_{H,k}^{(\beta)} = \frac{1}{k} \sum_{i=1}^{k} \left(\frac{X_{n-k,n}}{X_{n-i+1,n}} \right)^{\beta-1} = X_{n-k,n}^{\beta-1} \frac{1}{k} \sum_{i=1}^{k} \frac{1}{X_{n-i+1,n}^{\beta-1}}$$

and

$$\mathrm{I} = \sqrt{k}(1 - \beta) \left[\left(\frac{X_{n-k,n}}{U(n/k)} \right)^{\beta-1} \int_{X_{n-k,n}/U(n/k)}^{1} \frac{n}{k} (1 - F_n(tU(n/k))) \frac{\mathrm{d}t}{t^\beta} \right],$$
$$\mathrm{II} = \sqrt{k}(1 - \beta) \left[\left(\frac{X_{n-k,n}}{U(n/k)} \right)^{\beta-1} \int_{1}^{\infty} \left(\frac{n}{k} (1 - F_n(tU(n/k))) - t^{-1/\gamma} \right) \frac{\mathrm{d}t}{t^\beta} \right],$$
$$\mathrm{III} = \sqrt{k} \frac{\gamma(1 - \beta)}{1 + \gamma(\beta - 1)} \left(\left(\frac{X_{n-k,n}}{U(n/k)} \right)^{\beta-1} - 1 \right).$$

To obtain (15), we rewrite $\hat{x}_{H,k}^{(\beta)}$ as an integral with respect to the empirical distribution function F_n

$$\hat{x}_{H,k}^{(\beta)} = X_{n-k,n}^{\beta-1} \int_{X_{n-k,n}}^{\infty} \frac{n}{k} \frac{1}{s^{\beta-1}} \mathrm{d}F_n(s). \tag{16}$$

For $\beta \neq 1$, we obtain by integration by parts

$$\int_{t}^{\infty} (1 - F(s)) \frac{\mathrm{d}s}{s^\beta} = -\frac{1}{1 - \beta} t^{1-\beta} (1 - F(t)) + \frac{1}{1 - \beta} \int_{t}^{\infty} s^{1-\beta} \mathrm{d}F(s),$$

and therefore

$$\int_t^\infty s^{1-\beta} \mathrm{d}F(s) = t^{1-\beta}(1-F(t)) + (1-\beta)\int_t^\infty (1-F(s))\frac{\mathrm{d}s}{s^\beta}.$$

Thus, setting $t = X_{n-k,n}$ and $F = F_n$, yields

$$\begin{aligned}\hat{x}_{H,k}^{(\beta)} &= \frac{n}{k} X_{n-k,n}^{\beta-1}\left[X_{n-k,n}^{1-\beta}\left(1-F_n(X_{n-k,n})\right) + (1-\beta)\int_{X_{n-k,n}}^\infty (1-F_n(s))\frac{\mathrm{d}s}{s^\beta}\right] \\ &= 1 + (1-\beta)\left(\frac{X_{n-k,n}}{U(n/k)}\right)^{\beta-1}\int_{X_{n-k,n}/U(n/k)}^\infty \frac{n}{k}\left(1-F_n(tU(n/k))\right)\frac{\mathrm{d}t}{t^\beta},\end{aligned}$$

where in the last step we have substituted s by $tU(n/k)$ inside the integral.
Together with

$$\int_1^\infty t^{-1/\gamma-\beta}\mathrm{d}t = \frac{1}{\beta-1+\gamma^{-1}} = \frac{\gamma}{1+\gamma(\beta-1)} \quad \text{for} \quad \beta > 1-\gamma^{-1},$$

we obtain (15).
First consider I. The asymptotic approximation in (14) yields

$$\begin{aligned}&\int_{X_{n-k,n}/U(n/k)}^1 \sqrt{k}\left(\frac{n}{k}\{1-F_n(tU(n/k))\} - t^{-1/\gamma}\right)\frac{\mathrm{d}t}{t^\beta} \\ &\le \left(1-\frac{X_{n-k,n}}{U(n/k)}\right)\sup_{t\in B}\left|\frac{1}{t^\beta}\left(C_n(\gamma,\rho,k)+o_p(1)\right)\right|,\end{aligned}$$

