
Volume 140 Issue 12 December 2010 ISSN 0378-3758
ELSEVIER
journal of statistical planning and inference
Special Issue :
In Honor of Emanuel Parzen on the Occasion of his 80th Birthday and Retirement from the Department of Statistics, Texas A&M University
ScienceDirect

Journal of Statistical Planning and Inference 140 (2010) 3693–3708

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Tail exponent estimation via broadband log density-quantile regression

Scott H. Holan [a,*], Tucker S. McElroy [b]

[a] Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100, USA
[b] Statistical Research Division, U.S. Census Bureau, 4700 Silver Hill Road, Washington, DC 20233-9100, USA

A R T I C L E I N F O

Available online 10 May 2010

Keywords:
Density-quantile
Extreme-value theory
Fourier series estimator
Quantile density
Tail index

A B S T R A C T

Heavy tail probability distributions are important in many scientific disciplines such as hydrology, geology, and physics and therefore feature heavily in statistical practice. Rather than specifying a family of heavy-tailed distributions for a given application, it is more common to use a nonparametric approach, where the distributions are classified according to the tail behavior. Through the use of the logarithm of Parzen's density-quantile function, this work proposes a consistent, flexible estimator of the tail exponent. The approach we develop is based on a Fourier series estimator and allows for separate estimates of the left and right tail exponents. The theoretical properties for the tail exponent estimator are determined, and we also provide some results of independent interest that may be used to establish weak convergence of stochastic processes. We assess the practical performance of the method by exploring its finite sample properties in simulation studies. The overall performance is competitive with classical tail index estimators, and, in contrast, with these our method obtains somewhat better results in the case of lighter heavy-tailed distributions.

1. Introduction

Heavy tail distributions naturally arise in many areas of science. Often it is impossible to choose an appropriate distribution for a given application *a priori*. For this reason, it is common in the literature to proceed nonparametrically via classifying the distribution by its tail behavior, and many such nonparametric tail index estimators have been proposed.

A classical estimator enjoying widespread use is due to Hill (1975). This method provides a robust estimator based on the asymptotics of extreme values, although the results are misleading when applied to data from the stable family (McCulloch, 1997). Alternatively, there is the Pickands (1975) estimator, which is easy to compute and invariant to certain shift and scale transformations, but also suffers from poor asymptotic efficiency. Several refinements have been suggested for both estimators (Gomes and Martins, 2001; Drees, 1995). In addition to the Hill and Pickands estimators and their refinements, there is the method of Csörgo et al. (1985), where the authors develop an estimate that is expressed as the convolution of a kernel with the logarithm of the quantile function; this includes as particular cases the estimates proposed by Hill (1975) and de Haan (1981). Further, de Haan and Resnick (1980) and Teugels (1981) provide examples of simple estimators based on order statistics. Alternatively, Hall and Welsh (1985) propose an estimator that assumes a

* Corresponding author.
E-mail addresses: holans@missouri.edu (S.H. Holan), tucker.s.mcelroy@census.gov (T.S. McElroy).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.04.035

general nonparametric model, in which it is assumed that the only available information comes from the asymptotic properties of the tail distribution. For a survey of recent research in this area, see Embrechts et al. (1997) and the references therein.

In contrast to these methods, Parzen (1979) suggested an alternative approach that uses the density-quantile function as a measure of "tail orderings". Later, Schuster (1984) refined Parzen's classification scheme and provided a connection with the limit in probability of extreme spacings, and Rojo (1996) developed an approach that relaxed the smoothness conditions required in Schuster (1984).

Our work utilizes the general approach of Parzen (1979), using the logarithm of the density-quantile function to separately estimate the left and right tail exponent. First, we estimate the log density-quantile function using nonparametric kernel methods. Second, we regress the resulting function on a Fourier expansion of the density-quantile, only assuming the asymptotic tail behavior so that the nonparametric flavor is preserved. Note that the estimation of the tail exponent via the log density-quantile in our approach bears a similarity to spectral estimation of the fractional differencing parameter of a long memory process, and so our work is similar in spirit to that of Hurvich and Brodsky (2001).

The remainder of this paper is organized as follows. In Section 2 we introduce Parzen's density-quantile function and develop the general framework for our method. Additionally, this section presents a rigorous derivation for a mapping between α and ν, respectively, the classical tail index and the tail exponent proposed by Parzen. Section 3 describes the statistical estimator of the tail exponent, while Section 4 contains theoretical results that establish its asymptotic behavior; we establish consistency as well as asymptotic normality under some more restrictive assumptions. The methodology we develop is tested in Section 5; extensive simulations provide an indication of the mean square error of our estimator for finite sample sizes under different underlying distributions. There we demonstrate the effectiveness of our estimator's ability to characterize light–heavy tail behavior while using *default* tuning parameters and compare the performance of our approach against the Hill, Pickands and DEdH (Dekkers et al., 1989) estimators. Section 6 contains a discussion: although our method is generally competitive with the classical estimates, the performance is superior in the case of lighter heavy-tailed distributions. This is an important contrast, indicating that our approach would be more successful with slightly heavy-tailed data (such as encountered in econometrics and finance) rather than extremely heavy-tailed data (such as insurance data). For convenience of exposition all proofs are left to the Appendix, which also contains some required results on stochastic processes.

2. Tail exponents and indices

Parzen (1979) discusses an approach to classifying tail behavior of probability laws, which considers the limiting behavior of the density-quantile function $fQ(u)$ as u approaches 0 or 1. Using the notation of Parzen (2004), suppose F is a continuous cumulative distribution function (cdf), let $f=F'$ denote the probability density function (pdf) and let Q denote the quantile function; then $F(Q(u))=u$ for all $u\in[0,1]$. Then

$$f(Q(u))Q'(u)=1, \tag{1}$$

where $f(Q(u))=fQ(u)$ is the density-quantile function and $q(u)=Q'(u)$ is the quantile density. Therefore, (1) implies $fQ(u)=1/q(u)$. Furthermore, let $J(u)$ denote the score function, which is defined by

$$J(u)=-(fQ(u))'=-\frac{f'(Q(u))}{fQ(u)}. \tag{2}$$

Following Parzen (2004), we assume that the representation near 0 and 1 is given by regularly varying functions:

$$fQ(u)=u^{\nu_0}L_0(u),\quad u\in[0,1/2), \tag{3}$$

$$fQ(u)=(1-u)^{\nu_1}L_1(1-u),\quad u\in(1/2,1], \tag{4}$$

where L_0, L_1 are slowly varying functions at zero. That is, for $i=0,1$, L_i satisfies the condition that for a fixed $y>0$

$$\frac{L_i(yu)}{L_i(u)}\to 1\quad \text{as } u\to 0.$$

Note that in Parzen (2004) the relations (3) and (4) only hold asymptotically as $u\to 0$ and $u\to 1$, respectively. However, by redefining L_i we can easily obtain the exact relations (3) and (4). In this context we call ν_0 and ν_1 the left and right tail exponents, respectively, and they are used as a measure of tail behavior (if we do not want to distinguish left from right and speak generically, we just refer to ν). Note that these exponents can be obtained explicitly via

$$\nu_0=\lim_{u\to 0^+}\frac{\log fQ(u)}{\log u}, \tag{5}$$

$$\nu_1=\lim_{u\to 1^-}\frac{\log fQ(u)}{\log(1-u)}, \tag{6}$$

since it is easily shown that $\log L(u)/\log u \to 0$ as $u \to 0$ if L is slowly varying at zero.[1] A distribution is considered to be heavy-tailed if $v > 1$, as established in the discussion below.

