Journal of Statistical Planning and Inference 170 (2016) 27–45

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Tail index estimation with a fixed tuning parameter fraction

Tucker McElroy [a,*], Chaitra H. Nagaraja [b]

[a] U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233, USA

[b] Gabelli School of Business, Fordham University, 5 Columbus Circle, Rm. 1116, New York, NY 10023, USA

ARTICLE INFO

Article history:
Received 19 August 2014
Received in revised form 4 August 2015
Accepted 19 August 2015
Available online 3 September 2015

Keywords:
Extreme value theory
Heavy tails
Stable distributions

ABSTRACT

Semi-parametric tail index estimators, such as the Hill, Harmonic Moment, Pickands, and Dekkers, Einmahl and de Haan estimators, rely upon a tuning parameter that typically grows with sample size n. Proper selection of this tuning parameter $k = k(n)$ is crucial for good practical performance, although asymptotic theory dictates that $1/k + k/n \to 0$ as $n \to \infty$. A similar issue presents itself in the bandwidth literature in spectral density estimation and recent research shows that the use of asymptotic distributions when the bandwidth is a fixed ratio of sample size yields improved approximations to finite-sample distributions. Here, we study some semi-parametric tail index estimators utilizing the same perspective where $k = bn$ and $b \in (0, 1)$ is a fixed constant. This allows us to derive asymptotic bias and variance expressions which are compatible with the small-b conventional theory. Our simulations corroborate that the finite-sample bias and variance are well described by the asymptotic bias and variance quantities arising from our fixed bandwidth ratio theory.

Published by Elsevier B.V.

1. Introduction

Estimation of the tail index of a heavy-tailed distribution is a long-standing statistical problem, with numerous applications such as in finance or insurance, for example. Embrechts et al. (1997) and Gomes and Guillou (2014) provide an overview of the topic, while Mandelbrot and Hudson (2008) give a historical perspective. For specific applications, see Crovella et al. (1998) for heavy tails in internet traffic; Huisman et al. (2001) for heavy-tailed financial returns; and Lachtman et al. (2008) for extreme events in seismology. Early efforts focused on regression-based techniques to obtain tail index estimators; these relied upon a graphical method of estimation developed by Mandelbrot (1963, 1982) and are linked to the Hill estimator. Viharos (1999) later proposed a weighted least-squares method and Aban and Meerschaert (2004) devised a generalized least squares-based procedure. Csörgő and Viharos (1997) derived the asymptotic properties of these regression-based estimators.

In this paper, we focus upon a related class of semi-parametric tail index estimators – in particular, the Hill, Harmonic Moment, Pickands, and DEdH estimators (Hill, 1975, Beran et al., 2014, Pickands, 1975, and Dekkers et al., 1989 respectively) – each of which depend on a selected number of order statistics. We chose these estimators to represent a range of approaches to the tail index estimator problem, even though the Pickands estimator is rarely used and, for our purposes, mainly of historical interest.

The range of order statistics used for each type of estimator is governed by a tuning parameter k that grows with sample size n. That is, the asymptotic theory which motivates the definition of these estimators requires that $1/k + k/n \to 0$

* Corresponding author.
E-mail addresses: tucker.s.mcelroy@census.gov (T. McElroy), cnagaraja@fordham.edu (C.H. Nagaraja).

http://dx.doi.org/10.1016/j.jspi.2015.08.005
0378-3758/Published by Elsevier B.V.

as $n \to \infty$. However, generally, a finite sample of data $X_1, X_2, \ldots, X_n$ has n fixed for most analyses, so that viewing k as a sequence depending on sample size is only a theoretical heuristic. As results depend heavily on the choice of k – see the discussion in Embrechts et al. (1997) – its proper selection is critical. High values of k tend to decrease the asymptotic variance, but unfortunately they also tend to increase the bias. Moreover, balancing the bias and variance requires knowledge of the data distribution, which usually is unavailable.

Many procedures have been proposed for optimal selection of the tuning parameter k, which are reviewed in Gomes and Guillou (2014). To get a better grasp on the optimal selection of k for a given sample, it is necessary to have additional knowledge of the underlying stochastic structure, e.g., second-order conditions on slowly-varying functions appearing in the quantile function.

The selection of k finds a natural analogy in the large time series literature on optimal bandwidth selection for kernel estimates of the spectral density, elucidated in Parzen (1959), among others. The bandwidth parameter is also a tuning parameter of the estimator, which can likewise be viewed as semi-parametric, with optimality of the tuning achieved through some additional knowledge of the underlying stochastic structure. More recently, a fixed bandwidth-ratio literature has been advanced in Kiefer et al. (2000), using the perspective that actual bandwidth can always be expressed as some fraction of the sample size n. This has spurred a number of related papers which seem to be more useful than the classical theory in describing the finite-sample distributions of the estimators (see Sun et al., 2008 as an example). Simulation studies show that this new asymptotic distribution formulation (which is pivotal in the case of spectral density estimation) provides a superior approximation to the finite-sample distribution. Selection of the optimal tuning fraction was studied in Dekkers and de Haan (1993) – in which the authors adopt the classical perspective that $b \to 0$ – and it was there shown that the optimal fraction can be expressed as a power of sample size, with various constants; from their examples, it is clear that the exact formulation depends heavily on the particular distribution under consideration. Our results are similar in spirit, although the bias and variance expressions are seen to be comparatively simple expressions involving the quantile function.

Motivated by this perspective, we apply the same approach to tail index estimation. That is, we suppose that the tuning parameter k is some fixed ratio b of the sample size, with $k = bn$ and $b \in (0, 1)$. (In the context of the limit theory for order statistics, this approach has a precedent in the "central order property" of Smirnov, 1949, whose work was extended by Pancheva and Gacovska, 2014.) The classical theory always has $b \to 0$ to ensure consistent and unbiased estimation for the Hill, Harmonic Moment, Pickands, and DEdH estimators, but we will allow b to be fixed in our asymptotic analysis. The idea is that an asymptotic formulation with fixed b, rather than b negligible, will offer an improved approximation to the finite-sample distribution, even though this will result in an inconsistent estimator. Essentially, we show that the estimators are asymptotically normal (we let the bias and variance functions be denoted as $B(b)$ and $\mathrm{Var}(b)$, respectively). Furthermore, compared to the classical values of $B(0)$ and $\mathrm{Var}(0)$, our methods result in better approximations to the actual finite-sample bias and variance. (We show below that $B(0) = 0$, and that $\mathrm{Var}(0)$ corresponds to the classical theory's variance, for each of the estimators.) While neither $B(b)$ nor $\mathrm{Var}(b)$ is pivotal, the latter is estimable, and thus superior inference for the tail index can be based on $\widehat{\mathrm{Var}}(b)$ in lieu of $\widehat{\mathrm{Var}}(0)$.

The approach of this paper is in the spirit of penultimate approximations, which are discussed in Gomes and de Haan (1999), Gomes and Guillou (2014), and Reis et al. (2015). Suppose we have a statistic T_n and parameter ξ, and $g(T_n, \xi)$ is a pivot. We may know that the pivot's cumulative distribution function (cdf) $F_{g(T_n,\xi)}$ converges to some F_ξ, but perhaps there exists a family of cdfs F_{ξ_n} with $F_{\xi_n} \to F_\xi$, but with F_{ξ_n} yielding a closer approximation to $F_{g(T_n,\xi)}$. If the F_{ξ_n} are computable or estimable, then they are penultimate approximations to the pivot, and are to be preferred over F_ξ as an approximation. Our results take the form

$$\sqrt{k}(T_n - \xi - B(b)) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, b\mathrm{Var}(b)),$$

which means that the cdf of the pivot $\sqrt{k}(T_n - \xi)$ has a penultimate approximation with cdf (evaluated at x) given by $\Phi([x - \sqrt{nb}B(b)]/\sqrt{b\mathrm{Var}(b)})$, where Φ is the standard normal cdf, if we assume that $b \to 0$. That is, if one adopts the classical perspective (that $b \to 0$), then our limit results can be viewed as a sequence of penultimate approximations for the pivot.

One way to investigate whether the fixed tuning parameter ratio asymptotics provide any advantage – in terms of generating a penultimate approximation of the finite-sample distribution – is via simulation. In this paper, we compute the estimators with various choices of k on simulated data, and apply both the classical asymptotic theory and our own theory; we obtain the finding that the new asymptotic theory gives highly accurate bias and variance expressions. These asymptotic bias and variance expressions can be jointly minimized with respect to b on a case by case basis, so that we can see how the optimal tuning fraction depends upon each particular distribution. By considering many classes of heavy-tailed distributions, we can draw some general conclusions about the optimal b for the Hill, Harmonic Moment, Pickands, and DEdH estimators.

This paper presents theory for the fixed tuning parameter fraction, providing new results for the Hill, Harmonic Moment, Pickands, and DEdH estimators in Section 2. We make comparisons to the classical small-b tuning parameter fraction case as well. In Section 3, we apply this theory to simulated heavy-tailed data and assess the results against the derived bias and asymptotic normality results. Moreover, we show that using the optimal tuning fraction results in superior finite sample performance. In order to be useful, one needs to estimate the bias $B(b)$; bias estimation is discussed in Gomes and Guillou (2014). One can also utilize the new expressions from this paper to construct estimators of bias, which is outlined in Section 3. Section 4 concludes and the Appendix contains all proofs.

Table 1
Theoretical tail index values for selected distributions.

Distribution	pdf	Parameters	Tail index (α)
Burr	$c\nu\left\{x^{c-1}/(1+x^c)^{\nu+1}\right\}$	$\nu > 0, c > 0$	$c\nu$
Cauchy	$1/\left(\pi\gamma\left[1+\{((x-x_0)/\gamma)\}^2\right]\right)$	$x_0 \in \mathbb{R}, \gamma > 0\,(\alpha = 1, \beta = 0$ in stable class)	1
Log-gamma	$\left(\nu^\gamma\,(\log x)^{\gamma-1}\right)/\left(\Gamma(\gamma)x^{\nu+1}\right)$	$\gamma > 0,\ \nu > 0$	ν
Pareto	$\gamma x_0^\gamma/x^{\gamma+1}$	$x_0 > 0, \gamma > 0$	γ
Stable class	–	$\alpha \in (0, 2], \beta \in [-1, 1], x_0 \in \mathbb{R}, \gamma > 0$	α
t	$\frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}(1+x^2/\nu)^{-(\nu+1)/2}$	$\nu > 0$	ν

2. New asymptotic results for tail index estimators

Consider an independent and identically distributed (i.i.d.) sample $X_1, \ldots, X_n$ drawn from the upper right tail distribution function $\overline{F}(x) = x^{-\alpha}L(x)$ where $x > 0$. Here, $\alpha \in (0, \infty)$ with smaller values of α corresponding to more extreme events. Also, let $L \in \mathcal{L}$, the set of functions slowly-varying at infinity (Embrechts et al., 1997). We will concentrate our treatment on the upper right tail of the distribution, as is common in the classical tail index estimation literature (Hill, 1975, Pickands, 1975, Dekkers et al., 1989, etc.); extensions to the lower left tail can be made in principle. The quantile function is the generalized inverse of the cumulative distribution function, mapping $u \in [0, 1]$ to $\overline{\mathbb{R}} = [-\infty, \infty]$. Let $\overline{F}^{-1}(u)$ denote this generalized inverse for the right tail so that $Q(1-u) = \inf\{x : \overline{F}(x) \geq u\}$. Our interest naturally focuses on values of u close to zero (this represents the right tail). It follows from Proposition 1.5.15 of Bingham et al. (1987) that

$$Q(1-u) = u^{-1/\alpha}K(u) \tag{1}$$

for $u \in (0, 1/2)$ where K is slowly-varying as $u \to 0$. Note that α controls the tails, but the center of the distribution is determined by K; in subsequent results, we show how the form of K has an impact on the finite-sample bias and variance of common tail index estimators. Let $\overline{Q}$ denote the right quantile function, namely $\overline{Q}(u) = Q(1-u)$; also, let $P(x) = K(1/x)$ for $x > 1$, which will be useful subsequently.

The task of tail index estimation is inference for α in (1). It is well-known that there is a trade-off between bias and variance in estimators dependent upon a tuning parameter (e.g., Hill and Pickands estimators—see discussion in Embrechts et al., 1997). Many such estimators depend upon a number of order statistics k, assumed to grow with sample size, and yet $k = o(n)$. A larger value of k decreases the variance while also increasing the bias. We propose to study this issue by adapting the machinery of fixed-bandwidth ratio asymptotics to this setting. Therefore, we shall assume that $k = \lfloor bn \rfloor$ for $b \in (0, 1)$, a fixed proportion. (Then $k/n \to b$ as $n \to \infty$; as with the fixed-b bandwidth literature, we shall write $k/n = b$.) Heuristically speaking, $b = 0$ corresponds to the small-bandwidth theory presently used. Our first goal is to develop asymptotic results for the Hill, Harmonic Moment, Pickands, and DEdH estimators under this modified framework.