where $B := \left[\frac{X_{n-k,n}}{U(n/k)}, 1\right]$ and

$$C_n(\gamma,\rho,k) := W_n(t^{-1/\gamma}) - \sqrt{k}A_0(n/k)t^{-\gamma^{-1}}\frac{t^{\rho/\gamma}-1}{\gamma\rho}.$$

Thus, due to

$$\sqrt{k}\left(\left(\frac{X_{n-k,n}}{U(n/k)}\right)^{-1/\gamma-\beta+1} - 1\right) - \gamma\left(-1/\gamma-\beta+1\right)W_n(1) \xrightarrow{P} 0,$$

we obtain

$$\begin{aligned}\sqrt{k}\int_{X_{n-k,n}/U(n/k)}^1 t^{-1/\gamma-\beta}\mathrm{d}t &= -\sqrt{k}\frac{1}{-\gamma^{-1}-\beta+1}\left[\left(\frac{X_{n-k,n}}{U(n/k)}\right)^{-1/\gamma-\beta+1} - 1\right] \\ &\stackrel{d}{=} -\gamma W_n(1) + o_p(1).\end{aligned}$$

This yields

$$(1-\beta)\sqrt{k}\int_{X_{n-k,n}/U(n/k)}^{1}\frac{n}{k}\{1-F_n(tU(n/k))\}\frac{\mathrm{d}t}{t^{\beta}}+(1-\beta)\gamma W_n(1)\xrightarrow{P}0.$$

Moreover, since $\frac{X_{n-k,n}}{U(n/k)}\xrightarrow{P}1$, we can use Slutsky's theorem to show that

$$(1-\beta)\sqrt{k}\left(\frac{X_{n-k,n}}{U(n/k)}\right)^{\beta-1}\int_{X_{n-k,n}/U(n/k)}^{1}\frac{n}{k}\left(1-F_n(tU(n/k))\right)\frac{\mathrm{d}t}{t^{\beta}}$$
$$+(1-\beta)\gamma W_n(1)\xrightarrow{P}0.$$

Next, consider II. Uniform convergence in (14) allows to write

$$\begin{aligned}\mathrm{II}&\stackrel{d}{=}(1-\beta)\left(\left(\frac{X_{n-k,n}}{U(n/k)}\right)^{\beta-1}\int_1^{\infty}W_n(t^{-1/\gamma})\frac{\mathrm{d}t}{t^{\beta}}\right.\\&\quad\left.+\sqrt{k}A_0(n/k)\int_1^{\infty}t^{-1/\gamma}\frac{t^{\gamma/\rho}-1}{\rho\gamma}\frac{\mathrm{d}t}{t^{\beta}}+o_p(1)\right)\\&\stackrel{d}{=}(1-\beta)\int_1^{\infty}W_n(t^{-1/\gamma})\frac{\mathrm{d}t}{t^{\beta}}+(1-\beta)\sqrt{k}A_0(n/k)\int_1^{\infty}t^{-1/\gamma}\frac{t^{\gamma/\rho}-1}{\rho\gamma}\frac{\mathrm{d}t}{t^{\beta}},\end{aligned}$$

where in the last step we again make use of Slutsky's theorem. Note,

$$\int_1^{\infty}W_n(t^{-1/\gamma})\frac{\mathrm{d}t}{t^{\beta}}=\gamma\int_0^1 W_n(s)s^{-\gamma-1+\gamma\beta}\mathrm{d}s$$

and

$$\int_1^{\infty}t^{-1/\gamma}\frac{t^{\gamma/\rho-1}}{\rho\gamma}\frac{\mathrm{d}t}{t^{\beta}}=\frac{1}{(1-\rho+\gamma(\beta-1))(1+\gamma(\beta-1))},$$

where we implicitly assume that $\beta>1-1/\gamma$. Under the additional assumption $\lim_{n\to\infty}\sqrt{k}A_0(n/k)=\lambda$, we end up with

$$\mathrm{II}\stackrel{d}{=}(1-\beta)\gamma\int_0^1 W_n(s)s^{-\gamma-1+\gamma\beta}\mathrm{d}s+\frac{\lambda(1-\beta)}{(1-\rho+\gamma(\beta-1))(1+\gamma(\beta-1))}.$$