Examples. The uniform distribution on the unit interval has $F(x)=x$, and the density-quantile is $fQ(u)=1$ for $u \in [0,1]$. Thus $v_0 = v_1 = 0$. The exponential distribution has $F(x) = 1-e^{-\lambda x}$, so that the density-quantile is $fQ(u) = \lambda(1-u)$ for a positive rate λ. So the right tail exponent is $v_1 = 1$. Finally, the Cauchy distribution has $F(x) = (\arctan x)/\pi + \frac{1}{2}$, and the density-quantile is $fQ(u) = (1/\pi)\sin^2(\pi u)$. Using a Taylor series expansion, we find that $v_0 = v_1 = 2$.

Now taking the logarithm of (3) and (4), the resulting equations suggest using a regression estimator for the tail exponents. However, the form of L_0 and L_1 are generally not explicitly known, so we will utilize the Fourier representation of the logarithm of L_0 and L_1 in the space $\mathbb{L}_2[0,1]$ of square integrable functions defined on [0,1]; the following lemma justifies our approach.

Lemma 1. *If K is a slowly varying function at infinity and $L(x)=K(1/x)$ for $x \in (0,1)$, then $\log L$ is square integrable.*

Hence each L_i can be written as

$$L_i(u) = \exp\left\{\theta_{i,0} + 2\sum_{k=1}^{\infty}\theta_{i,k}\cos(2\pi k u)\right\},$$

and its order p truncation is

$$L_i^{(p)}(u) = \exp\left\{\theta_{i,0} + 2\sum_{k=1}^{p}\theta_{i,k}\cos(2\pi k u)\right\}. \tag{7}$$

Note that expanding these functions in terms of a Hilbert space basis implies that the coefficients will tend to decay as a function of the index k. Now $L_i^{(p)}$ is also slowly varying, and is asymptotic to a constant as the argument tends to zero (whereas for L_i, this need not be the case, since the coefficients $\theta_{i,k}$ need not be summable, only square summable). Additionally, since the system $\mathcal{S} = \{1, 2\cos(2\pi u), 2\cos(2\pi 2u), \ldots\}$ (the Fourier representation) is complete for the class of functions on $\mathbb{L}_2[0,1]$, $L_i^{(p)}$ converges to L_i in mean square as $p \to \infty$ (Mallat, 2001). That is, the system $\mathcal{S}$ forms an orthogonal basis for $\mathbb{L}_2[0,1]$. Note that defining $\log L_i$ in terms of its Fourier representation is nonparametric, and hence avoids the need to specify a functional form (i.e., a model) for L_i.

We now provide a mapping between the "classical" tail index and "Parzen" tail exponent, so that our approach can be compared with and embedded into the classical framework. For simplicity of exposition, we focus on the right tail index. Consider a heavy-tailed random variable X of right tail index $\alpha_1 > 0$, which is defined as follows. Letting F denote the cdf, we suppose that $1-F$ is regularly varying at ∞ of index $-\alpha_1$, i.e., $1-F(x) = x^{-\alpha_1}K(x)$ as $x \to \infty$, where K is a slowly varying function at ∞; compare with Embrechts et al. (1997, p. 75). Further, suppose that the probability density function f is ultimately monotone in its right tail, i.e., it is monotone on (z_1, ∞) for some z_1. Then by Theorem A3.7 of Embrechts et al. (1997, p. 568), for some slowly varying function L we have

$$f(x) \sim \alpha_1 x^{-(\alpha_1+1)}L(x) \quad \text{as } x \to \infty.$$

(We use the notation $a_n \sim b_n$ to denote that the limit of the ratio tends to unity.) Now, Parzen's right tail exponent is given by (6); we will directly calculate it in terms of α_1. Let $a_n = Q(1-1/n)$, so that we can write $a_n = n^{1/\alpha_1}P(n)$ for some slowly varying function P at ∞ (see Embrechts et al., 1997, p. 78 for a similar statement). It then follows that

$$fQ(1-1/n) = f(a_n) \sim \alpha_1 a_n^{-(\alpha_1+1)}L(a_n) = c\alpha_1 n^{-(1+1/\alpha_1)}P(n)^{-(1+\alpha_1)}L(a_n).$$

Again by Theorem A3.3 of Embrechts et al. (1997, p. 566), we have $\log P(n)/\log n \to 0$ and $\log L(a_n)/\log n \to 0$ as $n \to \infty$. Thus

$$v_1 = -\lim_{n\to\infty}\frac{\log fQ(1-1/n)}{\log n} = 1 + 1/\alpha_1.$$

The relation for the left tail index follows similarly. Thus, for $i=0,1$, we have

$$v_i = 1 + 1/\alpha_i \tag{8}$$

so that the formula holds for either left or right exponents. Another characterization of heavy-tailed distributions is through the extreme value index γ; see Drees (1998) for a discussion. It is well-known that $\alpha = 1/\gamma$ for classes of Pareto-like distributions, and it can be shown that $v = 1+\gamma$ for the distributions following (3) and (4) that are considered in this paper. The mapping between α and v is illustrated in the following examples.

Example 1. A stable variable has characteristic exponent $\delta \in (0,2]$, with $\delta = 2$ corresponding to the Gaussian distribution. When $\delta < 2$, the stable variable is heavy-tailed with classical $\alpha = \delta$. Note that $\delta = 1$ corresponds to the Cauchy distribution.

[1] The proof of this result is straight-forward for functions K that are slowly varying at infinity, using the representation Theorem A3.3 of Embrechts et al. (1997, p. 566). With the relation $L(x)=K(1/x)$, the above result is easily obtained; see McElroy and Politis (2007).

Thus, for the Cauchy distribution the Parzen tail exponent is 2, and more generally, for stable variables, we get all values between ∞ (the heaviest case) and 1.5.

Example 2. Another class of heavy-tailed variables is given by the Pareto, with $F(x)=1-(1+x)^{-\alpha}$ for $\alpha\in(0,\infty)$ and $x>0$. Thus $fQ(u)=\alpha(1-u)^{1+1/\alpha}$, and the corresponding Parzen tail exponents are $\nu_1=1+1/\alpha$ and $\nu_0=0$. For the left tail, observe that $Q(1/n)$ tends to the constant zero, and $f(0)$ is constant as well; finally $\log f(0)/\log n\to 0$. The right Parzen tail exponent attains any value between 1 and ∞.

Distributions with exponentially decaying tails, such as the Gaussian or exponential, do not fit into the heavy-tailed description for X, and thus this mapping does not apply. Heuristically, they correspond to $\alpha=\infty$, since their tails decay faster than any polynomial power of x, which corresponds to a Parzen tail exponent of 1.

3. Tail exponent estimators

Let $X_1,\ldots,X_n$ denote an iid sample. In the subsequent exposition we consider tail exponent estimators where we assume that the slowly varying function is given by (7) with p fixed and unknown. We suppose that the practitioner selects a value of p, say $\tilde{p}$, that is at least as large as the true value, and establish consistency (Theorem 1) and asymptotic normality (Theorem 2). It is possible to take the fully nonparametric perspective that $p=\infty$, letting $\tilde{p}$ increase as $n\to\infty$, in which case consistency can be proved so long as the θ_k coefficients decay sufficiently quickly; the details require some additional conditions and are presented separately in Theorem 3. For example, the Cauchy distribution ($\alpha=1$ in Example 1 of Section 2) has infinitely differentiable density quantile function, but $p_0=p_1=\infty$; on the other hand, the Pareto distribution (Example 2 of Section 2) has $\theta_{1,k}$ equal to $\log\alpha$ if k=0 and is zero otherwise—hence $p_0,p_1<\infty$.