Our results depend upon the quantile-density function, defined as $q(u) = \dot{Q}(u)$, and we assume that Q is sufficiently smooth such that q exists. A sufficient condition for this is that K is differentiable; in this case, P is also differentiable and $\lim_{u\to 0} u\dot{K}(u)/K(u) = -\lim_{x\to\infty} x\dot{P}(x)/P(x)$. For some of our results in this section, it is useful to assume that this limit is zero. To that end, recall that for any $P \in \mathcal{L}$ we have the representation $P(x) = c(x)\exp\{\int_z^x \eta(v)/v\,dv\}$ for $\eta(v) = o(1)$ and $c(x) \to c_0 > 0$ as $x \to \infty$ (cf., Embrechts et al., 1997). Therefore, in order to ensure that $\lim_{u\to 0} u\dot{K}(u)/K(u) = 0$, it is sufficient to assume that c is smooth and $\dot{c}(x) = o(1/x)$ as $x \to \infty$. The subclass of such slowly-varying functions P that are differentiable will be denoted as $\mathcal{L}_d$, and the further subclass with the property that $\dot{c}(x) = o(1/x)$ will be denoted as $\mathcal{L}_c$.

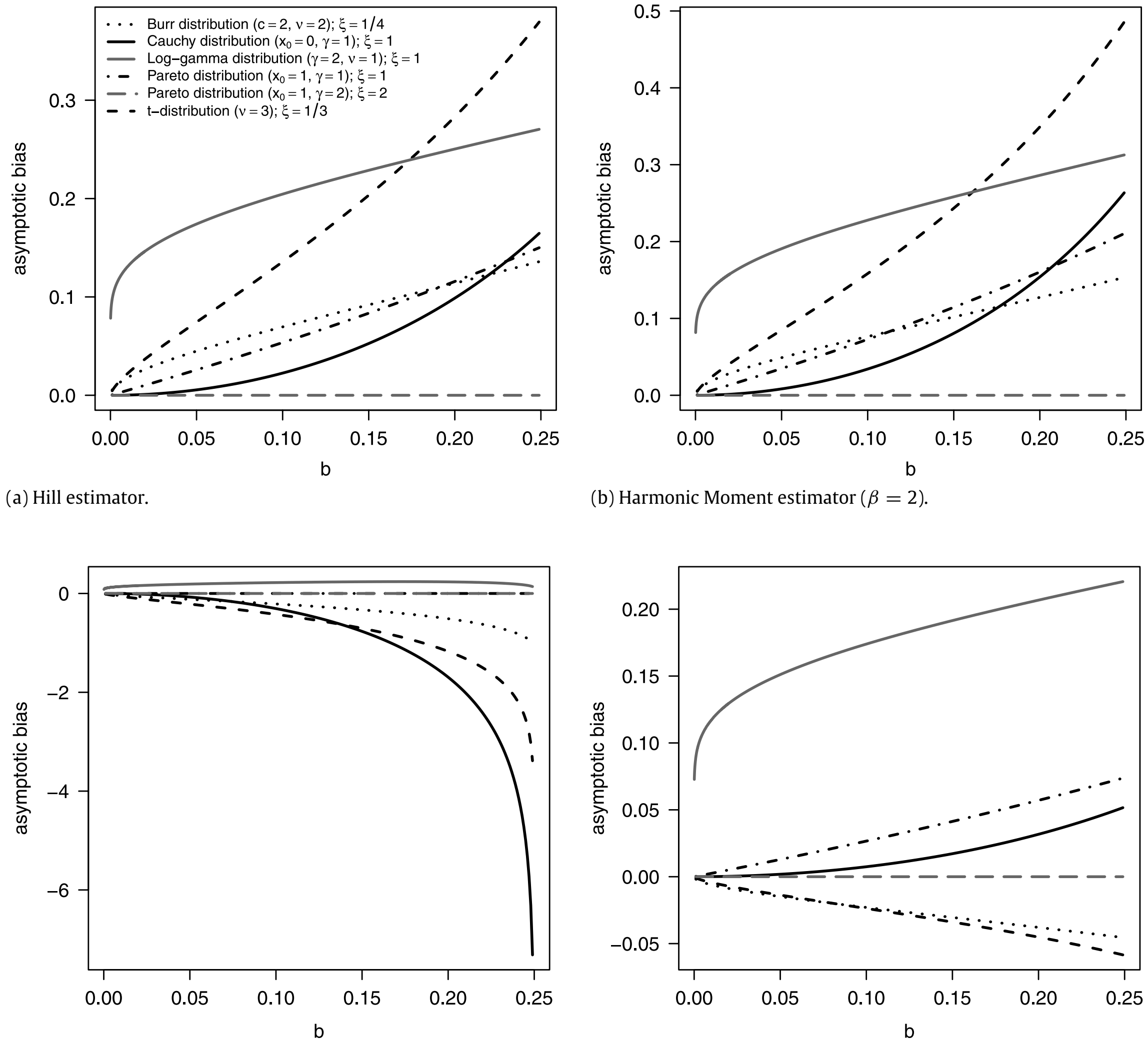
We present our results for the estimators, Hill, Harmonic Moment, Pickands, and DEdH, in the remainder of this section. To provide some context, Table 1 lists the tail index value α for a variety of distributions (see Embrechts et al., 1997 for additional distributions). Furthermore, we visualize the relationship between the asymptotic bias and choice of sample fraction b for each estimator in Fig. 1. These graphs show, for all four estimators, that the asymptotic bias goes to zero as $b \to 0$ for a range of distributions.

2.1. Theory for the Hill estimator

We begin with some definitions. The tail index in (1) is the parameter α, whereas the quantity $\xi = 1/\alpha$ can also be the subject of inference, sometimes called the extreme value index. It is more convenient for us to develop the asymptotic theory for ξ.

First, we define the log quantile moments:

$$H_{(\ell)} = \frac{1}{k}\sum_{j=1}^{k}\left\{\log\widehat{Q}(1-j/n) - \log\widehat{Q}(1-k/n)\right\}^{\ell}. \tag{2}$$



(a) Hill estimator.

(b) Harmonic Moment estimator ($\beta = 2$).

(c) Pickands estimator.

(d) DEdH estimator.

Fig. 1. Asymptotic bias as $b \to 0$ for estimates computed from six data distributions using the (a) Hill, (b) Harmonic Moment ($\beta = 2$), (c) Pickands, and (d) DEdH methods. Note: For the log-gamma distribution with $\gamma = 2$, $\nu = 1$, shown as a solid gray line on the plots, the asymptotic bias behaves as expected; however, it is difficult to show numerically, given the steep slope near zero.

Then, the Hill estimator of ξ is defined to be

$$\widehat{\xi}^{(H)} = H_{(1)}. \tag{3}$$

See Hill (1975) and Embrechts et al. (1997) for the formulation given here. Typically, the statistic is presented in terms of upper order statistics of the sample, but we have simply rewritten this in terms of the empirical quantile function $\widehat{Q}(u)$, the inverse of the empirical distribution function (i.e., namely, $\widehat{Q}(u) = X_{(\lfloor nu \rfloor)}$). By evaluating $\widehat{Q}(u)$ at the percentile $1 - j/n$, we obtain the upper right-hand side of the distribution.

Our asymptotic result for the Hill estimator (3) is given below. Although the Hill estimator is the $\beta = 1$ case of the Harmonic Moment estimator, the analysis of the Hill is slightly different so we consider these estimators in separate subsections. The asymptotic theory is formulated for a right quantile function $\overline{Q}$ that has a bounded derivative on some compact subset of $(0, 1/2]$; it is necessary to exclude the boundary point of zero because by (1) the derivative of our quantile function blows up at $u = 0$. This compact set might be written $[\epsilon, 1/2]$ with $\epsilon > 0$ and arbitrarily small. In practice, this device has no impact on data analysis as we typically think of ϵ as being smaller than $1/n$. Although for our results the limits of integration on integrals should run from ϵ to b instead of from 0 to b, we use the latter bounds because the discrepancy is negligible and the resulting expressions are far more compact and interpretable. (See the discussion in the proof of Theorem 1 as well in the Appendix.)

Before we present our main result, it is helpful to define a few quantities that describe the asymptotic bias and variance of the Hill estimator. $\overline{K}_1$ is a function of the slowly-varying function $K(u)$, and corresponds to the case $\ell = 1$ of

$$\overline{K}_\ell = b^{-1} \int_0^b (\log K(u) - \log K(b))^\ell \, du. \tag{4}$$

This is the key term in the bias, as shown in Theorem 1. Also, we define integrals of the log quantile function as follows:

$$\tau_\ell = b^{-1} \int_0^b (\log Q(1-u) - \log Q(1-b))^\ell \, du. \tag{5}$$

Recall the function $x\dot{K}(x)/K(x)$, which we denote by $R(x)$ for short. Observe that if $P \in \mathcal{L}_c$, then $R(x) \to 0$ as $x \to 0$. Next, we introduce another quantity related to the derivative of the log quantile function:

$$\lambda = b \, \frac{q(1-b)}{Q(1-b)} = \xi - \frac{b\dot{K}(b)}{K(b)} = \xi - R(b). \tag{6}$$

Note that $\lambda = -b\frac{\partial}{\partial b} \log Q(1-b)$, and $\lambda \to \xi$ as $b \to 0$ so long as $P \in \mathcal{L}_c$. In addition, all the quantities $\overline{K}_1$, τ_1, τ_2, and λ depend upon b. Finally, let W denote a standard Brownian Motion on $[0, 1]$ and let B be the Brownian Bridge defined by $B(u) = W(u) - uW(1)$.

Theorem 1. *Suppose that the right quantile function $\overline{Q}$ has a bounded derivative on some compact interval $[\epsilon, 1/2]$ for a small $\epsilon > 0$ and with $P \in \mathcal{L}_d$ and that the tuning fraction $b \leq 1/2$ is fixed. Then, the Hill estimator, computed from an i.i.d. sample with quantile function (1), has the following asymptotic behavior as $n \to \infty$. The asymptotic bias of the Hill estimator is*

$$B^{(H)}_{\widehat{\xi}} = \overline{K}_1, \tag{7}$$

and the bias-corrected estimator satisfies

$$\sqrt{n}\left(\widehat{\xi}^{(H)} - \xi - B^{(H)}_{\widehat{\xi}}\right) \overset{\mathcal{L}}{\Longrightarrow} b^{-1} \int_0^b \frac{q(1-u)}{Q(1-u)} B(1-u)\, du - b^{-1}\lambda\, B(1-b).$$

The limit, denoted by U_1, is Gaussian with mean zero and variance

$$b^{-1}(\tau_2 - \tau_1^2) + (b^{-1} - 1)(\lambda - \tau_1)^2. \tag{8}$$

Also, $\lim_{b\to 0} B^{(H)}_{\widehat{\xi}} = 0$ and when $P \in \mathcal{L}_c$, we have $\lim_{b\to 0} b\,\mathrm{Var}[U_1] = \xi^2$.

Remark 1. The asymptotic bias is clearly laid out by Theorem 1; it tends to zero as $b \to 0$, which corroborates the small-bandwidth asymptotic theory (see Fig. 1(a) where the asymptotic bias is plotted against the sample fraction b for several distributions). However, for $b > 0$ the result is useful as we can see the exact role of K – the unexplained part of the distribution – in the asymptotic bias. Embrechts et al. (1997) discusses the question of when and how ruin occurs; the framework of (1), and the bias and variance results of Theorem 1, illustrate how the whole distribution – not just the tails – has an impact (through the form of K) on tail index estimation.

Remark 2. Focusing on the stochastic portion of the limit, its variance depends on quantities that can be consistently estimated from the data. Multiplying the variance of U_1 by b is equivalent to normalizing the original statistic by $\sqrt{k}$ instead of $\sqrt{n}$; so we should expect to obtain the same asymptotic variance as the small b theory dictates. This is indeed the case—cf., Theorem 6.4.6 of Embrechts et al. (1997), which, however, is stated in terms of α rather than ξ.