Considering the remaining term, we obtain by the delta method

$$\mathrm{III}=\frac{\gamma(1-\beta)}{1+\gamma(\beta-1)}\sqrt{k}\left(\left(\frac{X_{n-k,n}}{U(n/k)}\right)^{\beta-1}-1\right)=\frac{\gamma^2(1-\beta)^2}{\gamma(1-\beta)-1}W_n(1).$$

Altogether, we get

$$
\begin{aligned}
&\sqrt{k}\left(\hat{x}_{H,k}^{(\beta)} - 1/(1+\gamma(\beta-1))\right) \\
&\overset{d}{\to} -\gamma(1-\beta)W_n(1) + \gamma(1-\beta)\int_0^1 W_n(s)s^{-\gamma-1+\gamma\beta}\mathrm{d}s \\
&\quad + \frac{\lambda(1-\beta)}{(1-\rho+\gamma(\beta-1))(1+\gamma(\beta-1))} + \frac{\gamma^2(1-\beta)^2}{\gamma(1-\beta)-1}W_n(1) \\
&= \frac{\gamma(1-\beta)}{\gamma(1-\beta)-1}W_n(1) + (1-\beta)\gamma\int_0^1 W_n(s)s^{-\gamma-1+\gamma\beta}\mathrm{d}s \\
&\quad + \frac{\lambda(1-\beta)}{(1-\rho+\gamma(\beta-1))(1+\gamma(\beta-1))}.
\end{aligned}
$$

Now, since $\gamma(1-\beta)/(\gamma(1-\beta)-1)W_n(1) + (1-\beta)\gamma\int_0^1 W_n(s)s^{-\gamma-1+\gamma\beta}\mathrm{d}s$ is a linear combination of Gaussian random variables, we obtain the desired central limit theorem after computing its variance.

$$
\begin{aligned}
&E\left\{\left[(\gamma(1-\beta)-1)^{-1}W(1) + \int_0^1 W(s)s^{-\gamma-1+\gamma\beta}\mathrm{d}s\right]^2\right\} \\
&= (\gamma(1-\beta)-1)^{-2}E[W^2(1)] + 2(\gamma(1-\beta)-1)^{-1}\int_0^1 E[W(1)W(s)]s^{-\gamma-1+\gamma\beta}\mathrm{d}s \\
&\quad + \int_0^1\int_0^1 E[W(t)W(s)]t^{-\gamma-1+\gamma\beta}\mathrm{d}t\; s^{-\gamma-1+\gamma\beta}\mathrm{d}s \\
&=: S_1 + S_2 + S_3.
\end{aligned}
$$

Obviously, $S_1 = (\gamma(1-\beta)-1)^{-2}$. Moreover, due to $E[W(s)W(t)] = \min(s,t)$, we obtain

$$
S_2 = 2(\gamma(1-\beta)-1)^{-1}\int_0^1 s^{-\gamma+\gamma\beta}\mathrm{d}s = -2(\gamma(1-\beta)-1)^{-2}.
$$

Furthermore, we have

$$
S_3 = 2\int_0^1\int_0^s t^{\gamma+\gamma\beta}dt\; s^{\gamma-1+\gamma\beta}\mathrm{d}s = 2(\gamma(1-\beta)-1)^{-1}(2\gamma(1-\beta)-1)^{-1},
$$

provided $\beta > 1-(2\gamma)^{-1}$. Note that for $\beta \le 1-(2\gamma)^{-1}$ S_3 is not defined.