In order to estimate ν_0 and ν_1, we consider $\log fQ(u)$ for $u\in(0,u_l]$ and $u\in[u_r,1)$, respectively, where $u_l\le\frac{1}{2}$ and $u_r\ge\frac{1}{2}$ can be chosen by the practitioner. Then applying (7) to (3) and (4) yields

$$\log fQ(u)=\nu_0\log(u)+\theta_{0,0}+2\sum_{k=1}^{p_0}\theta_{0,k}\cos(2\pi ku),\quad u\in(0,u_l], \tag{9}$$

$$\log fQ(u)=\nu_1\log(1-u)+\theta_{1,0}+2\sum_{k=1}^{p_1}\theta_{1,k}\cos(2\pi k(1-u)),\quad u\in[u_r,1). \tag{10}$$

It is important to note that in (9) and (10), log $fQ(u)$ is defined for $u\in(0,u_l]$ and $u\in[u_r,1)$ respectively.[2] There are two separate equations, one for the left and one for the right, since there are two possibly different slowly varying functions, where each have their own Fourier expansion. Here we also allow for the possibility of two different truncation orders p_0 and p_1. Now even though $L_i^{(p_i)}$ is a good approximation to L_i only in an aggregate $\mathbb{L}_2$ sense, this has little impact in practice since our main objective is estimation of ν_i. This flexible form for the slowly varying function is an advantage, since the practitioner is not forced to use an explicit "model" (e.g., compare Remark 2.4 of McElroy and Politis, 2008).

More strikingly, we exclude the percentiles u=0 and 1 so that the logarithmic expressions in (9) and (10) are well-defined. This means that the actual extremes (i.e., the maximum and minimum of the distribution) are omitted from our estimating equations, and hence *the sample extremes will not occur* in our estimates. This is not a problem for our estimator, since it relies not on the rate of convergence of certain statistics involving extremes (this can be viewed as the basis for the Hill estimator), but on the functional relationship of the whole (left or right) tail of the density-quantile function, as described in (3), (4) and (5), (6). A similar idea is at work in the long memory parameter estimation of Hurvich and Brodsky (2001), where periodogram ordinates at low frequencies are considered while frequency zero is omitted. The kernel smoothed estimator $\widehat{q}(u)$ of the quantile density that we use (detailed below) excludes some of the extreme order statistics, but in practice the tuning parameters can be set so that only the extremes are excluded.

Since $fQ(u)$ is unknown, estimation proceeds by first estimating the log density-quantile, and then using ordinary least squares regression via (9) and (10) to obtain estimates of the tail exponents ν_i. Specifically, let $\widehat{q}_n(u_j)$ denote an estimator of the quantile density $q(u)$ obtained from the data, where u_j=$(j-.5)/n$ and j=$0,1,\ldots,n$. Using the fact that $fQ(u)$=$1/q(u)$, we have by definition $\widehat{fQ}(u_j)=1/\widehat{q}_n(u_j)$, and thus $\log\widehat{fQ}(u_j)=-\log\widehat{q}_n(u_j)$ can be substituted in (9) and (10). We develop the exposition for the left tail exponent ν_0, noting that right tail exponent estimation follows analogously.

Let $y=\log\widehat{fQ}(\underline{\mathbf{u}})$ be the given log density-quantile estimate written as a column vector, where $\underline{\mathbf{u}}=(u_1,u_2,\ldots,u_l)'$ and $G_k=\cos(2\pi k\underline{\mathbf{u}})$. Note that we are now thinking of l as an integer so that u_l is of the form $(l-.5)/n$, which is mainly just a notational convenience; this l is fixed and pre-specified (e.g., $l=\lfloor n/2+.5\rfloor$). Thus $y=(\log\widehat{fQ}(u_1),\ldots,\log\widehat{fQ}(u_l))'$ and $G_k=(\cos(2\pi ku_1),\ldots,\cos(2\pi ku_l))'$. Then for any fixed positive integer $\tilde{p}_0$ chosen by the practitioner, define the $l\times\tilde{p}_0+2$ dimensional matrix $X=[G^*,G_0,2G_1,\ldots,2G_{\tilde{p}_0}]$, where $G^*=\log(\underline{\mathbf{u}})$. Furthermore, define $\widehat{\beta}_{\tilde{p}_0}=(\widehat{\nu}_0,\widehat{\theta}_{0,0},\widehat{\theta}_{0,1},\ldots,\widehat{\theta}_{0,\tilde{p}_0})'=$

[2] The assumption that $u\le u_l$ and $u\ge u_r$ is flexible, given that the user chooses these values, and is meant to restrict the estimators of the left and right tail exponents to the lower and upper quantiles, respectively. That is, we do not use large (small) order statistics when estimating the left (right) tail exponent.

$(X'X)^{-1}X'y$, which can be viewed as the ordinary least squares estimator of $\beta_{\tilde{p}_0} = (v_0, \theta_{0,0}, \theta_{0,1}, \ldots, \theta_{0,\tilde{p}_0})'$, where $\theta_{0,k} = 0$ if $k > p_0$. Letting e_j denote the $\tilde{p}_0 + 2$ dimensional vector with 1 in the jth position and zeros elsewhere, we have

$$\widehat{v}_0 = e_1' \widehat{\beta}_{\tilde{p}_0} = e_1'(X'X)^{-1}X'y. \tag{11}$$

This is an explicit formula for the estimate of the left tail index v_0, which depends on having y, the vector of estimates of the log density-quantile—but how should one choose $\widehat{q}(u)$, the estimate of the quantile density function (qdf)? There have been many research efforts aimed at this problem; see Cheng and Parzen (1997), Xiang (1994) and Falk (1986) for a more detailed discussion.

Letting the order statistics of the sample be denoted $X_{(1;n)} < X_{(2;n)} < \cdots < X_{(n;n)}$, a simple estimator of $q(u)$ is expressed in terms of the sample spacings

$$\widehat{q}_n(u_j) = n\{X_{(j+1;n)} - X_{(j;n)}\}, \tag{12}$$

for $u_j = (j - .5)/n$ and $j = 1, 2, \ldots, n-1$; see Parzen (1982). Based on simulation results, using (12) as a qdf estimator appears to provide an approximately unbiased estimate of v_i, but the resulting variability is too large to be competitive. A better class of statistics is provided by kernel quantile density estimators, first introduced by Parzen (1979). Let F_n denote the empirical distribution function (edf) of the sample, and its inverse will be $Q_n = F_n^{-1}$, called the empirical quantile function (eqf). One expression for the kernel quantile density estimator is given by

$$\widehat{q}_n(u) = \frac{1}{h_n^2} \int_0^1 Q_n(t) K'\left(\frac{t-u}{h_n}\right) dt \tag{13}$$

for some kernel K with derivative K'. Xiang (1994) suggests the quantile density estimator $\widehat{q}_n(t) = (1/nh_n^2) \sum_{i=1}^n K'((i/n - t)/h_n) X_{(i;n)}$ as an alternative that is easier to calculate than (13). More generally than (13), we will consider kernel qdf estimates of the form

$$\widehat{q}_n(u) = \frac{d}{du} \int_0^1 Q_n(t) K_n(u,t) d\mu_n(t), \quad u \in (0,1). \tag{14}$$

$K_n(u,t)$ is a kernel depending on n, e.g., $K_n(u,t) = K(h_n^{-1}(t-u))h_n^{-1}$. We will focus on kernel-smoothed estimators $\widehat{q}_n(u)$ that satisfy assumptions K_1–K_7 of Cheng (1995). For convenience we list these assumptions here. Also let $U = [a,b]$ denote an arbitrarily fixed subinterval of (0,1), while the measure μ_n and the kernel K_n satisfy appropriate variational properties discussed below.

K_1. For each n, $0 < \mu_n([0,1]) < \infty$ (but may depend on n), and $\mu_n(\{0,1\}) = 0$.
K_2. For each n and each (u,t), $K_n(u,t) \geq 0$, and for each $u \in U$, $\int_0^1 K_n(u,t)\, d\mu_n(t) = 1$.
K_3. For each n, $\int_0^1 t K_n(u,t)\, d\mu_n(t) = u$, $u \in U$.
K_4. There is a sequence $\delta_n \downarrow 0$ such that $\sup_{u \in U} | \int_{u-\delta_n}^{u+\delta_n} K_n(u,t)\, d\mu_n(t) - 1| \downarrow 0$ as $n \uparrow \infty$.