We now make some connections between this result and the more classical "small b" results. The small b theory indicates that the limit is expressible as $\xi \int_0^1 u^{-1}A(u)\,du$ for some Brownian Bridge process A; so, we should expect that $\sqrt{b}\,U_1 \overset{\mathcal{L}}{\Longrightarrow} \xi \int_0^1 u^{-1}A(u)\,du$ as $b \to 0$. The limit of the Hill estimator is expressed this way in Theorem 2 of McElroy and Jach (2012). When $P \in \mathcal{L}_c$, it can be shown that

$$\sqrt{b}\left(\int_0^1 R(bu)\,du - R(b)\right) \to 0$$

as $b \to 0$. Then, letting $\widetilde{B}(u) = B(1-u)$ in U_1, noting that $\widetilde{B}$ is also a Brownian Bridge process, we obtain

$$\sqrt{b}\,U_1 = o_P(1) + \xi b^{-1/2}\left(\int_0^1 u^{-1}C(bu)\,du - C(b)\right),$$

where C is a Brownian Motion, such that $\widetilde{B}(u) = C(u) - uC(1)$. We can define a rescaled Brownian Motion via $C_b(u) = C(bu)/\sqrt{b}$, so that $C_b \overset{\mathcal{L}}{\Longrightarrow} C_0$ as $b \to 0$, another standard Brownian Motion. Hence,

$$\sqrt{b}\, U_1 \overset{\mathcal{L}}{\Longrightarrow} \xi \left(\int_0^1 u^{-1} C_0(u)\, du - C_0(1) \right) = \int_0^1 \frac{\xi}{u} A(u)\, du$$

for a Brownian Bridge A. This agrees with the limit in Kulik and Soulier (2011) and McElroy and Jach (2012).

There have been some developments on computing the asymptotic Hill bias, based on the small tuning fraction theory. Dacorogna et al. (1995), who extend the results from Hall (1990), give an asymptotic expression for the bias for distribution functions of the form

$$\overline{F}(x) = ax^{-\alpha}\ (1 + cx^{-\beta}) \tag{9}$$

where $\alpha, \beta > 0$ and $a, c \in \mathbb{R}$. Note that this corresponds to our general formulation with $L(x) = a\,(1 + cx^{-\beta})$, which is seen to be in $\mathcal{L}$ with unit limit. As reported in Huisman et al. (2001), an asymptotic bias expression in this case is given by

$$B_0(b) = -\frac{c\beta}{\alpha(\alpha + \beta)}\ (a/b)^{-\frac{\beta}{\alpha}}$$

where we have substituted b for $k(n)/n$. Our own theory, based on Theorem 1, focuses on the case that $b > 0$; nevertheless, it is natural to expect that $\lim_{b\to 0} \left(B_{\widehat{\xi}}^{(H)} / B_0(b) \right) = 1$. In fact, for distributions of the form (9), we show that

$$B_{\widehat{\xi}}^{(H)} = -\frac{\beta}{\alpha}\ \frac{a\, c\, b^{\frac{\beta}{\alpha}} K^{-(\alpha+\beta)}(b)}{\alpha + \beta} \tag{10}$$

in the Appendix. It can also be shown that $K(0)$ exists and is equal to $a^{1/\alpha}$, so that asymptotically as $b \to 0$, (10) is approximately $B_0(b)$.

2.2. Theory for the Harmonic Moment estimator

The Harmonic Moment estimator (Fabián and Stehlík, 2009) is similar to the Hill estimator, and has been exposited in Beran et al. (2014). For $\beta \neq 1$ (the $\beta = 1$ corresponds to Hill), it is defined by

$$\widehat{\xi}^{(HM)} = \frac{1}{\beta - 1} \left(\left[\widehat{Q}(1-b)^{\beta-1}\, \frac{1}{k} \sum_{j=1}^{k} \widehat{Q}(1-(j-1)/n)^{1-\beta} \right]^{-1} - 1 \right). \tag{11}$$

As with the Hill estimator, we view k as a fixed function of n, so that $k = bn$ with $b \in (0, 1/2)$. The assumptions on the quantile function are similar, but the quantities needed to describe the asymptotic bias and variance differ slightly. Also, as in Beran et al. (2014), we require β to satisfy $2(1-\beta) < \alpha$ which also implies that $1 - \beta < \alpha$. Then,

$$Q(1-x)^{1-\beta} = x^{-(1-\beta)/\alpha} K(x)^{1-\beta} \tag{12}$$

is integrable in a neighborhood of zero, and hence its integral over x is a neighborhood of zero which is well-defined. Now, define the function $G(x) = (\beta - 1)^{-1} Q(1-x)^{1-\beta}$, and write $\langle G^p \rangle = b^{-1} \int_0^b G(x)^p\, dx$ as a shorthand. Then, we have the analogues of the τ_ℓ quantities defined for the Hill estimator:

$$\upsilon_\ell = \langle G^\ell \rangle - G(b)^\ell. \tag{13}$$

Moreover, note that the derivative of $G(x)$ is $Q(1-x)^{-\beta} q(1-x)$ which we denote by $g(x)$. Using the product rule, we obtain

$$g(x) = (\beta - 1)\, G(x)\, x^{-1}\, (\xi - R(x)). \tag{14}$$

Thus, $bg(b)/[(\beta-1)G(b)] = \xi - R(b)$ which is an analogue of λ in (6) in our discussion of the Hill estimator. When $P \in \mathcal{L}_c$, the limit of this analogue of λ as $b \to 0$ is also ξ because $R(b) \to 0$. With these notations, we can present the following theorem.

Theorem 2. *Suppose that the right quantile function $\overline{Q}$ has a bounded derivative on some compact interval $[\epsilon, 1/2]$ for a small $\epsilon > 0$ and with $P \in \mathcal{L}_d$. Assume the tuning fraction $b \leq 1/2$ is fixed. Then, the Harmonic Moment estimator, computed from an i.i.d. sample with quantile function (1), has the following asymptotic behavior as $n \to \infty$. Let $c = G(b)/\langle G \rangle$. The asymptotic bias of the Harmonic Moment estimator is*

$$B_{\widehat{\xi}}^{(HM)} = \frac{c - 1}{\beta - 1} - \xi \tag{15}$$

and the bias-corrected estimator satisfies

$$\sqrt{n}\left(\widehat{\xi}^{(HM)} - \xi - B_{\widehat{\xi}}^{(HM)}\right) \overset{\mathcal{L}}{\Longrightarrow} \frac{\langle G\rangle^{-1}}{1-\beta}\left(g(b)B(1-b) - c\,b^{-1}\int_0^b g(x)\,B(1-x)\,dx\right).$$

The limit, denoted by V_1*, is Gaussian with mean zero and variance*

$$\left((b^{-1}-1)(bg(b)+c\upsilon_1)^2 + b^{-1}c^2(\upsilon_2-\upsilon_1^2)\right)\,(1-\beta)^{-2}\langle G\rangle^{-2}. \tag{16}$$

Also, $\lim_{b\to 0} B_{\widehat{\xi}}^{(HM)} = 0$ *and* $\lim_{b\to 0} b\,\mathrm{Var}[V_1] = \xi^2(1+(\beta-1)\xi)^2/(1+2(\beta-1)\xi)$.

Remark 3. The asymptotic bias and variance agree with the small-b asymptotic theory described in Beran et al. (2014), and is further corroborated by our numerical results (see Fig. 1(b) where we let $\beta = 2$, following Stehlík et al. (2010)).

2.3. Theory for the Pickands estimator

We again formulate our inference problem in terms of the extremal index. The Pickands estimator is defined via

$$\widehat{\xi}^{(P)} = (\log 2)^{-1}\,\log\left(\frac{\widehat{Q}(1-k/n)-\widehat{Q}(1-2k/n)}{\widehat{Q}(1-2k/n)-\widehat{Q}(1-4k/n)}\right). \tag{17}$$

See Pickands (1975) and Embrechts et al. (1997) for more discussion. Observe from (17) that we require $k \le n/4$. Consequently, we view k as a fixed function of n, namely $k = bn$ with $b \in (0, 1/4]$. We adopt similar assumptions on the quantile function as we did for the Hill estimator theory. We introduce the abbreviations $\Delta qB(x) = q(1-x)B(1-x) - q(1-2x)B(1-2x)$ and $\Delta Q(x) = Q(1-x) - Q(1-2x)$ for clarity purposes.

Theorem 3. *Suppose that the right quantile function* $\overline{Q}$ *has a bounded derivative on some compact interval* $[\epsilon, 1/4]$ *for a small* $\epsilon > 0$ *and with* $P \in \mathcal{L}_d$*. Furthermore, assume the tuning fraction* $b \le 1/4$ *is fixed. Then, the Pickands estimator, computed from an i.i.d. sample with quantile function (1), has the following asymptotic behavior as* $n \to \infty$*. The asymptotic bias of the Pickands estimator is*

$$B_{\widehat{\xi}}^{(P)} = (\log 2)^{-1}\,\log\left(\frac{K(b)-2^{-\xi}K(2b)}{K(2b)-2^{-\xi}K(4b)}\right), \tag{18}$$

and the bias-corrected estimator satisfies

$$\sqrt{n}\left(\widehat{\xi}^{(P)} - \xi - B_{\widehat{\xi}}^{(P)}\right) \overset{\mathcal{L}}{\Longrightarrow} (\log 2)^{-1}\left(\frac{\Delta qB(b)}{\Delta Q(b)} - \frac{\Delta qB(2b)}{\Delta Q(2b)}\right).$$

The limit, denoted by V*, is Gaussian with mean zero and variance given by*

$$\begin{aligned}(\log 2)^{-2}\Bigg\{&b(1-b)\left(\frac{q(1-b)}{\Delta Q(b)}\right)^2 + 2b(1-2b)\left(\frac{q(1-2b)}{\Delta Q(b)}\right)^2\\
&+2b(1-2b)\left(\frac{q(1-2b)}{\Delta Q(2b)}\right)^2 + 4b(1-4b)\left(\frac{q(1-4b)}{\Delta Q(2b)}\right)^2\\
&-2b(1-2b)\,\frac{q(1-b)q(1-2b)}{(\Delta Q(b))^2} - 2b(1-2b)\,\frac{q(1-b)q(1-2b)}{\Delta Q(b)\Delta Q(2b)}\\
&+2b(1-4b)\,\frac{q(1-b)q(1-4b)}{\Delta Q(b)\Delta Q(2b)} + 4b(1-2b)\,\frac{(q(1-2b))^2}{\Delta Q(b)\Delta Q(2b)}\\
&-4b(1-4b)\,\frac{q(1-2b)q(1-4b)}{\Delta Q(b)\Delta Q(2b)} - 4b(1-4b)\,\frac{q(1-2b)q(1-4b)}{(\Delta Q(2b))^2}\Bigg\}.\end{aligned} \tag{19}$$

Also, $\lim_{b\to 0} B_{\widehat{\xi}}^{(P)} = 0$ *and when* $P \in \mathcal{L}_c$*, we have*

$$\lim_{b\to 0} b\,\mathrm{Var}[V] = \frac{1}{(\log 2)^2}\,\frac{\xi^2\left(2^{(2\xi+1)}+1\right)}{4\left(2^{\xi}-1\right)^2}.$$

Remark 4. Again, the asymptotic bias is clearly laid out by Theorem 3, and the small-b asymptotic theory is corroborated by our $b \to 0$ results (see Fig. 1(c)). The limiting variance expression in this case matches the classical result given in Theorem 6.4.1 of Embrechts et al. (1997).

2.4. Theory for the DEdH estimator

Finally, we analyze the DEdH estimator. A motivation for developing this estimator was an improvement to the bias, although the asymptotic variance in the small b case is actually somewhat higher. Furthermore, the estimator is consistent when ξ is negative (unlike the Hill estimator). The DEdH estimator of ξ is defined by

$$\widehat{\xi}^{(D)} = 1 + H_{(1)} + \frac{1}{2}\left(\frac{H_{(1)}^2}{H_{(2)}} - 1\right)^{-1} \tag{20}$$

where $H_{(1)}$ and $H_{(2)}$ are defined via (2). See Dekkers et al. (1989) for more details. As with the Hill estimator, we need to introduce additional expressions which play a role in the asymptotic bias and variance formulas. Recall that K_1 and K_2 is defined via (4). Now, we define $\overline{K\ell}$ to be a function of the slowly-varying function K given by:

$$\overline{K\ell} = b^{-1}\int_0^b \log(u/b)\,\log\{K(u)/K(b)\}\,du. \tag{21}$$

We find that the asymptotic means of $H_{(1)}$ and $H_{(2)}$ are τ_1 and τ_2, given in Eqs. (5), but the asymptotic variance of the DEdH depends on τ_3 and τ_4 as well in a complex manner. Now we can state our main result.