Finally, we obtain

$$\begin{aligned}
&\mathrm{Var}\left(\frac{\gamma(1-\beta)}{\gamma(1-\beta)-1}W_n(1)+(1-\beta)\gamma\int_0^1 W_n(s)s^{-\gamma-1+\gamma\beta}\mathrm{d}s\right)\\
&=\gamma^2(1-\beta)^2\Big[(\gamma(1-\beta)-1)^{-2}-2(\gamma(1-\beta)-1)^{-2}\\
&\quad+2(\gamma(1-\beta)-1)^{-1}(2\gamma(1-\beta)-1)^{-1}\Big]\\
&\quad=\frac{\frac{1}{\gamma}(1-\beta)^2}{\left(1-\gamma^{-1}-\beta\right)^2\left(2-2\beta-\gamma^{-1}\right)}.
\end{aligned}$$

Now, we are ready to state the central limit theorem for $\hat{x}_{H,k}^{(\beta)}$. Suppose k is intermediate, $\beta > 1-1/(2\gamma)$ and $\lim_{n\to\infty}\sqrt{k}A_0(n/k)=\lambda$. Then

$$\sqrt{k}\left(\hat{x}_{H,k}^{(\beta)}-1/(1+\gamma(\beta-1))\right)\xrightarrow{d}N\left(\lambda\tilde{\mu}_\beta,\tilde{\sigma}_\beta^2\right), \tag{17}$$

where

$$\tilde{\mu}_\beta(\gamma,\rho):=\tilde{\mu}_\beta=(1-\beta)/\left[(1-\rho+\gamma(\beta-1))(1+\gamma(\beta-1))\right]$$

and

$$\tilde{\sigma}_\beta^2(\gamma):=\tilde{\sigma}_\beta^2=\frac{\gamma^2(1-\beta)^2}{(1+\gamma(\beta-1))^2(1+2\gamma(\beta-1))}.$$

We can construct an estimator for γ using the map $g(\theta)=\frac{1}{\beta-1}\left(\frac{1}{\theta}-1\right)$, since $g(1/(1+\gamma(\beta-1)))=\gamma$. Applying the delta method yields

$$\sqrt{k}\left(\frac{1}{\beta-1}\left(\frac{1}{\hat{x}_{H,k}^{(\beta)}}-1\right)-\gamma\right)\xrightarrow{d}N\left(\frac{\lambda(1+\gamma(\beta-1))}{1-\rho+\gamma(\beta-1)},\frac{\gamma^2(1+\gamma(\beta-1))^2}{1+2\gamma(\beta-1)}\right).$$

□

Proof of Corollary 1 The proof of (a) and (c) is straightforward. The assertion (b) follows from

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\mathrm{AMSE}(1)>0 \quad \text{for } \rho<0. \tag{18}$$

□

Proof of Theorem 3 The idea of the proof is based on de Haan and Ferreira (2006, p. 79). Setting $t=n/k$ we minimize the approximation of $\mathrm{AMSE}(H_{n,k}^{(\beta)})$, i.e.

$$\arg\min_{t>0}\left(n^{-1}t\sigma_\beta^2+A^2(t)\mu_\beta^2\right). \tag{19}$$

Since $A(t)$ is a regularly varying function with index ρ, there exists a positive decreasing function $s \in RV_{2\rho-1}$ such that as $t \to \infty$,

$$A^2(t) \sim \int_t^\infty s(u)\mathrm{d}u.$$

Thus, for any $c > 1$ and sufficiently large t we have

$$n^{-1}t\sigma_\beta^2 + c^{-1}\mu_\beta^2 \int_t^\infty s(u)\mathrm{d}u < n^{-1}t\sigma_\beta^2 + A^2(t)\mu_\beta^2 < n^{-1}t\sigma_\beta^2 + c\mu_\beta^2 \int_t^\infty s(u)\mathrm{d}u. \tag{20}$$