Conditions K_5–K_7 concern the derivative $K_n'(u,t) = \partial K_n(u,t)/\partial u$. Let S_n be the (unique) closed subset of (0,1) such that $\mu_n\{(0,1) \backslash S_n\} = 0$ and $\mu_n\{(0,1) \backslash S_n'\} > 0$ for any $S_n' \subset S_n$. For the sequence δ_n in K_4, let $I_n(u) = [u - \delta_n, u + \delta_n]$, $I_n^c(u) = (0,1) \backslash I_n(u)$, for $u \in U$. Define $\Lambda(u; K_n) = \int_{I_n(u)} |K_n'(u,t)|\, d\mu_n(t)$, $u \in U$; and for a well-defined function g on (0,1), let $R(g; K_n) = \sup_{u \in U} \int_{I_n^c(u)} |g(t) K_n'(u,t)|\, d\mu_n(t)$.

K_5. For each n $\sup_{u \in U} \int_0^1 |K_n'(u,t)|\, d\mu_n(t) < \infty$ (but may depend on n).
K_6. (a) For each n and each $u \in U$, $K_n(u,t) \equiv 0$, $t \in I_n^c(u)$; or (b) $S_n \subseteq [\varepsilon, 1-\varepsilon] \subset (0,1)$, with $U \subset [\varepsilon, 1-\varepsilon]$ for some $0 < \varepsilon < \frac{1}{2}$.
K_7. For the δ_n sequence in K_4, $\delta_n^2 \sup_{u \in U} \Lambda(u; K_n) \to 0$ and $R(1; K_n) \to 0$ as $n \uparrow 0$.

One estimator that satisfies these conditions is the "boundary-modified Bernstein polynomial". Let ε be such that $U \subset [\varepsilon, 1-\varepsilon] \subset (0,1)$, $L_\varepsilon = 1 - 2\varepsilon$ and $t_j = \varepsilon + (j/k)L_\varepsilon$, $j = 0, 1, \ldots, k$, where k is user-defined. Then the kth degree boundary-modified Bernstein polynomial qdf estimator on U can be expressed as

$$\widehat{q}_n^B(u) = \frac{1}{L_\varepsilon^k} \sum_{j=0}^{k-1} \frac{Q_n(t_{j+1}) - Q_n(t_j)}{1/k} \binom{k-1}{j} (u-\varepsilon)^j (1-\varepsilon-u)^{k-1-j}. \tag{15}$$

Letting $k = k_n \uparrow \infty$ as $n \uparrow \infty$, Cheng (1995) shows assumptions K_1–K_7 are satisfied; that paper also demonstrates that (15) is of type (13) (and hence of type (14) as well). Finally, with $y = \log \widehat{fQ}(u) = -\log \widehat{q}_n^B(u)$ we can form an estimate of v_0 via (11).

Remark 1. Using the boundary-modified Bernstein polynomial requires the choice of user-selected parameters. The values $k = n$ and $\varepsilon = .001$ performed well in simulation, and satisfy the necessary assumptions (see Section 5 for further discussion).

4. Asymptotic results

When estimating the qdf in the context of tail exponent estimation, the issue of how to choose the percentiles $\underline{\mathbf{u}}$ arises. For notational convenience in the asymptotic results below, we will suppose them to be of the form $u_j = j/n$, with j ranging between 1 and l for the left tail index, but $j=r, r+1,\ldots,n-1$ for the right tail index, where l and r are user-selected parameters such that $u_l \le \frac{1}{2}$ and $u_r \ge \frac{1}{2}$ for all n. Although in practice we take our percentiles of the form $u_j = (j-.5)/n$, there is no difference to the asymptotics. Results are stated in terms of a closed percentile set $U=[a,b]$, where a and b are chosen according to whether we are estimating the left or right tail exponent.

In the asymptotic results of this section, we consider both the case that the true p_0, p_1 are finite, and that they are infinite. When both are finite (e.g., satisfied by the Pareto distribution), the deterministic bias (arising from using a finite $\tilde{p}$ in our regressions) is negligible asymptotically, and hence consistency and asymptotic normality results (Theorems 1 and 2, respectively) can be formulated. However, when the true order is infinite, the bias does not vanish automatically, and we must make the additional assumption that $\log L_0$ and $\log L_1$ are smooth—equivalently, the coefficients θ_k are summable—in order to obtain consistency (Theorem 3 below). (The Cauchy distribution, for example, has the requisite smoothness.) However, the presence of this deterministic bias destroys the asymptotic normality, essentially altering the rate of convergence; therefore no central limit theorem is presented in this case.

For the following consistency results, we suppose that the quantile density function $q(u)$ is estimated with a kernel-smoothed estimator $\widehat{q}_n(u)$ (14), as in Cheng (1995)—note that this includes the more specific class (13) and the boundary-modified Bernstein polynomial estimator (15) as well. The kernel that such an estimator relies upon must satisfy some basic assumptions, such as K_1–K_7 of Cheng (1995), provided in Section 3. One example of such an estimator is given by (15). Additionally, some regularity conditions on the quantile density are also necessary: assumptions Q_1, Q_2, and Q_3 of Cheng (1995). For convenience these latter assumptions are discussed below.

Q_1 (Smoothness). The qdf is twice differentiable on (0,1).
Q_2 (Controlled tail). There exists a $\gamma > 0$ such that $\sup_{u\in(0,1)} u(1-u)|J(u)|/fQ(u) \le \gamma$.
Q_3 (Tail monotonicity). Either $q(0) < \infty$ or $q(u)$ is nonincreasing in some interval $(0,u_*)$, and either $q(1) < \infty$ or $q(u)$ is nondecreasing in some interval $(u^*,1)$.

These conditions are a bit stronger than the basic assumptions discussed in Section 2. Taking the lower percentiles, we have $q(u) = u^{-\nu_0}/L_0(u)$ for $u < \frac{1}{2}$, so that Q_1 is satisfied if L_0 is twice differentiable in $(0,\frac{1}{2})$. Q_2 is automatically satisfied using (2), since the limits (5) and (6) exist. Q_3 may or may not be satisfied in general, depending on the form of L_0; certainly, the assumption of Q_1 and Q_3 places no burdensome restriction on the slowly varying function L_0.

Since $\log fQ(u) = -\log q(u)$ and $\log \widehat{fQ}(u) = -\log \widehat{q}(u)$, we can write regression equations using (9) and (10) for the left and right tail exponents ν_0 and ν_1:

$$\log \widehat{fQ}(u_j) = \nu_0 \log(u_j) + \theta_{0,0} + 2\sum_{k=1}^{p_0} \theta_{0,k} \cos(2\pi k u_j) + \varepsilon(u_j),$$

$$\log \widehat{fQ}(1-u_j) = \nu_1 \log(u_j) + \theta_{1,0} + 2\sum_{k=1}^{p_1} \theta_{1,k} \cos(2\pi k u_j) + \varepsilon(1-u_j),$$

where $\varepsilon(u) = -\log\{\widehat{q}(u)/q(u)\}$ is the "residual" process. Then we have the following consistency theorem.