Theorem 4. *Suppose that the right quantile function $\overline{Q}$ has a bounded derivative on some compact interval $[\epsilon, 1/2]$ for a small $\epsilon > 0$ and with $P \in \mathcal{L}_d$. Assume that the tuning fraction $b \leq 1/2$ is fixed as well. Then, the DEdH estimator, computed from an i.i.d. sample with quantile function (1), has the following asymptotic behavior as $n \to \infty$. The asymptotic bias of the DEdH estimator is*

$$B_{\widehat{\xi}}^{(D)} = \overline{K}_1 + \frac{2\overline{K}_1^2 - \overline{K}_2 + 2\xi\left(\overline{K\ell} + 2\overline{K}_1\right)}{2\left\{-\xi^2 + 2\xi\left(\overline{K\ell} + \overline{K}_1\right) + \overline{K}_1^2 - \overline{K}_2\right\}}. \tag{22}$$

Let $\alpha_1 = 1 - \tau_1\tau_2(\tau_2 - \tau_1^2)^{-2}$ and $\alpha_2 = 0.5\tau_1^2(\tau_2 - \tau_1^2)^{-2}$. The bias-corrected estimator satisfies

$$\sqrt{n}\left(\widehat{\xi}^{(D)} - \xi - B_{\widehat{\xi}^{(D)}}\right) \overset{\mathcal{L}}{\Longrightarrow} \alpha_1 U_1 + \alpha_2 U_2$$

with U_1 defined in Theorem 1, and U_2 given by

$$U_2 = 2b^{-1}\int_0^b \frac{q(1-u)}{Q(1-u)}\,\{\log Q(1-u) - \log Q(1-b)\}\,B(1-u)\,du - 2b^{-1}\tau_1\,\lambda B(1-b).$$

The limit $\alpha_1 U_1 + \alpha_2 U_2$ is Gaussian with mean zero and variance

$$b^{-1}\left\{\alpha_1^2(\tau_2 - \tau_1^2) + 2\alpha_1\alpha_2(\tau_3 - \tau_1\tau_2) + \alpha_2^2(\tau_4 - \tau_2^2)\right\} + (b^{-1} - 1)\,\{\alpha_1(\lambda - \tau_1) + \alpha_2(2\lambda\tau_1 - \tau_2)\}^2. \tag{23}$$

Also, $\lim_{b\to 0} B_{\widehat{\xi}}^{(D)} = 0$ and when $P \in \mathcal{L}_c$, we have $\lim_{b\to 0} b\,\mathrm{Var}[\alpha_1 U_1 + \alpha_2 U_2] = 1 + \xi^2$.

The asymptotic bias and variance again agree with the small tuning fraction results—it is known from Dekkers et al. (1989) that the asymptotic variance behaves like $(1+\xi^2)/b$ when a vanishing tuning parameter fraction is utilized (see Fig. 1(d)). Also, we can give a limiting expression for $\sqrt{b}\,(\alpha_1 U_1 + \alpha_2 U_2)$ as we did for the Hill estimator. First, one can show that

$$\sqrt{b}\,U_2 = o_P(1) - 2\xi^2\int_0^1 \frac{\log u}{u}A(u)\,du$$

as $b \to 0$, where A is the same Brownian Bridge process as in the derivation of the limiting U_1 variable. Then $\sqrt{b}$ multiplied by the limiting distribution of the DEdH statistic as $b \to 0$ equals

$$(1 - 2/\xi)\xi\int_0^1 u^{-1}A(u)\,du + \frac{1}{2\xi^2}\,(-2\xi^2)\int_0^1 \frac{\log u}{u}\,A(u)\,du = \int_0^1 \frac{\xi - 2 - \log u}{u}\,A(u)\,du.$$

This has variance $1+\xi^2$, as shown in Theorem 2 of McElroy and Jach (2012).

3. Numerical calculation of asymptotic bias and variance

We explore the theoretical results in this section for five heavy-tailed distributions: Cauchy, Burr, log-gamma, Pareto, and t (see Table 1). Note that the function L (given by $F(x) = x^{-\alpha}L(x)$) in each case is asymptotic to a constant, except in the case of the log-gamma distribution. For the latter case, it can be shown that $K(u) \propto (-\log u)^{(\beta-1)/\alpha}$ given that $L(x) \propto (\log x)^{\beta-1}$, and hence the log-gamma distribution belongs to the class $\mathcal{L}_c$.

In Section 3.1, we compare the bias and variance of the estimates from simulated data with the asymptotic results derived in the previous section. Note that for all four estimators, the asymptotic bias and variance can be computed from knowing K for any specific distribution. We examine the problem of choosing an optimal sample fraction b^*, selected by minimizing the asymptotic mean squared error in Section 3.2. Finally, in Section 3.3, we discuss possible applications of our work.

We will work with the asymptotic variances that would arise if we took rate $\sqrt{k}$ instead of $\sqrt{n}$ in the convergence results. This means we must multiply the variances in (8), (16), (19), and (23) by b, resulting in:

$$\mathrm{Var}_H = (\tau_2 - \tau_1^2) + (1-b)(\lambda - \tau_1)^2, \tag{24}$$

$$\mathrm{Var}_{HM} = \left((1-b)(bg(b) + c\upsilon_1)^2 + c^2(\upsilon_2 - \upsilon_1^2)\right)\ (1-\beta)^{-2}\langle G\rangle^{-2}, \tag{25}$$

$$\begin{aligned}
\mathrm{Var}_P \;=\; & b(\log 2)^{-2}\left\{ b(1-b)\left(\frac{q(1-b)}{\Delta Q(b)}\right)^2 + 2b(1-2b)\left(\frac{q(1-2b)}{\Delta Q(b)}\right)^2 \right.\\
& + 2b(1-2b)\left(\frac{q(1-2b)}{\Delta Q(2b)}\right)^2 + 4b(1-4b)\left(\frac{q(1-4b)}{\Delta Q(2b)}\right)^2 \\
& - 2b(1-2b)\,\frac{q(1-b)q(1-2b)}{(\Delta Q(b))^2} - 2b(1-2b)\,\frac{q(1-b)q(1-2b)}{\Delta Q(b)\Delta Q(2b)} \\
& + 2b(1-4b)\,\frac{q(1-b)q(1-4b)}{\Delta Q(b)\Delta Q(2b)} + 4b(1-2b)\,\frac{(q(1-2b))^2}{\Delta Q(b)\Delta Q(2b)} \\
& \left. - 4b(1-4b)\,\frac{q(1-2b)q(1-4b)}{\Delta Q(b)\Delta Q(2b)} - 4b(1-4b)\,\frac{q(1-2b)q(1-4b)}{(\Delta Q(2b))^2} \right\},
\end{aligned} \tag{26}$$

$$\mathrm{Var}_D = \left\{\alpha_1^2(\tau_2 - \tau_1^2) + 2\alpha_1\alpha_2(\tau_3 - \tau_1\tau_2) + \alpha_2^2(\tau_4 - \tau_2^2)\right\} + (1-b)\ \{\alpha_1(\lambda - \tau_1) + \alpha_2(2\lambda\tau_1 - \tau_2)\}^2. \tag{27}$$

3.1. Simulation study

In our simulation study, we examine the average bias and variance generated by the four methods of estimating $\xi = 1/\alpha$ for different sample sizes (from 100 to 10,000) and distributions. A total of 2000 data sets were simulated from each distribution and sample size combination. We set the sample fraction $b = 0.15$ for all cases. (We will look more at the selection of b in the next section.) These numerically computed values were then compared with the asymptotic bias and variances (for variances, we use (24) through (27)).

The results for the Burr, Cauchy, and t distributions are graphed in Fig. 2; results for the log-gamma and Pareto distributions are plotted in Fig. 3. In the first column of graphs, the numerical bias is plotted for the five estimators: Hill, Harmonic Moment ($\beta = 0.8$ and $\beta = 2$), Pickands, and DEdH. In order to compare the numerical value to the asymptotic results, the asymptotic bias is denoted by the dotted lines for each estimator. The second column of graphs follows the same format but plots the results for variances instead. Here, the Monte Carlo variability is evident.

We find that even for small sample sizes, the asymptotic variance is well approximated, especially for the Hill, Harmonic Moment, and Pickands estimators. Capturing the asymptotic bias accurately, however, seems to require larger sample sizes, particularly for the DEdH estimator (although that bias is roughly zero for DEdH). However, even with this error, it is superior to assuming, erroneously, that the bias is zero, which is the classical approach. For all six distributions, the Pickands estimator had the highest asymptotic variance followed by the DEdH. However, the bias simulations showed no such pattern consistently across distributions. The Harmonic Moment and Hill estimator results tended to be similar (not necessarily zero, though). The Pickands has the worst bias especially for the Cauchy, $t(2)$, and Burr distributions.

3.2. Optimal sample fraction b^*

Given a heavy-tailed distribution such as one listed in Table 1, we can numerically compute the sample fraction b which minimizes the asymptotic mean squared error (asymptotic MSE). We follow the suggestion of De Haan and Peng (1998) so that the optimal b for a given sample size n, denoted by $b^*(n)$, is computed by minimizing $\mathrm{Var}/(bn) + \mathrm{Bias}^2$. Here, "Bias" is (7), (15), (18), or (22), while "Var" is (24), (25), (26), or (27) for the Hill, Harmonic Moment, Pickands, and DEdH estimators respectively. Fig. 4 shows the pattern of optimal $b^*(n)$ for sample sizes from 100 to 10,000 for four distributions: Burr with $c = 2$ and $\nu = 2$; Cauchy with $x_0 = 0$ and $\gamma = 1$; log-gamma with $\gamma = 2$ and $\nu = 1$; and t with $\nu = 3$.

The optimal fraction generally decreases as the sample size increases because the squared bias becomes the primary component of the asymptotic MSE, which is not surprising. Furthermore, as expected with tail index estimators, using a larger number of observations in the calculations (i.e., increasing b) decreases the variance of the estimate but, unfortunately, also increases the bias. Dekkers and de Haan (1993) and De Haan and Peng (1998) note correctly that determining the optimal sample fraction b is difficult when the distribution is unknown.

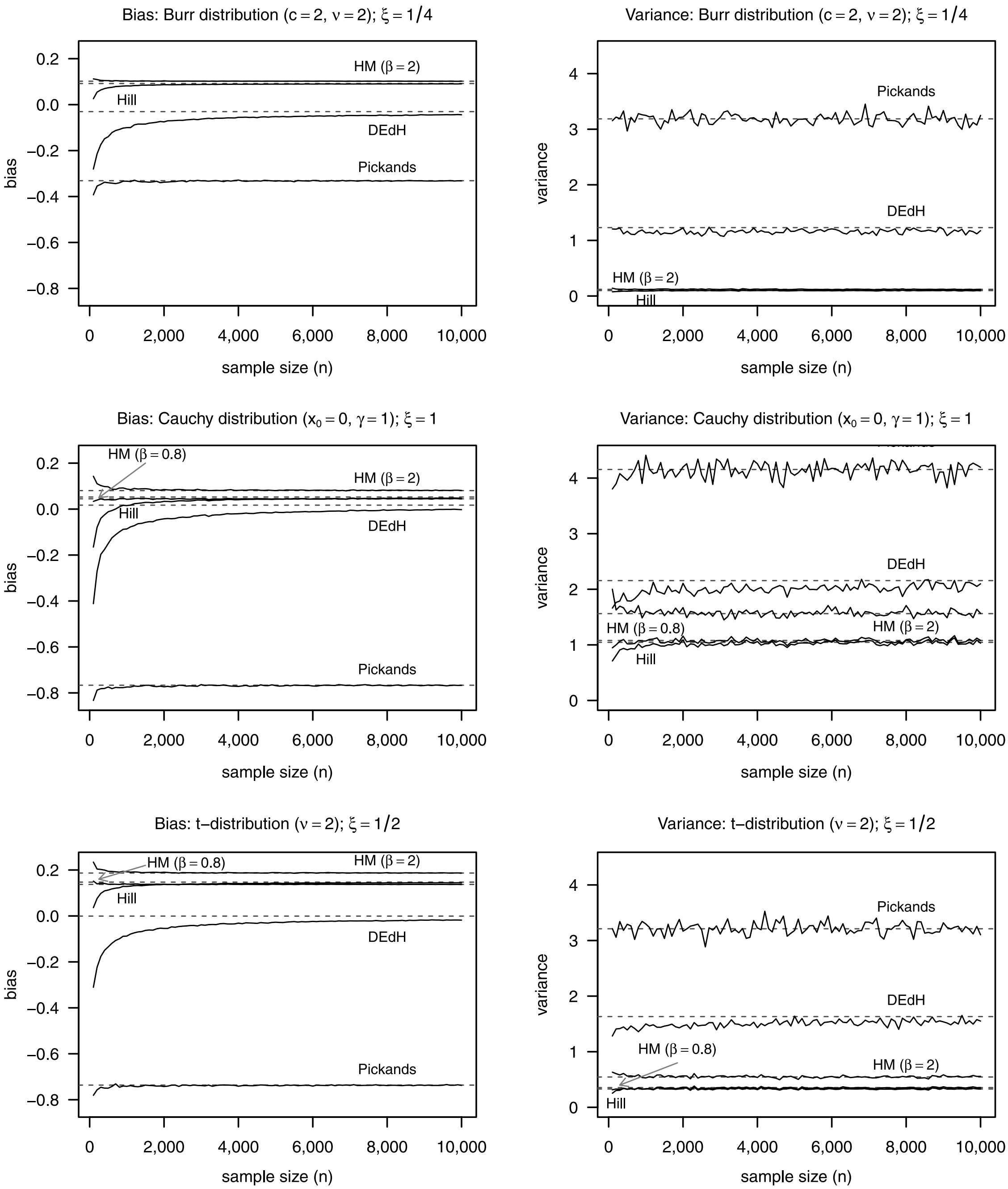


Fig. 2. Numerical bias and variance of ξ estimators (solid lines) for data simulated from the Burr, Cauchy, and $t(2)$ distributions when $b = 0.15$ with corresponding asymptotic values (dotted lines). Note: The Harmonic Moment estimator when $\beta = 0.8$ is omitted from the Burr distribution simulations because it does not satisfy the requirement that $2(1-\beta) < \alpha$.