The minimum of the right hand side is given by

$$\frac{\sigma_\beta^2}{cn\mu_\beta^2} = s(t) \quad \Longleftrightarrow \quad t = s^{\leftarrow}\left(\frac{\sigma_\beta^2}{cn\mu_\beta^2}\right) = s^{\leftarrow}(n^{-1}c^{-1}\gamma^2\tau_\beta).$$

Similarly, we obtain for the left hand side $t = s^{\leftarrow}(n^{-1}c\gamma^2\tau_\beta)$. Therefore, the infimum in (19) is attained at $t_0^{(\beta)} := s^{\leftarrow}(n^{-1}\gamma^2\tau_\beta)$. Replacing t by n/k yields

$$k_0^{(\beta)} \sim \frac{n}{s^{\leftarrow}(n^{-1}\gamma^2\tau_\beta)}.$$

Note that the optimal sequence of order statistics depends on β. Now, considering the asymptotic distribution of $H_{n,k_0^{(\beta)}}^{(\beta)}$, we can write

$$\sqrt{k_0^{(\beta)}}\left(H_{n,k_0^{(\beta)}}^{(\beta)} - \gamma\right) \overset{d}{\approx} \sigma_\beta Z + \sqrt{k_0^{(\beta)}} A\left(n/k_0^{(\beta)}\right)\mu_\beta, \tag{21}$$

where Z is normal distributed. To obtain an expression of $\mathrm{AMSE}(H_{n,k_0}^{(\beta)})$, we have to evaluate $\sqrt{k_0^{(\beta)}} A\left(n/k_0^{(\beta)}\right)$ for large n. Therefore, note that

$$k_0^{(\beta)} A^2\left(n/k_0^{(\beta)}\right) \sim \frac{n}{t_0^{(\beta)}} \cdot \int_{t_0^{(\beta)}}^\infty s(u)\mathrm{d}u.$$

Now,

$$vs^{\leftarrow}(v) + \int_{s^{\leftarrow}(v)}^\infty s(u)\mathrm{d}u = \int_0^v s^{\leftarrow}(u)\mathrm{d}u$$

leads to

$$k_0^{(\beta)} A^2\left(n/k_0^{(\beta)}\right) \sim \frac{n}{s^{\leftarrow}(n^{-1}\gamma^2\tau_\beta)}\left[\int_0^{\gamma^2\tau_\beta/n} s^{\leftarrow}(u)\mathrm{d}u - \frac{\gamma^2\tau_\beta}{n}s^{\leftarrow}(\gamma^2\tau_\beta/n)\right]$$
$$= \gamma^2\tau_\beta\left(\underbrace{\frac{\int_0^{\gamma^2\tau_\beta/n} s^{\leftarrow}(u)\mathrm{d}u}{n^{-1}\gamma^2\tau_\beta\, s^{\leftarrow}(n^{-1}\gamma^2\tau_\beta)}}_{=:D} - 1\right).$$

We need the following result in order to proceed.

Proposition 1 (Karamata's theorem) *Suppose* $f \in RV_\alpha$. *If* $\alpha < -1$, *then*

$$\lim_{t\to\infty} \frac{tf(t)}{\int_t^\infty f(s)\mathrm{d}s} = -\alpha - 1. \tag{22}$$

Since $(1/s)^{\leftarrow} \in RV_{1/(1-2\rho)}$ and $s^{\leftarrow}(1/x) = (1/s)^{\leftarrow}(x)$, we obtain

$$\lim_{x\to\infty} \frac{\int_0^{1/x} s^{\leftarrow}(u)\mathrm{d}u}{x^{-1}s^{\leftarrow}\left(x^{-1}\right)} = \lim_{x\to\infty} \frac{\int_x^\infty (1/s)^{\leftarrow}(u)\frac{\mathrm{d}u}{u^2}}{x^{-1}(1/s)^{\leftarrow}(x)} = \frac{1-2\rho}{-2\rho}.$$