Theorem 1 (*Consistency*). *Suppose that the density-quantile function $q(u)$ satisfies Q_1, Q_2, Q_3, and we construct a kernel-smoothed estimator $\widehat{q}(u)$ with kernel satisfying K_1–K_7 of Cheng (1995). Moreover, suppose that we consider each regression with the percentiles restricted to some closed subset $U=[a,b]$. Also suppose that p_0, p_1 are finite but with $\tilde{p}_i > p_i$ for $i=0,1$. Then the estimates $\widehat{\nu}_0$ and $\widehat{\nu}_1$ are consistent.*

Not only are the estimates we obtain consistent, but as the following theorem shows, our estimates are also asymptotically normal under some additional assumptions. For this result, we suppose that $q(u)$ is estimated by a kernel estimator given by convolution as in (13), as opposed to the more general (14); of course, this includes the boundary-modified Bernstein polynomial estimator (15) as well. Furthermore, to establish the result we need the following additional notation: let $G^*(u) = \log(u)$ and $G_k(u) = \cos(2\pi k u)$, and let the vector $(w^*, w_0, \ldots, w_{\tilde{p}_i})$ denote the first row of the limiting inverse matrix of $(X'X)/n$. Then we define for $i=0,1$

$$G(u) = w^* G^*(u) + w_0 G_0(u) + 2w_1 G_1(u) + \cdots + 2w_{\tilde{p}_i} G_{\tilde{p}_i}(u). \tag{16}$$

We require an additional assumption on the kernel K:

$$K_8. \quad \sup_{u\in U} \left| h_n^{-1} K\left(\frac{s-u}{h_n}\right) - h_n^{-1} K\left(\frac{t-u}{h_n}\right) \right| \le C_n |t-s|^{\beta} \quad \text{and} \quad |K''(x)| \le C/|x|$$

for some constant $C > 0$, and $|x|$ sufficiently large. The C_n's are positive constants with $\sup_{n\ge 1} C_n < \infty$, and the rate β can be any positive number.

Theorem 2 (*Asymptotic normality*). *Suppose the same assumptions as in Theorem* 1 *hold, and in addition suppose that the kernel is symmetric and differentiable on* $[-1,1]$, *and satisfies assumption* K_8. *Let* $G(u)$ *be given by* (16) *and denote its derivative by* $g(u)$. *Let* h_n *be chosen such that* $nh_n^2 \to \infty$ *and* $nh_n^4 \to 0$, *but* $h_n \to 0$ *as* $n \to \infty$. *Also suppose that* p_0, p_1 *are finite but with* $\tilde{p}_i > p_i$ *for* $i=0,1$. *Then*

$$\sqrt{n}(\widehat{v}_i - v_i) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0,V)$$

where the variance V *is given by* (A.6).

Remark 2. The result of Theorem 2 is corroborated by simulation results in Section 5. It is difficult to use this result for the construction of confidence intervals, since the limiting variance is complicated and depends on the unknown function $q(u)$. Since the data are independent and heavy-tailed, it may be possible to use subsampling (Politis et al., 1999) or the jackknife (Shao and Tu, 1995) to estimate the variance.

We now provide a result on consistency for large p_0 and p_1. However, in this case we must reformulate our model equations (9) and (10) such that our regressors are orthogonal. Consider the Fourier expansion of $\log L_i(a+(b-a)u)$ for $u \in [0,1]$, which by change of variable amounts to

$$\log L_i(u) = \theta_{i,0} + 2\sum_{k=1}^{p_i} \theta_{i,k} \cos\left(2\pi k \frac{u-a}{b-a}\right), \quad u \in [a,b].$$

We then let the regressor functions be $G_0(u)=1$ and $G_j(u) = 2\cos(2\pi k(u-a)/(b-a))$ for $j \geq 1$, which are orthonormal over $u \in U = [a,b]$ with respect to the inner product $\langle \cdot,\cdot \rangle_{\mathbb{L}^2(U)}$ given by $\langle f,g \rangle_{\mathbb{L}^2(U)} = .5(b-a)^{-1}\int_a^b f(u)g(u)\,du$. Moreover $\{G_k\}$ is complete for $\mathbb{L}^2(U)$, so that $\log = \sum_{k \geq 0} \alpha_k G_k$ with

$$\alpha_k = \langle \log, G_k \rangle_{\mathbb{L}^2(U)} = \begin{cases} \dfrac{b\log b - a\log a}{2(b-a)} - 1/2, & k=0, \\ -\dfrac{1}{2\pi k}\displaystyle\int_0^{2\pi k} \frac{\sin u}{2\pi k a + (b-a)u}\,du, & k \geq 1. \end{cases}$$

These coefficients decay at rate k^{-1}. Then the design matrix becomes $X = [G^*, G_0, G_1, \ldots, G_{\tilde{p}_0}]$, following the notation of Section 3. The percentiles are restricted to U, but now the columns are asymptotically orthonormal (and linearly independent). This is because

$$\langle f,g \rangle_n := n^{-1} \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} f(j/n)g(j/n) = O(n^{-1}) + 2(b-a)\langle f,g \rangle_{\mathbb{L}^2(U)}$$

by Riemann integration (if f and g are bounded). Then under some additional conditions, the following consistency result holds as $\tilde{p}_i \to \infty$.

Theorem 3. *Assume the conditions of Theorem* 1, *but with the different regressors described above. Both* p_0 *and* p_1 *can be infinite, but assume that* $\log L_i(u)$ *is continuously differentiable, and that* $\tilde{p}_i$ *is chosen as a function of* n *such that* $\tilde{p}_i \log \tilde{p}_i / n \to 0$ *as* $n \to \infty$. *Then* $\widehat{v}_i \overset{P}{\longrightarrow} v_i$.

5. Empirical study

The theory we propose applies to distributions having both symmetric and asymmetric tails; popular tail index estimators, such as the Hill, Pickands and DEdH estimators, must be carefully adapted to the asymmetric case (Hill, 1975; Pickands, 1975; Dekkers et al., 1989). So one benefit of our method is the ease of estimating a left or right tail exponent (index) when the distribution is asymmetric. In this section we evaluate the utility and finite sample performance of our estimators through an extensive simulation study. This illustrates the asymptotic theory and provides an empirical comparison with several established methods under various distributional assumptions. In particular, we compare our estimator with the Hill, Pickands and DEdH estimators under several different bandwidths b, using the Burr, Student t and $\alpha-$stable distributions. We are able to demonstrate that, under a *default* specification of tuning parameters, our estimator is competitive with or superior to the other estimators investigated when estimating the tail exponent (index) for somewhat lighter heavy-tailed distributions.

The simulation study we undertake uses (15) as an estimate of $q(u)$. Recall that (15) is of form (13), and hence satisfies the conditions of Theorems 1–3. Specifically, we choose the number of grid points (i.e., the u_j) equal to the sample size and choose $k=n$, $\varepsilon = .001$, and $\tilde{p}_0 = \tilde{p}_1 = 1, \ldots, 4$. Other values of k and ε were investigated, yielding comparable results. We present the results with $k=n$ and $\varepsilon = .001$ since this constitutes our recommended *default* tuning parameters with $\tilde{p}_0 = \tilde{p}_1 = 1$. As noted previously, this choice of parameters does not necessarily constitute an *optimal* choice of tuning parameters. However, this choice performs well in practice and is thus provided here as a recommended *default*.

The method we propose requires the user to specify 4 "user-defined" tuning parameters. In contrast, the Hill, Pickands and DEdH require only one (the bandwidth). Although our estimator necessitates that the user specify more tuning parameters than in the case of the Hill, Pickands and DEdH, as noted, our recommended *default* specification can be used to alleviate this need and has performed well in practice.

In order to simulate data from a distribution of a specified tail index we utilize the Burr (with $\kappa=\tau=1$—this is actually a Pareto, so that Theorems 1 and 2 apply), Student *t*, and $\alpha-$stable distributions. The preceding distributions all have symmetric tails and we calculate the tail exponent, ν, using our formulation from (8). This procedure was carried out for several values of the tail exponent ν, using 1000 repetitions and a sample size of 1000. The tail exponents ν were chosen between 1.05 and 3 ($.5 \le \alpha \le 20$) for the Burr and *t* distributions, and between 1.556 and 3 ($.5 \le \alpha \le 1.8$) for the $\alpha-$stable distribution. (Since the $\alpha-$stable distribution is only defined for $\alpha \le 2$, we take the maximum index to be $\alpha = 1.8$, or $\nu = 1.556$.) We present results for both the left and right tail exponent.