3.3. Bias estimation and reduction

In order to make a penultimate approximation via the results of this paper, we need to estimate $B(b)$ and $\mathrm{Var}(b)$. For the case of the Hill estimator, we propose a method suggested by the expansion used in Holan and McElroy (2010). Suppose $\log K(u)$ for $u \in (0, 1/2)$ is expanded in a Fourier basis such that

$$K(u) = \exp\left\{ \sum_{\ell \geq 0} \eta_\ell \cos(4\pi \ell u) \right\} \tag{28}$$

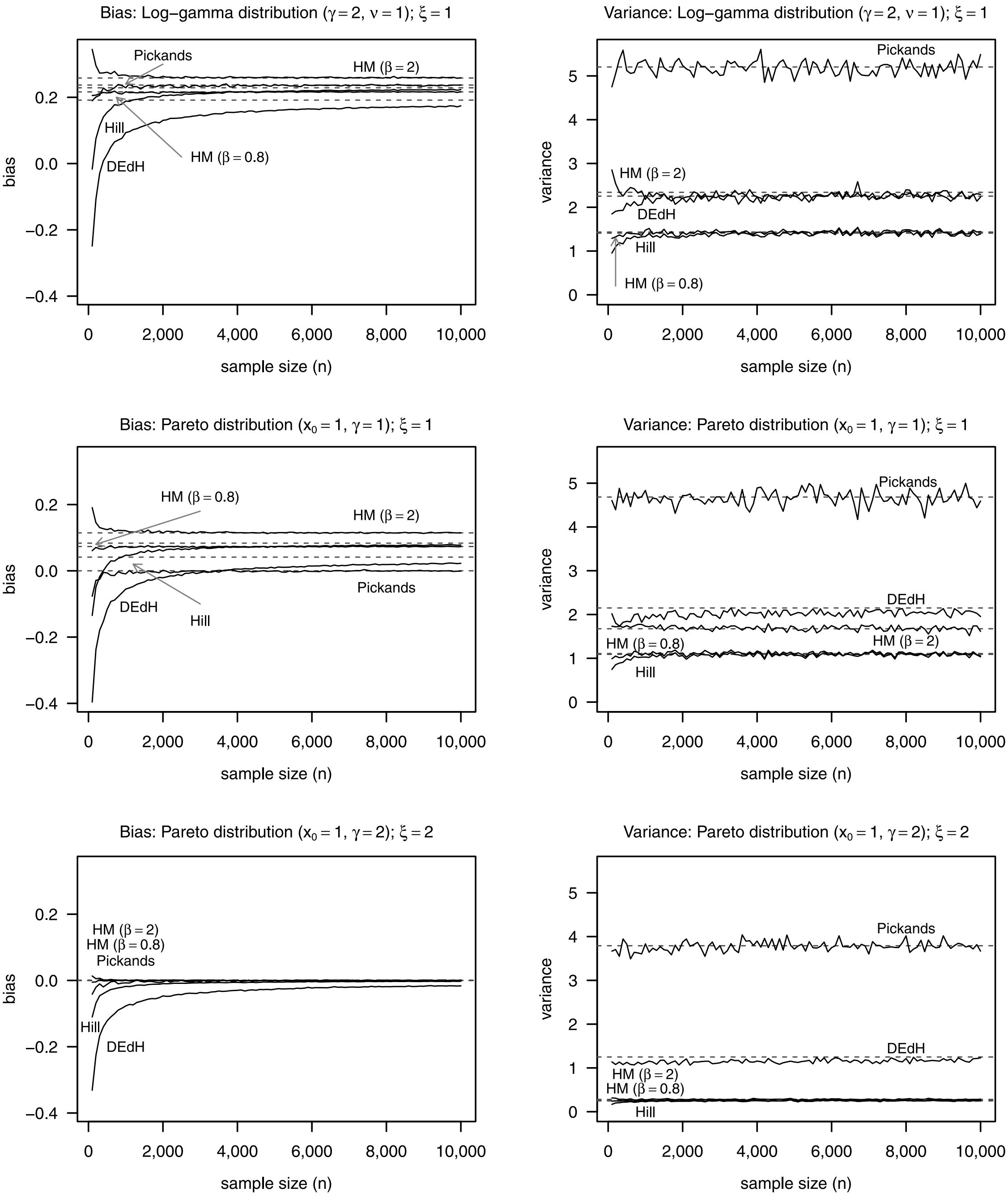


Fig. 3. Numerical bias and variance of ξ estimators (solid lines) for data simulated from the log-gamma and two Pareto distributions when $b = 0.15$ with corresponding asymptotic values (dotted lines).

where $\{\eta_\ell\}$ are real coefficients. Then,

$$\overline{K}_1 = \sum_{\ell \geq 1} \eta_\ell \left(\frac{\sin(4\pi \ell b)}{4\pi \ell b} - \cos(4\pi \ell b) \right) \tag{29}$$

is the bias, and it can be estimated if we know the η_ℓ coefficients. Set $x_\ell(b) = \sin(4\pi \ell b)/(4\pi \ell b) - \cos(4\pi \ell b)$ for $\ell \geq 1$ and $x_0(b) = 1$, so that $B_{\widehat{\xi}}^{(H)} = \sum_{\ell \geq 1} \eta_\ell x_\ell(b)$. The asymptotic mean of $\widehat{\xi}^{(H)}$ is $\xi + B_{\widehat{\xi}}^{(H)}$, which suggests regressing the estimator on $x_0(b), x_1(b), \ldots$ for various choices of b (possibly using polynomial powers of b). The errors in this regression are correlated, suggesting the use of GLS, but we might still use OLS for simplicity. In such a regression, the target is the coefficient of $x_0(b)$,

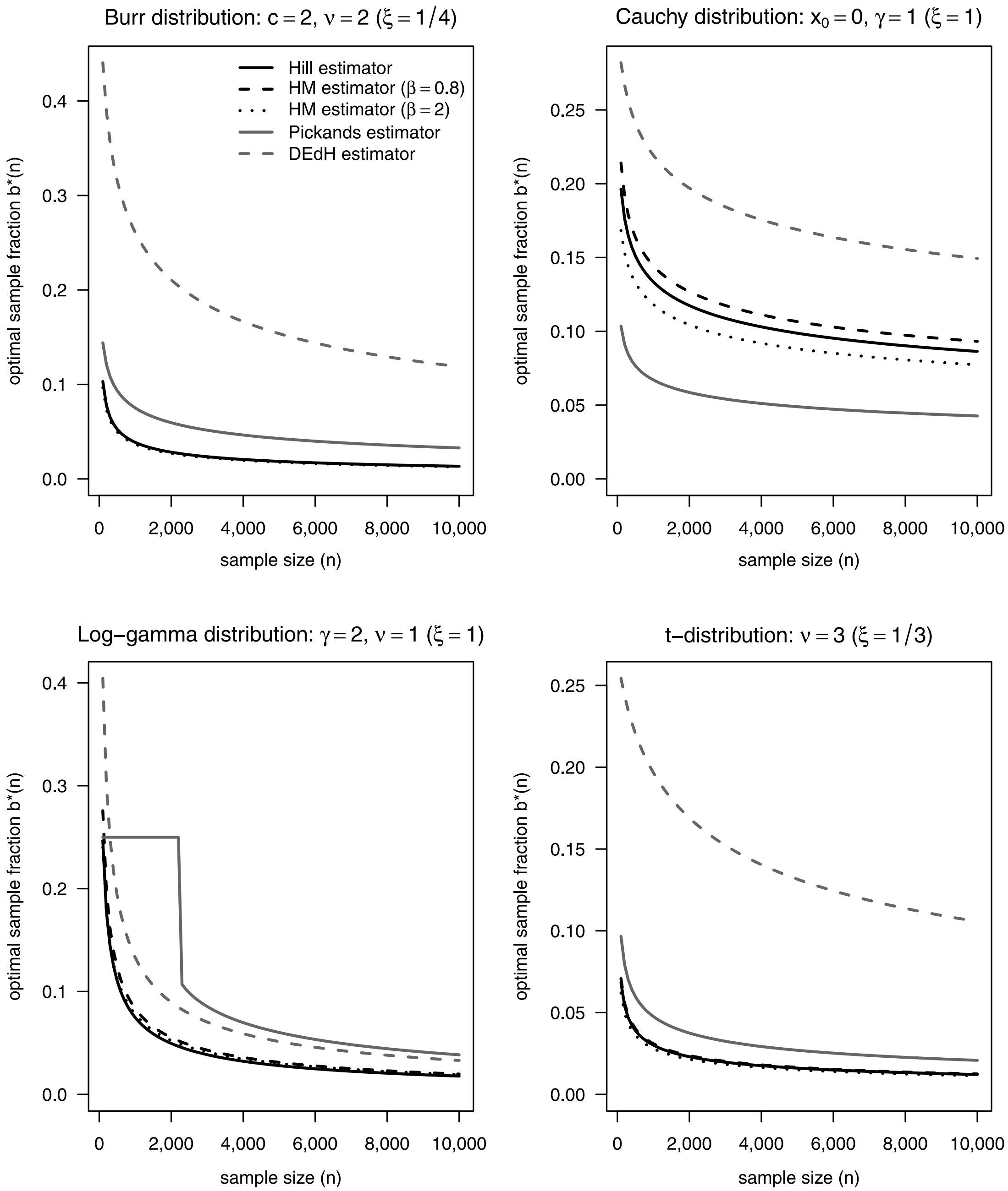


Fig. 4. Sample fraction $b^*(n)$ which minimizes the asymptotic MSE; Hill estimator (solid), Harmonic Moment with $\beta = 0.8$ (dashed), Harmonic Moment with $\beta = 2$ (dotted), Pickands (solid, gray), and DEdH (dashed, gray) for various distributions. Note: the Harmonic Moment estimator with $\beta = 0.8$ is omitted from the Burr distribution analysis as it does not satisfy the $2(1-\beta) < \alpha$ condition.

which is ξ. This is a modification of the method proposed by Huisman et al. (2001) who use values of b as ad hoc linear regressors.

4. Conclusion

We contend here that developing the asymptotic theory for tail index estimators such as the Hill, Harmonic Moment, Pickands, and DEdH – from the perspective that the tuning parameter is a fixed fraction b of sample size – permits one to better understand the actual bias and variance of finite-sample estimators. The asymptotic theory of this new approach agrees well with conventional results that adopt a vanishing tuning parameter fraction (i.e., b vanishing); this is demonstrated through our simulation studies.

The problem of conducting inference for a parameter ultimately revolves around the issue of describing the finite-sample distribution of its estimator, and asymptotic theory is only as valuable as its ability to approximate this finite-sample distribution. The classical theory assumes that b is vanishing, which results in a consistent estimator; unfortunately, the asymptotic bias (which is zero) and the variance stemming from this theory yield a poor approximation to the finite-sample distribution whenever the ratio of k to n is non-negligible (e.g., $b = 0.15$). Our fixed-b theory does not yield a consistent estimator, but the asymptotic bias and variance nevertheless yield a superior approximation to the finite-sample distribution because the known value of b is directly accounted for.

Unfortunately, the new limit distributions are not pivotal (unlike in the spectral density estimation literature) and the bias function is not obviously estimable; the variance, however, can be estimated through plug-in estimates of Q and q. Since one can quickly compute the tail index estimates over a range of b values, it is natural to regress these estimates on polynomial powers of b (or with regressors suggested by (29)), with the intercept corresponding to the true parameter. Future research will pursue this application further.