Note, $x^{-2}(1/s)^{\leftarrow}(x) \in RV_{1/(1-2\rho)-2}$. Hence, for $n\to\infty$, $n^{-1}\gamma^2\tau_\beta \to 0$ and

$$\lim_{n\to\infty} D = \frac{1-2\rho}{-2\rho}.$$

Altogether, we obtain for $\rho < 0$,

$$\lim_{n\to\infty} k_0^{(\beta)} A^2\left(n/k_0^{(\beta)}\right) = \gamma^2\tau_\beta\left(\frac{1-2\rho}{-2\rho} - 1\right) = \frac{\gamma^2\tau_\beta}{-2\rho}$$

and therefore

$$\sqrt{k_0^{(\beta)}}\left(H^{(\beta)}_{n,k_0^{(\beta)}} - \gamma\right) \xrightarrow{d} N\left(\frac{\mathrm{sign}(A)}{\sqrt{-2\rho}}\sigma_\beta, \sigma_\beta^2\right).$$

For Hill's estimator there exist a corresponding result (de Haan and Ferreira 2006)

$$\sqrt{k_0^{(1)}}\left(H_{n,k_0^{(1)}} - \gamma\right) \xrightarrow{d} N\left(\frac{\mathrm{sign}(A)\gamma}{\sqrt{-2\rho}}, \gamma^2\right).$$

Inserting $k_0^{(\beta)}$ into the approximation of $\mathrm{AMSE}(H^{(\beta)}_{n,k_0^{(\beta)}})$ yields

$$\mathrm{AMSE}\left(H^{(\beta)}_{n,k_0^{(\beta)}}\right) = \frac{1}{k_0^{(\beta)}}\sigma_\beta^2\left(1 - \frac{1}{2\rho}\right) = \frac{s^{\leftarrow}\left(n^{-1}\gamma^2\tau_\beta\right)}{n}\sigma_\beta^2\left(1 - \frac{1}{2\rho}\right).$$

Hence, without explicit knowledge of the function s it is not possible to minimize $\mathrm{AMSE}(H^{(\beta)}_{n,k_0^{(\beta)}})$ with respect to β. However, we can use the regular variation of s to compare the $\mathrm{AMSE}(H^{(\beta)}_{n,k_0^{(\beta)}})$ for two different values of β. This was already done in de Haan and Peng (1998) for different tail index estimators including Hill's method. □

Proof of Corollary 2 Considering the ratio of $\mathrm{AMSE}(H^{(\beta)}_{n,k_0^{(\beta)}})$ and $\mathrm{AMSE}(H^{(1)}_{n,k_0^{(1)}})$ yields

$$\mathrm{eff}_0(\beta, 1) = \frac{s^{\leftarrow}\left(n^{-1}\gamma^2\tau_\beta\right)}{s^{\leftarrow}\left(n^{-1}\gamma^2\tau_1\right)} \frac{(1+\gamma(\beta-1))^2}{1+2\gamma(\beta-1)}.$$

Since $s^{\leftarrow} \in RV_{1/(2\rho-1)}$, we have

$$\lim_{n\to\infty} \frac{s^{\leftarrow}\left(n^{-1}\gamma^2\tau_\beta\right)}{s^{\leftarrow}\left(n^{-1}\gamma^2\tau_1\right)} = \left((1-\rho)^{-2}\tau_\beta\right)^{1/(2\rho-1)}.$$

Therefore, we obtain

$$\mathrm{eff}_0(\beta, 1) = \left((1-\rho)^{-2}\gamma^2\tau_\beta\right)^{1/(2\rho-1)} \frac{(1+\gamma(\beta-1))^2}{1+2\gamma(\beta-1)}.$$

Moreover, we have

$$\frac{\partial}{\partial\beta} eff_0(\beta, 1)\bigg|_{\beta=1} \frac{\partial}{\partial\beta}\mathrm{eff}_0(\beta, 1)\bigg|_{\beta=1} = \frac{1}{2\rho-1}\frac{2\gamma\rho}{(1-\rho)} > 0,$$

since $\rho < 0$. Thus, for any distribution with regularly varying tail there exists some $\beta < 1$ such that $H^{(\beta)}_{n,k_0^{(\beta)}}$ outperforms $H^{(1)}_{n,k_0^{(1)}}$. □