Table 1 shows the results of the empirical study for the Burr ($\kappa=\tau=1$) distribution. Specifically, we compare the mean square error (MSE) for all the estimators under consideration. The most notable attribute of our estimator is that for $\nu \le 1.2$ ($\alpha \ge 5$) our estimator is universally superior to the Hill, Pickands, and DEdH in terms of lower MSE. Further, for $\nu \le 1.556$ ($\alpha \ge 1.8$) our estimator was competitive with those investigated under default tuning parameter specifications (i.e., $\tilde{p}_0=\tilde{p}_1=1$).

Similarly Table 2 displays the results of the empirical investigation under the *t* distribution. Although our results under this distribution do not universally out-perform that of the Hill, Pickands, and DEdH, our estimator is superior to Hill (b=100,200), Pickands (b=50,100) and DEdH (b=50,100,200) for $\nu \le 1.2$ ($\alpha \ge 5$) and competitive for $\nu \le 1.333$ ($\alpha \ge 3$). Again, this is noteworthy as these results are presented under a default specification of the tuning parameters for our estimator. Additionally, in practice we would not know the family of distributions under investigation; however through exploratory data analysis we may have indication as to the heaviness of the tail. Therefore in cases where the distribution is deemed to have a "light–heavy" tail our estimator (under default tuning parameters) provides a favorable alternative to popular tail index estimators in the literature.

The last distribution we investigated was the $\alpha-$stable family (Table 3). Since the $\alpha-$stable family is not defined for $\nu < 1.5$ ($\alpha > 2$), we would not expect our estimator's performance (in terms of MSE) to exceed that of its competitors. Although in this case our estimator is not preferable, it is important to note that in practice one would most likely be able to discern *a priori* (through exploratory data analysis) that one was estimating a heavier tail than is recommended under our approach.

An additional aspect of our estimator that was investigated empirically was the agreement of the distribution of the estimator with normality. One particular assessment we employed was visual inspection of the histogram of the distribution of our estimators with the standard normal pdf superimposed (Fig. 1). Although we only display one example histogram (for Burr ($\kappa=\tau=1$), $\nu = 1.1$) this figure was representative of the other simulations (and are thus not displayed). One thing to note (as depicted by this figure) is that although our estimator was consistently in close agreement with the normal distribution there were several instances where the other estimators appeared to violate normality.

In summary, it seems that tail exponent (index) estimation in the density-quantile framework performs better when estimating lighter heavy-tailed distributions. The superior estimation is due to the fact that the low tail thickness corresponds to values of ν close to unity, whereas α is tending to infinity. Specifically, it will be easier to estimate values

Table 1
Tail exponent (index) estimation using LDQ (log density-quantile) for ν_0 and ν_1, as well as the Hill, Pickands, and DEdH estimators.

ν (α)	**MSE—Burr distribution** $\kappa=\tau=1$															
	LDQ—ν_0				**LDQ**—ν_1				**Hill**			**Pickands**		**DEdH**		
	p=1	2	3	4	1	2	3	4	b=50	100	200	50	100	50	100	200
3 (.5)	**.446**	.654	.965	1.146	**.167**	.237	.352	.415	.080	.038	.031	.146	.072	.098	.049	**.029**
2.25 (.8)	**.120**	.182	.282	.345	**.066**	.095	.144	.179	.031	.020	.057	.109	**.005**	.054	.027	.027
2 (1)	**.073**	.108	.167	.208	**.049**	.071	.107	.132	.023	**.022**	.081	.094	.045	.048	.024	.027
1.833 (1.2)	**.057**	.084	.131	.163	**.037**	.052	.078	.098	**.019**	.027	.104	.086	.042	.034	.020	.027
1.667 (1.5)	**.041**	.060	.093	.115	**.028**	.044	.068	.086	**.017**	.036	.133	.088	.042	.034	.019	.024
1.556 (1.8)	**.030**	.045	.070	.088	**.023**	.035	.054	.068	.021	.048	.161	.078	.038	.029	**.017**	.025
1.5 (2)	**.025**	.039	.059	.075	**.022**	.032	.049	.062	.022	.052	.177	.075	.036	.030	**.017**	.024
1.333 (3)	**.019**	.030	.046	.058	**.016**	.024	.040	.054	.035	.078	.231	.072	.036	.026	**.016**	.021
1.25 (4)	**.017**	.025	.040	.049	**.014**	.023	.036	.049	.045	.096	.262	.065	.034	.028	**.016**	.019
1.2 (5)	**.015**	.024	.037	.048	**.013**	.022	.035	.048	.055	.111	.287	.065	.034	.028	**.016**	.020
1.182 (5.5)	**.014**	.023	.037	.047	**.013**	.020	.034	.045	.057	.114	.297	.067	.031	.025	**.015**	.018
1.167 (6)	**.013**	.022	.034	.044	**.012**	.019	.031	.043	.059	.120	.303	.072	.033	.024	**.014**	.018
1.1 (10)	**.012**	.019	.030	.040	**.011**	.017	.030	.042	.074	.139	.340	.067	.036	.029	**.015**	.016
1.067 (15)	**.011**	.017	.027	.035	**.011**	.017	.027	.038	.083	.153	.358	.070	.033	.027	**.015**	.017
1.05 (20)	**.012**	.018	.028	.037	**.011**	.018	.030	.040	.086	.158	.371	.066	.035	.028	**.015**	**.015**

The simulations were drawn from the Burr distribution with $\kappa=\tau=1$. The LDQ estimator used here is given by (15) with k=n, $\varepsilon = .001$. The simulations consisted of 1000 repetitions of sample size 1000. Note that the bold font entries denote ν_0 and ν_1 for p=1 (the *default* specification) along with the minimum estimate among the Hill, Pickands, and DEdH estimators.

Table 2
Tail exponent (index) estimation using LDQ (log density-quantile) for ν_0 and ν_1, as well as the Hill, Pickands, and DEdH estimators.

ν (**df**)	**MSE—*t*-distribution**															
	LDQ—ν_0				**LDQ**—ν_1				**Hill**			**Pickands**		**DEdH**		
	$p=1$	2	3	4	$p=1$	2	3	4	$b=50$	100	200	$b=50$	100	$b=50$	100	200
3 (.5)	**.422**	.503	.833	.958	**.206**	.217	.361	.419	.078	.039	**.019**	.167	.094	.101	.049	.024
2.25 (.8)	**.133**	.145	.246	.286	**.079**	.082	.132	.157	.032	.016	**.011**	.107	.110	.053	.026	.014
2 (1)	**.090**	.105	.176	.207	**.052**	.058	.093	.113	.020	**.011**	.014	.090	.142	.045	.021	.012
1.833 (1.2)	**.064**	.075	.126	.154	**.041**	.053	.080	.103	.014	**.008**	.020	.088	.144	.042	.019	.011
1.667 (1.5)	**.038**	.049	.083	.101	**.027**	.039	.006	.079	.009	**.007**	.031	.097	.179	.033	.016	.009
1.556 (1.8)	**.029**	.042	.071	.088	**.025**	.040	.058	.078	**.007**	.008	.041	.109	.192	.034	.017	.010
1.5(2)	**.026**	.037	.062	.079	**.021**	.037	.051	.071	**.006**	.009	.049	.100	.198	.032	.016	.009
1.333 (3)	**.019**	.031	.045	.059	**.019**	.037	.046	.062	**.007**	.019	.080	.104	.216	.033	.019	.013
1.25 (4)	**.018**	.030	.038	.050	**.021**	.037	.043	.058	**.010**	.027	.102	.121	.224	.034	.020	.017
1.2 (5)	**.017**	.030	.036	.047	**.022**	.041	.044	.059	**.014**	.035	.120	.124	.211	.035	.022	.022
1.182 (5.5)	**.018**	.028	.036	.047	**.025**	.042	.046	.063	**.015**	.037	.125	.129	.217	.035	.024	.025
1.167 (6)	**.019**	.031	.037	.050	**.024**	.040	.043	.059	**.017**	.040	.132	.129	.214	.035	.021	.025
1.1 (10)	**.024**	.037	.041	.052	**.028**	.046	.048	.064	**.025**	.054	.158	.132	.214	.043	.028	.031
1.067 (15)	**.026**	.038	.041	.052	**.032**	.051	.052	.067	**.030**	.062	.176	.144	.204	.044	.032	.039
1.05 (20)	**.025**	.037	.039	.049	**.034**	.051	.055	.070	**.032**	.068	.183	.133	.214	.050	.036	.041