It seems the philosophy outlined in this paper might be useful in other areas of semi-parametric statistics wherein a suite of estimators is computed, each being dependent on a tuning parameter. Typically, such tuning parameters (perhaps the order of truncation of some basis expansion, or a bandwidth parameter, etc.) may be viewed as a quantity $k(n)$ that tends to infinity with sample size n, but also such that $k(n)/n \to 0$. If the framework of this paper is applied to such problems, one would have $k(n) = bn$ for a fixed fraction $b \in (0, 1)$, which would result in statistical properties such as a non-vanishing bias.

Disclaimer

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Appendix

Proof of Theorem 1 (*Hill Estimator*)**.** Let U denote the (connected) compact interval $[\epsilon, 1/2]$ for which $\overline{Q}$ has a bounded derivative. By an abuse of notation, we write the lower limit of the integrals below as zero, instead of the boundary implied by the closed interval U (see the comments at the end of the proof for more details). We first note that the empirical quantile function is defined continuously such that $\widehat{Q}(u)$ is the lth order statistic, where $l = \lfloor nu \rfloor$. Thus

$$\widehat{\xi}^{(H)} = \frac{n}{k} \int_0^{k/n} \log \widehat{Q}(1-x)\, dx - \log \widehat{Q}(1-k/n),$$

so that up to negligible error we obtain $b^{-1} \int_0^b \log \widehat{Q}(1-x)\, dx - \log \widehat{Q}(1-b)$. Let us consider the same expression, but with Q in place of $\widehat{Q}$:

$$\tau_1 = \frac{1}{b} \int_0^b \log Q(1-x)\, dx - \log Q(1-b) = \xi - \left(\log K(b) - \frac{1}{b} \int_0^b \log K(x)\, dx \right), \tag{A.1}$$

using the fact that $\log b - b^{-1} \int_0^b \log x dx = 1$. With $T_n = \widehat{\xi}^{(H)} - \tau_1$ (which is just the functional $b^{-1} \int_0^b dx - \Delta_{1-b}$ applied to $\log \widehat{Q}(1-x) - \log Q(1-x)$ with Δ_{1-b} the Dirac delta function at $1-b$) we have $\sqrt{n}(\widehat{\xi}^{(H)} - \xi) = \sqrt{n}(\tau_1 - \xi) + \sqrt{n}\, T_n$. Hence $\tau_1 - \xi = \overline{K}_1$ is the asymptotic bias, denoted $B_{\widehat{\xi}^{(H)}}$ in Theorem 1.

The bias can be further analyzed as follows: by Theorem A3.3 of Embrechts et al. (1997), we can write $K(u) = c(1/u) \exp\{\int_a^{1/u} \delta(s)/s\, ds\}$ for some $a > 0$, and both $c(x) \to c_0 \in (0, \infty)$ and $\delta(x) \to 0$ as $x \to \infty$. Then,

$$\log K(b) - \frac{1}{b} \int_0^b \log K(x)\, dx = \log c(1/b) - b^{-1} \int_0^b \log c(1/x)\, dx + b^{-1} \int_0^b \int_{b^{-1}}^{x^{-1}} \delta(s) s^{-1}\, ds\, dx$$

and the first two terms clearly tend to zero as $b \to 0$. The third term can be handled as follows. For any given $\epsilon > 0$, let b be so small such that $|\delta(s)| < \epsilon$ for all $s > b^{-1}$. Then

$$\left| \int_0^b \int_{b^{-1}}^{x^{-1}} \delta(s) s^{-1}\, ds\, dx \right| \leq \int_0^b \epsilon (\log b - \log x)\, dx = \epsilon b,$$

which shows the second term of the bias tends to zero as $b \to 0$.

Next, we establish the convergence of $\sqrt{n}\, T_n$. We can apply Theorem 4 of Holan and McElroy (2010) with a trivial kernel to deduce

$$\sqrt{n}\, \left(\widehat{Q}(\cdot) - Q(\cdot) \right) \stackrel{\mathcal{L}}{\Longrightarrow} q \cdot B \tag{A.2}$$

as $n \to \infty$, with the convergence taking place on the space of continuous functions on U. This result is also established in Gihman and Skorohod (1980, p. 437). Now,

$$\sqrt{n}\left(\log \widehat{Q}(u) - \log Q(u)\right) = \sqrt{n} \log\left(1 + \left[\widehat{Q}(u) - Q(u)\right] Q^{-1}(u)\right)$$

which is $o_P(1)$ (for each u) plus $\sqrt{n}(\widehat{Q}(u) - Q(u))Q^{-1}(u)$. Hence, to get the limit of $\sqrt{n}\,T_n$, we need to apply the integration functional $b^{-1}\int_0^b \cdot\, Q^{-1}(x)\,dx$, as well as $\Delta_{1-b}Q^{-1}(1-b)$ to (A.2), and obtain the convergence

$$\sqrt{n}\,T_n \overset{\mathcal{L}}{\Longrightarrow} \frac{1}{b}\int_0^b \frac{q(1-u)}{Q(1-u)} B(1-u)\,du - \frac{q(1-b)}{Q(1-b)} B(1-b).$$

A change of variable now produces the stated result. To compute the variance of the Gaussian limit, we use known properties of Brownian Bridge and simplify. For the limit of $b\mathrm{Var}[U_1]$, observe that $\lambda - \tau_1 = \overline{K}_1 - R(b)$ tends to zero. Also, we have the relations,

$$\tau_1 = \xi + \overline{K}_1 \tag{A.3}$$

$$\tau_2 = 2\xi^2 - 2\xi\overline{K\ell} + \overline{K}_2. \tag{A.4}$$

Hence, $\tau_2 - \tau_1^2 = \xi^2 - 2\xi\overline{K\ell} + \overline{K}_2 - \overline{K}_1^2$ which tends to ξ^2 as $b \to 0$.

To account for the lower bound of U explicitly, we must instead consider the statistic

$$\widehat{\xi}_\epsilon^{(H)} = \frac{1}{b-\epsilon}\int_\epsilon^b \log \widehat{Q}(1-u)\,du - \log \widehat{Q}(1-b)$$

and make this modification throughout the proof. So $\overline{K}_1 = (b-\epsilon)^{-1}\int_\epsilon^b \log(K(u)/K(b))\,du$, and $\tau_1 = \xi - \overline{K}_1$; also,

$$\sqrt{n}\,T_n \overset{\mathcal{L}}{\Longrightarrow} \frac{1}{b-\epsilon}\int_\epsilon^b \frac{q(1-u)}{Q(1-u)} B(1-u)\,du - \frac{q(1-b)}{Q(1-b)} B(1-b).$$

Note that ϵ functions as a second tuning parameter, similar to b. We do not prove a limit theory for $\widehat{\xi}^{(H)} = \widehat{\xi}_0^{(H)}$, but rather for $\widehat{\xi}_\epsilon^{(H)}$; when $\epsilon < 1/n$, for fixed n we have $\widehat{\xi}_0^{(H)} = \widehat{\xi}_\epsilon^{(H)}$. Thus, we shall apply the limit results for $\widehat{\xi}_\epsilon^{(H)}$ to $\widehat{\xi}^{(H)}$ with this understanding. Finally, because the limiting bias and variance is continuous in ϵ, and because ϵ can be made arbitrarily small (e.g., $\epsilon = 10^{-10}$), we replace ϵ by zero in the bias and variance expressions—this abuse of notation has no practical cost when $\epsilon = 10^{-10}$, and makes for easier reading of the results. These comments pertain to the Pickands and DEdH estimators as well. □

Derivation of (10). Letting $\overline{Q}(u) = Q(1-u)$, we see it is the inverse of $\overline{F}(x)$. Then, by change of variable we have

$$\begin{aligned} b^{-1}\int_0^b \log \overline{Q}(u)\,du &= b^{-1}\int_{\overline{Q}(0)}^{\overline{Q}(b)} \log x\, d\overline{F}(x) \\ &= b^{-1}\left(\log x\,\overline{F}(x)\Big|_{\overline{Q}(0)}^{\overline{Q}(b)} - \int_{\overline{Q}(0)}^{\overline{Q}(b)} x^{-1}\,\overline{F}(x)\,dx\right) \\ &= b^{-1}\left(b\,\log \overline{Q}(b) + \int_{\overline{Q}(b)}^{\infty} a\,x^{-(\alpha+1)}\,(1 + cx^{-\beta})\,dx\right) \\ &= \log \overline{Q}(b) + \frac{a}{b}\,\overline{Q}^{-\alpha}(b)\left(\frac{1}{\alpha} + \frac{c\,\overline{Q}^{-\beta}(b)}{\alpha+\beta}\right), \end{aligned}$$

which uses the fact that $\overline{Q}(0) = \infty$ and the stated form of $L(x)$. Hence

$$\overline{K}_1 = \frac{a}{b}\,\overline{Q}^{-\alpha}(b)\left(\frac{1}{\alpha} + \frac{c\,\overline{Q}^{-\beta}(b)}{\alpha+\beta}\right) - \frac{1}{\alpha}$$

and can be further simplified by setting $z = \overline{Q}(b)$. In this case, $b = \overline{F}(z) = az^{-\alpha}(1+cz^{-\beta})$ which can be substituted into $\overline{K}_1$:

$$\begin{aligned} \overline{K}_1 &= \frac{a\,z^{-\alpha}}{az^{-\alpha}(1+cz^{-\beta})}\left(\frac{1}{\alpha} + \frac{c\,z^{-\beta}}{\alpha+\beta}\right) - \frac{1}{\alpha} \\ &= -\frac{\beta\,c\,z^{-\beta}}{\alpha(\alpha+\beta)(1+cz^{-\beta})} \\ &= -\frac{\beta\,c\,a\,z^{-(\alpha+\beta)}}{\alpha(\alpha+\beta)\,b}. \end{aligned}$$

Finally, substituting $z = \overline{Q}(b) = b^{-1/\alpha}K(b)$ produces the stated formula for $\overline{K}_1$. □

Proof of Theorem 2 (*Harmonic Moment Estimator*)**.** Using the step function structure of the sample quantile function, the estimator can be rewritten as

$$\widehat{\xi}^{(HM)} = \frac{1}{\beta - 1} \left(b^{-1} \int_0^b \widehat{Q}(1-x)^{1-\beta}\, dx \right)^{-1} \left(\widehat{Q}(1-b)^{1-\beta} - \left(b^{-1} \int_0^b \widehat{Q}(1-x)^{1-\beta}\, dx \right) \right).$$

We also introduce the notations $\widehat{G}(x) = (\beta-1)^{-1}\widehat{Q}(1-x)^{1-\beta}$ and $\langle \widehat{G}^p \rangle = b^{-1} \int_0^b \widehat{G}(x)^p\, dx$. First, we analyze the integral of the quantile function; using integration by parts, we have

$$\begin{aligned}\int_0^b Q(1-x)^{1-\beta}\, dx &= \frac{\alpha}{\alpha+\beta-1} \left(b^{(\alpha+\beta-1)/\alpha} K(b)^{1-\beta} + (\beta-1) \int_0^b Q(1-x)^{1-\beta} R(x)\, dx \right) \\ &= \frac{\alpha}{\alpha+\beta-1}\, bQ(1-b)^{1-\beta} \left(1 + (\beta-1) \frac{\int_0^b Q(1-x)^{1-\beta} R(x)\, dx}{bQ(1-b)^{1-\beta}} \right)\end{aligned}$$

where $R(x) = x\dot{K}(x)/K(x)$. This calculation shows that

$$\frac{\int_0^b Q(1-x)^{1-\beta}\, dx}{bQ(1-b)^{1-\beta}} \to \frac{\alpha}{\alpha+\beta-1} \tag{A.5}$$

as $b \to 0$, because (using L'Hôpital's rule) we see that $\int_0^b Q(1-x)^{1-\beta} R(x)\, dx$ tends to zero faster than $bQ(1-b)^{1-\beta}$. Next, we decompose the statistic into the asymptotic mean and the mean zero stochastic portions. Then, we will apply a Functional Central Limit Theorem (FCLT), based on the result $\sqrt{n}(\widehat{Q}(u) - Q(u)) \stackrel{\mathcal{L}}{\Longrightarrow} q(u)B(u)$. Using the delta method with the function $\delta(x) = x^{1-\beta}$, we obtain the FCLT

$$\sqrt{n} \left(\widehat{Q}(u)^{1-\beta} - Q(u)^{1-\beta} \right) \stackrel{\mathcal{L}}{\Longrightarrow} (1-\beta) Q(u)^{-\beta} q(u) B(u) \tag{A.6}$$

when $\beta \neq 1$. The statistic's decomposition is as follows: $(\beta-1)\widehat{\xi}$ equals

$$\langle \widehat{G} \rangle^{-1} \left(\widehat{G}(b) - G(b) - \langle \widehat{G} \rangle + \langle G \rangle \right) + \left(\langle \widehat{G} \rangle^{-1} - \langle G \rangle^{-1} \right) (G(b) - \langle G \rangle) + \langle G \rangle^{-1} (G(b) - \langle G \rangle).$$