Proof of Corollary 3 Using the results from Corollary 2, the proof is straightforward. □

Proof of $B^{(\beta)}_{n,k}$. To derive the limits of $B^{(\beta)}_{n,k}$, we will use

$$\hat{x}^{(\beta)}_{H,k} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-k,n}}{X_{n-i+1,n}}\right)^{\beta-1} \overset{P}{\to} \frac{1}{\gamma(\beta-1)+1}.$$

Simple calculations yield

$$\Delta H^{(\beta)}_{n,k}(x) = \frac{1}{\beta-1}\frac{1}{(b+a)}\left(1-\frac{a}{(k-1)^{-1}b}\right),$$

where

$$a := a(x) = \left(\frac{X_{n-k,n}}{x}\right)^{\beta-1} \quad \text{and} \quad b := \sum_{i=2}^{k}\left(\frac{X_{n-k,n}}{X_{n-i+1,n}}\right)^{\beta-1}.$$

For $\beta > 1$, we have $\lim_{x\to\infty} a(x) = 0$.
Moreover,

$$\frac{b}{k-1} = \frac{1}{k-1}\sum_{i=1}^{k-1}\left(\frac{X_{n-k,n-1}}{X_{n-i+1,n-1}}\right)^{\beta-1} \xrightarrow{P} \frac{1}{\gamma(\beta-1)+1}.$$

Thus,

$$B_{n,k}^{(\beta)} = \frac{1}{\beta-1}\frac{1}{b} = \frac{1}{\beta-1}\frac{1}{k-1}\frac{k-1}{b} \xrightarrow{P} 0.$$

For $\beta = 1$, we have

$$\Delta H_{n,k}^{(1)}(x) = \frac{1}{k}\sum_{i=2}^{k}\log(X_{n-i+1,n}) + \frac{1}{k}\log(x) - \left(\frac{1}{k-1}\sum_{i=2}^{k}\log(X_{n-i+1,n})\right)$$
$$= \frac{1}{k}\left(\log(x) - \frac{1}{k-1}\sum_{i=2}^{k}\log(X_{n-i+1,n})\right) \overset{x\to\infty}{\rightarrow} \infty = B_{n,k}^{(1)}.$$

For $1-(2\gamma)^{-1} < \beta < 1$, we have $a(x) \to \infty$ as $x \to \infty$ and

$$B_{n,k}^{(\beta)} = \frac{1}{1-\beta}\left(\frac{1}{k-1}\sum_{i=2}^{k}\left(\frac{X_{n-k,n}}{X_{n-i+1,n}}\right)^{\beta-1}\right)^{-1} \xrightarrow{P} \frac{\gamma(\beta-1)+1}{1-\beta} = \frac{1}{1-\beta} - \gamma.$$

□

Acknowledgments We would like to thank a referee and the associate editor for their constructive remarks that led to an improved presentation of the results. This research has been supported by the DFG (grant BE 2123/10-1).

References

Balkema, A., de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, *2*, 792–804.

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

Beirlant, J., Figueiredo, F., Gomes, M., Vandewalle, B. (2008). Improved reduced-bias tail index and quantile estimators. *Journal of Statistical Planning and Inference*, *138*, 1851–1870.

Beran, J., Schell, D. (2012). On robust tail index estimation. *Computational Statistics and Data Analysis*, *56*(11), 3430–3443.

Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.

Brazauskas, V., Serfling, R. (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal*, *4*, 12–27.

Cheng, S., Peng, L. (2001). Confidence intervals for the tail index. *Bernoulli*, *7*(5), 751–760.

Csörgő, S., Mason, D. M. (1985). Central limit theorems for sums of extreme values. *Mathematical Proceedings of the Cambridge Philosophical Society*, *98*(3), 547–558.