The simulations were drawn from the *t*-distribution. The LDQ estimator used here is given by (15) with $k=n$, $\varepsilon=.001$. The simulations consisted of 1000 repetitions of sample size 1000. Note that the bold font entries denote ν_0 and ν_1 for $p=1$ (the *default* specification) along with the minimum estimate among the Hill, Pickands, and DEdH estimators.

Table 3
Tail exponent (index) estimation using LDQ (log density-quantile) for ν_0 and ν_1, as well as the Hill, Pickands, and DEdH estimators.

ν (α)	**MSE—α-stable distribution**															
	LDQ—ν_0				**LDQ**—ν_1				**Hill**			**Pickands**		**DEdH**		
	$p=1$	2	3	4	$p=1$	2	3	4	$b=50$	100	200	$b=50$	100	$b=50$	100	200
3 (.5)	**.407**	.615	.902	1.064	**.179**	.260	.380	.452	.075	**.050**	.089	.184	.132	.098	.053	.062
2.25 (.8)	**.131**	.156	.257	.304	**.078**	.086	.139	.162	.032	**.019**	.031	.113	.057	.051	.026	.020
2 (1)	**.094**	.104	.177	.210	**.054**	.060	.098	.125	.019	**.010**	.014	.095	.134	.044	.020	.011
1.833 (1.2)	**.061**	.073	.122	.146	**.043**	.052	.082	.100	.015	.007	**.006**	.107	.293	.038	.019	.011
1.667 (1.5)	**.036**	.054	.099	.121	**.027**	.043	.064	.082	.016	.013	**.002**	.241	.585	.034	.023	.029
1.556 (1.8)	**.072**	.071	.076	.096	**.096**	.094	.079	.091	.045	.030	**.002**	.485	.819	.054	.072	.129

The simulations were drawn from the α–stable distribution. The LDQ estimator used here is given by (15) with $k=n$, $\varepsilon=.001$. The simulations consisted of 1000 repetitions of sample size 1000. Note that the bold font entries denote ν_0 and ν_1 for $p=1$ (the *default* specification) along with the minimum estimate among the Hill, Pickands, and DEdH estimators.

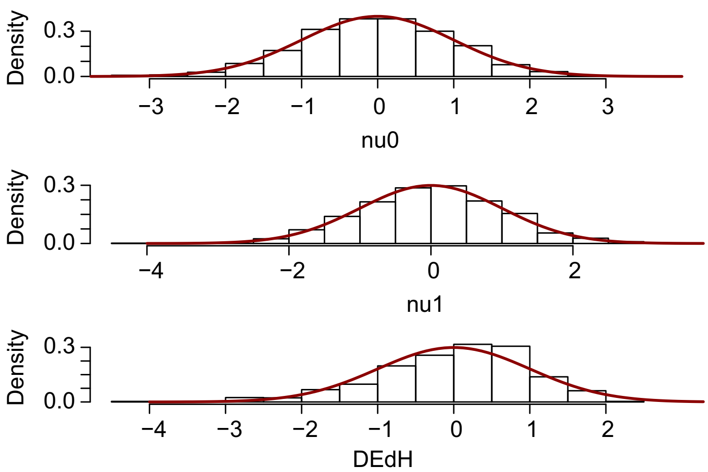

Fig. 1. This figure contains a histogram for the normalized ("studentized") distribution of the left and right tail exponent (index) estimators (ν_0 and ν_1) along with the DEdH estimator ($b=100$) for the Burr ($\kappa=\tau=1$) distribution with $\nu=1.1$ ($\alpha=10$) from a simulation with 1000 repetitions and of sample size 1000. The standard normal pdf is superimposed for reference.

close to one then around infinity, and so it is intuitive that the variance will tend to be lower. Thus our estimator improves for larger values of α, which is in contrast to the other estimators under investigation.

Finally, it should be noted that although the performance of our tail exponent estimator appears asymmetric with respect to the left and right tail exponent, this is a result of making default choices for the tuning parameters *a priori*. Specifically, the use of (15) as a starting point for our estimator yields a boundary bias which is hard to minimize for both tails simultaneously, but the performance of our estimator can be improved upon if one is willing to deviate from the default recommendations we have provided.

6. Discussion

In this paper we developed a new method of tail exponent (index) estimation. The approach we propose evolves naturally out of the density-quantile framework for classifying probability laws via tail behavior. Moreover, we argue that our method is rather flexible, allowing for separate left and right tail exponent (index) estimation when little or nothing is known about the distribution *a priori*; we impose the requirement that the data are iid and have tail behavior governed by (3) and (4). By making some additional particular assumptions on the data, we show that our tail exponent estimator is both consistent and asymptotically normal. Since less work has been done on the Parzen tail exponent, we provide a bridge to the classical tail index theory through our results in Section 2. Furthermore, in the development of the asymptotic theory we provide results of independent interest that can be used to establish weak convergence of stochastic processes. Although the method we propose is fairly general, it still requires some user defined choices. For example, the qdf estimator and its associated "tuning" parameters all need to be chosen by the practitioner. Even though we have made recommendations for suitable default choices, we do not provide optimal selection criteria here.

To illustrate the finite sample performance of our estimator we provide the results of an extensive empirical study. This study involves simulating from a Burr (with $\kappa=\tau=1$), Student t, and $\alpha-$stable distribution, making use of the equivalence formula (8), and includes the mean square error for several different values of v. Further, we compare our estimator to the Hill, Pickands and DEdH estimators for several different bandwidth specifications. The results indicate decent performance for the estimator we develop. Additionally, our estimator improves (i.e., has smaller mean square error) for values of v closer to unity (i.e., larger values of α), and thus is superior for getting at lighter heavy-tailed distributions. Histograms of the distribution of the normalized estimator illustrate asymptotic normality.

In summary, this is a method that should be useful when little is known about the heavy-tailed distributional family, but it is suspected that the tails are on the lighter side (such as occur, for example, in econometrics and finance). Parzen's (1982) approach allows one to easily describe light-tailed as well as heavy-tailed behavior, whereas classical approaches tend to be focused on only the latter.

Acknowledgements

This paper would not have been possible without the pioneering wisdom of Emanuel Parzen.