The third term is deterministic and constitutes the mean, because the first and second terms are asymptotically mean zero. We analyze this mean first; it is written $G(b)/\langle G \rangle - 1$, so that the bias is

$$(\beta-1)^{-1} \left(\frac{G(b)}{\langle G \rangle} - 1 \right) - \xi = (\beta-1)^{-1} \left(\frac{bQ(1-b)^{1-\beta}}{\int_0^b Q(1-x)^{1-\beta}\, dx} - 1 \right) - \xi$$

which yields (15). This bias tends to zero as b shrinks, which is seen as follows. Note that (A.5) implies that $G(b)/\langle G \rangle \to (\alpha+\beta-1)/\alpha$ as $b \to 0$. Plugging this in, we have

$$\lim_{b \to 0} B_{\widehat{\xi}}^{(HM)} = (\beta-1)^{-1} \left(\frac{\alpha+\beta-1}{\alpha} - 1 \right) - 1/\alpha = 0.$$

Next, we consider the stochastic terms. We have

$$\langle \widehat{G} \rangle^{-1} - \langle G \rangle^{-1} = -\frac{\langle \widehat{G} \rangle - \langle G \rangle}{\langle \widehat{G} \rangle \cdot \langle G \rangle} = o_P(n^{-1/2}) - \frac{\langle \widehat{G} \rangle - \langle G \rangle}{\langle G \rangle^2}.$$

This can be combined asymptotically with the first term (replacing $\langle \widehat{G} \rangle$ by $\langle G \rangle$ in the denominator) to obtain

$$\sqrt{n} \left(\widehat{\xi}^{(HM)} - \xi - B_{\widehat{\xi}}^{(HM)} \right) = o_P(1) + \frac{\sqrt{n}}{\beta-1} \langle G \rangle^{-1} \left([\widehat{G}(b) - G(b)] - (G(b)/\langle G \rangle) \left[\langle \widehat{G} \rangle - \langle G \rangle \right] \right).$$

So the limit in the FCLT (A.6) is $(1-\beta)g(1-u)B(u)$. Consequently, the FCLT stated in terms of $\widehat{G}$ will divide (A.6) by $\beta - 1$, so that

$$\sqrt{n} \left(\widehat{\xi}^{(HM)} - \xi - B_{\widehat{\xi}}^{(HM)} \right) \stackrel{\mathcal{L}}{\Longrightarrow} \frac{1}{1-\beta} \langle G \rangle^{-1} \left(g(b)B(1-b) - c\, b^{-1} \int_0^b g(x)\, B(1-x)\, dx \right),$$

where $c = G(b)/\langle G\rangle$. Next, we calculate the limiting variance. Using known properties of Brownian Bridge, we obtain the variance of the expression in parentheses is equal to

$$\left(b^{-2}\int_0^b\int_0^x g(x)g(y)(1-x)y\,dy\,dx + b^{-2}\int_0^b\int_x^b g(x)g(y)(1-y)x\,dy\,dx\right)c^2$$
$$-2cg(b)(1-b)b^{-1}\int_0^b g(x)x\,dx + g^2(b)(1-b)b$$

and the first term can be simplified to

$$\left(b^{-2}\int_0^b g(x)\int_0^x g(y)y\,dy\,dx + b^{-2}\int_0^b xg(x)\int_x^b g(y)\,dy\,dx - \left[\int_0^b g(x)x\,dx\right]^2\right)c^2$$
$$= \left(b(G(b)^2+\langle G^2\rangle) - 2bG(b)\langle G\rangle - b^2(G(b)-\langle G\rangle)^2\right)c^2$$

using $\int_0^b xg(x)\,dx = b(G(b)-\langle G\rangle)$. Then with the definitions of υ_1 and υ_2 given in (13), the whole variance simplifies to the expression stated in (16). Next, we show that b times the limiting variance tends, as $b \to 0$, to the classical quantity derived in Beran et al. (2014), namely $\xi^2(1+(\beta-1)\xi)^2/(1+2(\beta-1)\xi)$. First, we have

$$\frac{bg(b)+c\upsilon_1}{\langle G\rangle} = \frac{G(b)}{\langle G\rangle}\left(1-\frac{G(b)}{\langle G\rangle}+(\beta-1)(\xi-R(b))\right)$$

utilizing (14). The limit of the expression in parentheses is zero, and hence

$$\lim_{b\to 0} b(b^{-1}-1)(bg(b)+c\upsilon_1)^2(1-\beta)^{-2}\langle G\rangle^{-2} = \lim_{b\to 0}(1-b)(1-\beta)^{-2}\left(\frac{bg(b)+c\upsilon_1}{\langle G\rangle}\right)^2 = 0.$$

The other term contributing to the variance is

$$\frac{c^2(\upsilon_2-\upsilon_1^2)}{(1-\beta)^2\langle G\rangle^2} = c^2(1-\beta)^{-2}\left(\frac{\langle G^2\rangle}{G(b)^2}-1\right).$$

Using L'Hôpital again, similar to (A.5), we obtain

$$\frac{\int_0^b G(x)^2\,dx}{bG(b)^2} \to \frac{\alpha}{\alpha+2(\beta-1)}.$$

Using this,

$$\frac{\langle G^2\rangle}{G(b)^2} \to \frac{(\alpha+\beta-1)^2}{\alpha(\alpha+2(\beta-1))}.$$

Since $c \to (\alpha+\beta-1)/\alpha$, we obtain the desired result. □

Proof of Theorem 3 (*Pickands Estimator*). From the known weak convergence $\sqrt{n}\,(\widehat{Q}(u)-Q(u)) \stackrel{\mathcal{L}}{\Longrightarrow} q(u)B(u)$, we have

$$\sqrt{n}\left(\widehat{Q}(1-b)-\widehat{Q}(1-2b)-Q(1-b)+Q(1-2b)\right) \stackrel{\mathcal{L}}{\Longrightarrow} \Delta qB(b),$$
$$\sqrt{n}\left(\widehat{Q}(1-2b)-\widehat{Q}(1-4b)-Q(1-2b)+Q(1-4b)\right) \stackrel{\mathcal{L}}{\Longrightarrow} \Delta qB(2b)$$

as $n \to \infty$. From this it follows, with the Taylor expansion in probability of the logarithm, that

$$\widehat{\xi}^{(P)}\cdot\log 2 = o_P(1) + \log\left(\frac{Q(1-b)-Q(1-2b)}{Q(1-2b)-Q(1-4b)}\right) + \frac{\widehat{Q}(1-b)-\widehat{Q}(1-2b)-Q(1-b)+Q(1-2b)}{Q(1-b)-Q(1-2b)}$$
$$-\frac{\widehat{Q}(1-2b)-\widehat{Q}(1-4b)-Q(1-2b)+Q(1-4b)}{Q(1-2b)-Q(1-4b)}.$$

For the deterministic term, we use (1) to obtain

$$Q(1-b)-Q(1-2b) = b^{-\xi}\left(K(b)-2^{-\xi}K(2b)\right),$$
$$Q(1-2b)-Q(1-4b) = b^{-\xi}\left(2^{-\xi}K(2b)-4^{-\xi}K(4b)\right),$$

so that the log ratio equals $\xi\log 2$ plus the log of the term $[K(b)-2^{-\xi}K(2b)]/[K(2b)-2^{-\xi}K(4b)]$. Then, it follows that the asymptotic bias is equal to (18). For the remaining stochastic terms in the expansion of $\widehat{\xi}^{(P)}$, we can apply the weak convergence results above to obtain the result of the theorem. To compute the variance (19), note that $B(x)$ appears four times so that the variance involves sixteen terms; expanding and simplifying reduces the expression to the stated ten terms.

The limit of the bias is zero, which is seen by dividing both the numerator and denominator of the logged expression in (18) by $K(2b)$ and using the basic property of $\mathcal{L}$. For the limiting variance, we use $\lim_{b\to 0} R(b) = 0$ and the following calculations:

$$\frac{q(1-b)}{\Delta Q(b)} = b^{-1}\,\frac{-\xi + b\dot{K}(b)/K(b)}{1-2^{-\xi}\,K(2b)/K(b)} \sim b^{-1}\cdot\frac{-\xi}{1-2^{-\xi}}$$
$$\frac{q(1-2b)}{\Delta Q(b)} = b^{-1}\,\frac{-\xi\,2^{-(\xi+1)} + 2^{-\xi}\,b\dot{K}(2b)/K(2b)}{K(b)/K(2b)-2^{-\xi}} \sim b^{-1}\cdot\frac{-\xi\,2^{-(\xi+1)}}{1-2^{-\xi}}$$
$$\frac{q(1-2b)}{\Delta Q(2b)} = b^{-1}\,\frac{-\xi\,2^{-(\xi+1)} + 2^{-\xi}\,b\dot{K}(2b)/K(2b)}{2^{-\xi}-4^{-\xi}\,K(4b)/K(2b)} \sim b^{-1}\cdot\frac{-\xi\,2^{-(\xi+1)}}{2^{-\xi}-4^{-\xi}}$$
$$\frac{q(1-4b)}{\Delta Q(2b)} = b^{-1}\,\frac{-\xi\,4^{-(\xi+1)} + 4^{-\xi}\,b\dot{K}(4b)/K(4b)}{2^{-\xi}\,K(2b)/K(4b)-4^{-\xi}} \sim b^{-1}\cdot\frac{-\xi\,4^{-(\xi+1)}}{2^{-\xi}-4^{-\xi}}.$$

The variance is composed completely of these four terms, and so plugging them into (19) and letting $b\to 0$ produces the classical expression for the scaled variance. $\square$

Proof of Theorem 4 (*DEdH Estimator*)**.** We begin by expanding $\widehat{\xi}^{(D)}$:

$$\begin{aligned}\widehat{\xi}^{(D)} &= 1 + H_{(1)} + \frac{H_{(2)}}{2\left(H_{(1)}^2 - H_{(2)}\right)} \\ &= \left(\tau_1 + \frac{2\tau_1^2-\tau_2}{2\left(\tau_1^2-\tau_2\right)}\right) + \left(H_{(1)}-\tau_1\right) + \frac{H_{(2)}-\tau_2}{2\left(H_{(1)}^2-H_{(2)}\right)} - \frac{\tau_1}{2}\,\frac{\left(H_{(1)}^2-\tau_1^2\right)-\left(H_{(2)}-\tau_2\right)}{\left(\tau_1^2-\tau_2\right)\left(H_{(1)}^2-H_{(2)}\right)}.\end{aligned}$$

This breaks the statistic down into stochastic terms that are appropriately centered, together with a leading deterministic term. Now it follows from (A.3) and (A.4) that

$$\tau_1 + \frac{2\tau_1^2-\tau_2}{2\left(\tau_1^2-\tau_2\right)} - \xi = \overline{K}_1 + \frac{2(\xi^2+2\xi\overline{K}_1+\overline{K}_1^2) - (2\xi^2 - 2\xi\overline{K\ell}+\overline{K}_2)}{2(\xi^2+2\xi\overline{K}_1+\overline{K}_1^2 - [2\xi^2-2\xi\overline{K\ell}+\overline{K}_2])} = B_{\widehat{\xi}^{(D)}}.$$

This proves (22), since the other terms in the expansion of $\widehat{\xi}^{(D)}$ are centered which is seen as follows:

$$\widehat{\xi}^{(D)} - \xi - B_{\widehat{\xi}^{(D)}} = \left(H_{(1)}-\tau_1\right)\left(1 - \frac{\tau_2}{2}\,\frac{H_{(1)}+\tau_1}{\left(\tau_1^2-\tau_2\right)\left(H_{(1)}^2-H_{(2)}\right)}\right) + \frac{H_{(2)}-\tau_2}{2\left(H_{(1)}^2-H_{(2)}\right)}\left(1+\frac{\tau_2}{\tau_1^2-\tau_2}\right).$$

Now the stated convergence result of Theorem 4 will follow from the joint weak convergence

$$\sqrt{n}\left(H_{(1)}-\tau_1, H_{(2)}-\tau_2\right) \stackrel{\mathcal{L}}{\Longrightarrow} (U_1, U_2)\,.$$

In fact, it will follow that $\sqrt{n}$ times $(\widehat{\xi}^{(D)} - \xi - B_{\widehat{\xi}^{(D)}})$ converges in distribution to $\alpha_1 U_1 + \alpha_2 U_2$, where α_1, α_2 are defined in the statement of the theorem. Focusing on the second quantile moment, we have

$$\begin{aligned}H_{(2)} - \tau_2 = b^{-1}&\int_0^b \left(\log\widehat{Q}(1-x) - \log Q(1-x)\right)\left(\log\widehat{Q}(1-x)+\log Q(1-x)\right)\,dx \\ &- 2b^{-1}\int_0^b \left(\log\widehat{Q}(1-x)-\log Q(1-x)\right)\,dx\;\log\widehat{Q}(1-b) \\ &- 2b^{-1}\int_0^b \log Q(1-x)\,dx\;\left(\log\widehat{Q}(1-b)-\log Q(1-b)\right) \\ &+ \left(\log\widehat{Q}(1-b)-\log Q(1-b)\right)\left(\log\widehat{Q}(1-b)+\log Q(1-b)\right).\end{aligned}$$

Thus, together with the Taylor series expansion of $\log \widehat{Q} - \log Q$ used in the proof of Theorem 1, we can apply several integration and Dirac functionals to the basic convergence (A.2) – together with the analysis for $H_{(1)}$ – to obtain

$$\begin{aligned}\sqrt{n}\left(H_{(2)} - \tau_2\right) \stackrel{\mathcal{L}}{\Longrightarrow} 2b^{-1} &\int_0^b \frac{q(1-x)}{Q(1-x)} \log Q(1-x)\, B(1-x)\, dx \\ &- 2b^{-1} \int_0^b \frac{q(1-x)}{Q(1-x)} B(1-x)\, dx \,\log Q(1-b) \\ &- 2b^{-1} \int_0^b \log Q(1-x)\, dx \,\frac{q(1-b)}{Q(1-b)} B(1-b) \\ &+ 2\log Q(1-b)\, \frac{q(1-b)}{Q(1-b)} B(1-b).\end{aligned}$$

This is U_2, which completes the proof of the weak convergence.