Csörgő, S., Viharos, L. (1997). Asymptotic normality of least-squares estimators of tail indices. *Bernoulli, (3)*3, 351–370.

Csörgő, S., Deheuvels, P., Mason, D.M. (1985). Kernel estimates of the tail index of a distribution. *The Annals of Statistics*, *13*, 1050–1077.

Davis, R., Resnick, S. (1984). Tail estimates motivated by extreme value theory. *The Annals of Statistics*, *12*, 1467–1487.

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

de Haan, L., Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, *52*, 60–70.

de Haan, L., Resnick, S. (1998). On asymptotic normality of the hill estimator. *Communications in Statistics. Stochastic Models*, *14*, 849–866.

Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling extremal events*. New York: Springer.

Fabián, Z., Stehlík, M. (2009). *On robust and distribution sensitive Hill like method. IFAS research report 43*. Linz: Department for Applied Statistics, Johannes Kepler University Linz.

Ferreira, A., de Vries, C. (2004). *Optimal confidence intervals of the tail index and high quantiles. Discussion paper TI 2004-090/2*. The Netherlands: Tinbergen Institute.

Finkelstein, M., Tucker, H.G., Veeh, J.A. (2006). Pareto tail index estimation revisited. *North American Actuarial Journal*, *10*(1), 1–10.

Haeusler, E., Teugels, J.L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics*, *13*(2), 743–756.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B*, *44*(1), 37–42.

Hampel, F. (1968). Contributions to the theory of robust estimation. Ph.D. thesis, Berkeley: Department of Statistics, University of California.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986). *Robust statistics: the approach based on influence functions*. New York: Wiley.

Henry III, J.B. (2009). A harmonic moment tail index estimator. *Journal of Statistical Theory and Applications*, *8*(2), 141–162.

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

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

Knight, K. (2012). A simple modification of the Hill estimator with applications to robustness and bias reduction (preprint).

Lu, J.-C., Peng, L. (2002). Likelihood based confidence intervals for the tail index. *Extremes*, *5*, 337–352.

Mason, D. (1982). Laws of large numbers of sums of extreme values. *The Annals of Probability*, *10*, 168–177.

Peng, L., Welsh, A. (2002). Robust estimation of the generalized Pareto distribution. *Extremes*, *4*, 53–65.

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

Qi, Y. (2008). Bootstrap and empirical likelihood methods in extremes. *Extremes*, *11*, 81–97.

Reiss, R., Thomas, M. (2005). *Statistical analysis of extreme values (for insurance, finance, hydrology and other fields)* (3rd rev ed.). Basel: Birkhäuser.

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

Stehlík, M., Potocký, R., Waldl, H., Fabián, Z. (2010). On the favourable estimation of fitting heavy tailed data. *Computational Statistics*, *25*, 485–503.

Stehlík, M., Fabián, Z., Střelec, L. (2012). Small sample robust testing for normality against Pareto tails. *Communications in Statistics: Simulation and Computation*, *41*(7), 1167–1194.

Vandewalle, B. (2004). Some robust and semi-parametric methods in extreme value theory. Doctoral thesis, Leuven: Department of Mathematics, Katholieke Universiteit Leuven.

Vandewalle, B., Beirlant, J., Hubert, M. (2004). A robust estimator of the tail index based on an exponential regression model. In M. Hubert, G. Pison, A. Struyf, S. van Aelst (Eds.), *Theory and applications of recent robust methods* (pp. 367–376). Basel: Birkhauser.

Vandewalle, B., Beirlant, J., Christmann, A., Hubert, M. (2007). A robust estimator for tail index of Pareto-type distributions. *Computational Statistics and Data Analysis*, *51*, 6252–6268.

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

Worms, J., Worms, R. (2011). Empirical likelihood based confidence regions for first order parameters of heavy-tailed distributions. *Journal of Statistical Planning and Inference*, *141*(8), 2769–2786.