Our next step is to compute the variance. First, writing $B(b) = W(b) - bW(1)$ for a Brownian Motion $W(\cdot)$, several pages of algebra produce

$$\begin{aligned}\alpha_1 U_1 + \alpha_2 U_2 = {} & (B(1) - B(b))\, \left(\lambda\,(\alpha_1 + 2\tau_1\alpha_2) - \tau_1\alpha_1 - \tau_2\alpha_2\right) \\ & + B(b)\, \big(\alpha_1[(1-b^{-1})\lambda - b^{-1}\log \overline{Q}(b) - \tau_1] \\ & + \alpha_2[2(1-b^{-1})\lambda\tau_1 + b^{-1}\log^2 \overline{Q}(b) - \tau_2]\big) \\ & + b^{-1} \int_0^b \left[(\alpha_1 - 2\alpha_2 \log \overline{Q}(b))\, \log \overline{Q}(u) + \alpha_2 \log^2 \overline{Q}(u)\right] dB(u).\end{aligned}$$

Note that if we set $\alpha_1 = 1$ and $\alpha_2 = 0$, the calculation would correspond to the Hill limit; therefore at the end of the variance calculation, we will at once obtain the Hill variance. The above form is convenient because two of the three terms involve independent increments. The variance is then

$$\begin{aligned}&(1-b)\, \left(\lambda\,(\alpha_1 + 2\tau_1\alpha_2) - \tau_1\alpha_1 - \tau_2\alpha_2\right)^2 \\ &\quad + b\, \left(\alpha_1[(1-b^{-1})\lambda - b^{-1}\log \overline{Q}(b) - \tau_1] + \alpha_2[2(1-b^{-1})\lambda\tau_1 + b^{-1}\log^2 \overline{Q}(b) - \tau_2]\right)^2 \\ &\quad + b^{-2} \int_0^b \left[(\alpha_1 - 2\alpha_2 \log \overline{Q}(b))\, \log \overline{Q}(u) + \alpha_2 \log^2 \overline{Q}(u)\right]^2 du \\ &\quad + 2b^{-1} \int_0^b \left[(\alpha_1 - 2\alpha_2 \log \overline{Q}(b))\, \log \overline{Q}(u) + \alpha_2 \log^2 \overline{Q}(u)\right] du \\ &\quad \cdot \left(\alpha_1[(1-b^{-1})\lambda - b^{-1}\log \overline{Q}(b) - \tau_1] + \alpha_2[2(1-b^{-1})\lambda\tau_1 + b^{-1}\log^2 \overline{Q}(b) - \tau_2]\right).\end{aligned}$$

The first and second terms can be written as

$$\begin{aligned}&(1-b)\, \left(\alpha_1^2(\lambda - \tau_1)^2 + 2\alpha_1\alpha_2(\lambda - \tau_1)(2\lambda\tau_1 - \tau_2) + \alpha_2^2(2\lambda\tau_1 - \tau_2)^2\right), \quad \text{and} \\ &b\, \left([\alpha_1(\lambda - \tau_1) + \alpha_2(2\lambda\tau_1 - \tau_2)] - b^{-1}\left[\alpha_1(\lambda + \log \overline{Q}(b)) + \alpha_2(2\lambda\tau_1 - \log^2 \overline{Q}(b))\right]\right)^2\end{aligned}$$

respectively. Let $L_k = b^{-1} \int_0^b \log^k \overline{Q}(u)\, du$, and note that $\tau_1 = L_1 - \log \overline{Q}(b)$ and $\tau_2 - \tau_1^2 = L_2 - L_1^2$. Then, the third term is

$$\begin{aligned}&b^{-1}\, \left(\alpha_1^2 L_2 + \alpha_1\alpha_2(-4L_2 \log \overline{Q}(b) + 2L_3) + \alpha_2^2(4L_2 \log^2 \overline{Q}(b) + L_4 - 4L_3 \log \overline{Q}(b))\right) \\ &\quad = b^{-1}\, \left(\alpha_1^2 L_2 + \alpha_1\alpha_2(-4L_2[L_1 - \tau_1] + 2L_3) + \alpha_2^2(4L_2[\tau_2 - L_2 + 2L_1(L_1 - \tau_1)] + L_4 - 4L_3[L_1 - \tau_1])\right).\end{aligned}$$

Finally, the fourth term can be simplified to

$$\begin{aligned}&2\left([\alpha_1 - 2\alpha_2 \log \overline{Q}(b)]L_1 + \alpha_2 L_2\right)\, (\alpha_1[\lambda - \tau_1] + \alpha_2[2\lambda\tau_1 - \tau_2]) \\ &\quad - 2b^{-1}\, \left([\alpha_1 - 2\alpha_2 \log \overline{Q}(b)]L_1 + \alpha_2 L_2\right)\, \left(\alpha_1[\lambda + \log \overline{Q}(b)] + \alpha_2[2\lambda\tau_1 - \log^2 \overline{Q}(b)]\right).\end{aligned}$$

Combining the first, second, and fourth terms, along with $R_2 = L_2 - 2L_1 \log \overline{Q}(b)$, yields

$$\begin{aligned}&-(\alpha_1[\lambda - \tau_1] + \alpha_2[2\lambda\tau_1 - \tau_2])^2 + b^{-1}\, (\alpha_1[\lambda - \tau_1 + L_1] + \alpha_2[2\lambda\tau_1 - \tau_2 + R_2]) \\ &\quad \times (\alpha_1[\lambda - \tau_1 - L_1] + \alpha_2[2\lambda\tau_1 - \tau_2 - R_2]).\end{aligned}$$

Summing all four terms, we have $(b^{-1} - 1)\, (\alpha_1[\lambda - \tau_1] + \alpha_2[2\lambda\tau_1 - \tau_2])^2$ plus b^{-1} times

$$\alpha_1^2\,(\tau_2 - \tau_1^2) + 2\alpha_1\alpha_2(\tau_3 - \tau_1\tau_2) + \alpha_2^2(\tau_4 - \tau_2^2)$$

after much simplification using (5). This completes the derivation of (23). Since $\overline{K}_1, \overline{K}_2, \overline{K\ell}$ all tend to zero as $b \to 0$, the DEdH bias also tends to zero. Finally, the limiting variance is obtained using $\lambda \to \xi$ when $P \in \mathcal{L}_c$, along with $\tau_k \to k!\,\xi^k$.

Hence $\alpha_1 \to 1 - 2/\xi$ and $\alpha_2 \to 0.5\,\xi^{-2}$; plugging into (23) produces the limiting $1 + \xi^2$ when the variance is multiplied by b. $\square$

References

Aban, I.B., Meerschaert, M.M., 2004. Generalized least-squares estimators for the thickness of heavy tails. J. Statist. Plann. Inference 119, 341–352.
Beran, J., Schell, D., Stehlík, M., 2014. The harmonic moment tail index estimator: asymptotic distribution and robustness. Ann. Inst. Statist. Math. 66, 193–220.
Bingham, N., Goldie, C., Teugels, J., 1987. Regular Variation. Cambridge University Press, Cambridge.
Crovella, M., Taqqu, M., Betravos, A., 1998. Heavy-tailed probability distributions in the world wide web. In: Adler, R., Feldman, R., Taqqu, M. (Eds.), A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Birkhäuser, Boston, pp. 3–25.
Csörgő, S., Viharos, L., 1997. Asymptotic normality of least-squares estimators of tail indices. Bernoulli 3, 351–370.
Dacorogna, M., Müller, U., Pictel, O., de Vries, C., 1995. The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Data Sets. Discussion paper 95-70, Tinbergen Institute.
De Haan, L., Peng, L., 1998. Comparison of tail index estimators. Stat. Neerl. 52, 60–70.
Dekkers, A.L.M., de Haan, L., 1993. Optimal choice of sample fraction in extreme-value estimation. J. Multivariate Anal. 47, 173–195.
Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. Ann. Statist. 17, 1833–1855.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer, New York.
Fabián, Z., Stehlík, M., 2009. On Robust and Distribution Sensitive Hill Like Method. IFAS Research Paper Series 2009-43. http://www.jku.at/ifas/content/e108280/e146255/e146265/res43_steh.pdf.
Gihman, I., Skorohod, A., 1980. The Theory of Stochastic Processes. Springer, Berlin.
Gomes, M.I., de Haan, L., 1999. Approximation by penultimate extreme value distributions. Extremes 2, 71–85.
Gomes, M.I., Guillou, A., 2014. Extreme value theory and statistics of univariate extremes: a review. Internat. Statist. Rev. 1–30. http://dx.doi.org/10.1111/insr.12058.
Hall, P., 1990. Using the bootstrap to estimate mean square error and select smoothing parameters in non-parametric problems. J. Multivariate Anal. 32, 177–203.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163–1174.
Holan, S., McElroy, T., 2010. Tail exponent estimation via broadband log density-quantile regression. J. Statist. Plann. Inference 140, 3693–3708.
Huisman, R., Koedijk, K.G., Kool, C.J.M., Palm, F., 2001. Tail-index estimates in small samples. J. Bus. Econom. Statist. 19, 208–216.
Kiefer, N., Vogelsang, T., Bunzel, H., 2000. Simple robust testing of regression hypotheses. Econometrica 68, 695–714.
Kulik, R., Soulier, P., 2011. The tail empirical process of some long memory sequences. Stochastic Process. Appl. 121, 109–134.
Lachtman, J., Morgan, F., Aspinall, W., 2008. Temporal changes in the cumulative piecewise gradient of a variant of the gutenbergrichter relationship, and the imminence of extreme events. Earth-Sci. Rev. 87, 94–112.
Mandelbrot, B., 1963. The variation of certain speculative prices. J. Bus. 36, 394–419.
Mandelbrot, B., 1982. The Fractal Geometry of Nature. W.H. Freeman, San Francisco.
Mandelbrot, B., Hudson, R., 2008. The (Mis)behavior of Markets. Profile Books Ltd.
McElroy, T., Jach, A., 2012. Tail index estimation in the presence of long memory dynamics. Comput. Statist. Data Anal. 56, 266–282.
Pancheva, E.I., Gacovska, , 2014. Asymptotic behavior of central order statistics under monotone naturalizations. Theory Probab. Appl. 58 (1), 107–120.
Parzen, E., 1959. On choosing an estimate of the spectral density function of a stationary time series. Ann. Math. Statist. 28, 921–932.
Pickands III, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131.
Reis, P., Castro, L., Dias, S., Gomes, M., 2015. Penultimate approximations in statistics of extremes and reliability of large coherent systems. Methodol. Comput. Appl. Probab. 17, 189–206.
Smirnov, N.V., 1949. Limit distributions for the terms of a variational series. Tr. Mat. Inst. Steklova 25, 3–60. (in Russian); 1952. Amer. Math. Soc. Transl. 67, (in English).
Stehlík, M., Potocký, R., Waldl, H., Fabián, Z., 2010. On the favorable estimation for fitting heavy tailed data. Comput. Statist. 25, 484–503.
Sun, Y., Phillips, P.C.B., Jin, S., 2008. Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. Econometrica 76, 175–194.
Viharos, L., 1999. Weighted least-squares estimators of tail indices. Probab. Math. Statist. 19, 249–265.