Disclaimer: This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Appendix A. Results on stochastic processes

In order to prove Theorem 2, we need to establish a convergence of stochastic processes. This is done by first establishing some basic results on weak convergence, and then adapting these to the density-quantile estimate. Let $C[0,1]$ denote the space of continuous functions from $[0,1]$ into the real numbers; this is made into a metric space via the metric (4.1) of Karatzas and Shreve (1997). We commence with a result analogous to Theorem 4.9 of Karatzas and Shreve (1997, p. 62). First, let us define the concept of the *modulus of continuity* on $[0,1]$ for any $T \in (0,1)$:

$$m^T(\omega,\delta) = \max_{|s-t|\le\delta,\ 0\le s,t\le T} |\omega(s)-\omega(t)|.$$

Proposition 1 (Karatzas and Shreve, 1997). *A set $A \subseteq C[0,1]$ has compact closure if and only if the following two conditions hold*:

$$\sup_{\omega\in A} |\omega(0)| < \infty, \tag{A.1}$$

$$\lim_{\delta\downarrow 0} \sup_{\omega\in A} m^T(\omega,\delta) = 0 \quad \text{for every } T \in (0,1). \tag{A.2}$$

In condition (A.1), *the time point* 0 *can be replaced by the time point* 1.

Next we consider an adaptation of Theorem 4.10 of Karatzas and Shreve (1997, p. 63). By $\mathcal{B}(C[0,1])$ we denote the $\sigma-$field generated by open sets in $C[0,1]$. Recall that a sequence of probability measures $\{P_n\}_{n=1}^{\infty}$ is *tight*, by definition, if for every $\varepsilon>0$ there exists a compact set K in $C[0,1]$ such that $P_n(K)\geq 1-\varepsilon$ for all n. The following result gives two sufficient conditions for tightness that are easier to work with.

Proposition 2. *A sequence $\{P_n\}_{n=1}^{\infty}$ of probability measures on $(C[0,1],\mathcal{B}(C[0,1]))$ is tight if*

$$\lim_{\lambda\uparrow\infty}\sup_{n\geq 1} P_n[\omega:|\omega(0)|>\lambda]=0, \tag{A.3}$$

$$\lim_{\delta\downarrow 0}\sup_{n\geq 1} P_n[\omega:m^T(\omega,\delta)>\varepsilon]=0 \quad \forall T\in(0,1),\ \varepsilon>0. \tag{A.4}$$

In condition (A.3), *the time point* 0 *can be replaced by the time point* 1.

The preceding Propositions 1 and 2 are fairly standard, and may be used to establish the weak convergence of stochastic processes; a proof of the latter result can be found in Holan and McElroy (2009). In what follows, we consider the kernel quantile estimator related to (13)

$$\widehat{Q}_n(u)=\int_0^1 Q_n(t)h_n^{-1}K\left(\frac{u-t}{h_n}\right)dt,$$

which is introduced in Falk (1985). In a like manner, a deterministic approximation to the true $Q_n(u)$ is given by

$$\tilde{Q}_n(u)=\int_0^1 Q(t)h_n^{-1}K\left(\frac{u-t}{h_n}\right)dt.$$

Now Theorem 1.3 of Falk (1985) states that for any $u_1,u_2,\ldots,u_d\in U=[a,b]$ under certain conditions,

$$\sqrt{n}\{(\widehat{Q}_n(u_1)-\tilde{Q}_n(u_1)),\ldots,(\widehat{Q}_n(u_d)-\tilde{Q}_n(u_d))\}\overset{\mathcal{L}}{\Longrightarrow}\{W_{u_1},\ldots,W_{u_d}\}$$

as $n\to\infty$, where the W_{u_j}'s are jointly Gaussian with mean zero and covariance $q(u_i)q(u_j)u_i(1-u_j)$ for $u_i\leq u_j$. From this result we may guess that $\sqrt{n}(\widehat{Q}_n(u)-\tilde{Q}_n(u))$ as a stochastic process converges to the process $q(u)W(u)$, where $W(u)$ is a Brownian Bridge, since the respective finite-dimensional distributions converge. The following theorem gives conditions under which this convergence is true. We require an additional technical concept: let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space on which the random variables $X_1,X_2,\ldots$ are defined, and let P_n be the measure induced by $\sqrt{n}(\widehat{Q}_n(u)-\tilde{Q}_n(u))$ on the space $(C(U),\mathcal{B}(C(U)))$.

Theorem 4. *Suppose that Q has bounded derivative on U, and suppose that K has bounded support on U, integrates to one, and satisfies condition K_8. Then*

$$\sqrt{n}(\widehat{Q}_n(u)-\tilde{Q}_n(u))\overset{\mathcal{L}}{\Longrightarrow}q(u)W(u),$$

i.e., the induced measures P_n corresponding to $\sqrt{n}(\widehat{Q}_n(u)-\tilde{Q}_n(u))$ on the space $(C(U),\mathcal{B}(C(U)))$ converge weakly to a measure P, the distribution of $q(u)W(u)$.

Our next result develops some asymptotic theory for the regression estimate given by

$$\frac{1}{n}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\log\left(\frac{\widehat{q}_n(k/n)}{q(k/n)}\right)G(k/n). \tag{A.5}$$

Later we will need the following deterministic approximation to $q(u)$: $\tilde{q}_n(u)=\int_U Q(x)h_n^{-2}K'((u-x)/h_n)\,dx$. The function $G(u)$ is a fairly arbitrary regressor function. We formulate a general theory for the asymptotics of expressions (A.5), which may then be applied to obtain the asymptotics of the tail exponent estimators. Our main theorem is stated below:

Theorem 5. *Suppose that the quantile density function $q(u)$ satisfies Q_1, Q_2, Q_3, and we construct a kernel-smoothed $\widehat{q}_n(u)$ with kernel satisfying K_1–K_7 of Cheng (1995), as well as K_8 above. Let $G(u)$ be a regressor function with derivative $g(u)=G'(u)$, with g and G uniformly bounded on U. Let h_n be chosen such that $nh_n^2\to\infty$ and $nh_n^4\to 0$ but $h_n\to 0$ as $n\to\infty$. Then*

$$n^{-1/2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\log\left(\frac{\widehat{q}_n(k/n)}{q(k/n)}\right)G(k/n)\overset{\mathcal{L}}{\Longrightarrow}G(b)W(b)-G(a)W(a)-\int_a^b W(u)\left(g(u)-G(u)\frac{q'(u)}{q(u)}\right)du,$$

where $W(u)$ is a Brownian Bridge. The limiting variance is

$$V=\int_a^b G^2(u)\,du+\int_a^b\int_a^b G(u)G(v)\left(1+[(u\wedge v)-uv]\frac{q'(u)}{q(u)}\frac{q'(v)}{q(v)}\right)du\,dv. \tag{A.6}$$

Appendix B. Proofs

Proof of Lemma 1. By Theorem A3.3 of Embrechts et al. (1997), we have

$$\log L(x) = \log c(1/x) + \int_z^{1/x} \frac{\delta(u)}{u}\, du$$

with $c(1/x) \to c_0 \in (0,\infty)$ as $x \to 0$, $\delta(u) \to 0$ as $u \to \infty$, and z some positive number. Then

$$\int_0^1 \left(\int_z^{1/x} \frac{\delta(u)}{u}\, du \right)^2 dx = \int_1^\infty y^{-2} \left(\int_z^y \frac{\delta(u)}{u}\, du \right)^2 dy,$$

and $\int_z^y \delta(u)/u\, du = o(\log y)$ by L'Hopital's rule, so that the integrand above is $o(y^{-2}\log^2 y)$ as $y \to \infty$. This establishes square integrability. □

Proof of Theorem 1. We focus on the v_0 case, since the v_1 case is similar. It follows from basic linear regression that

$$\widehat{v}_0 - v_0 = e_1'(X'X)^{-1}X'\underline{\varepsilon}$$

with $\underline{\varepsilon}$ the vector of $\varepsilon(u_j)$ such that the percentiles all lie in the set U. This amounts to considering u_j with $\lceil na \rceil \le j \le \lfloor nb \rfloor$. Let $\gamma = X(X'X)^{-1}e_1$, so that

$$|\widehat{v}_0 - v_0| = \left| \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \gamma_j \varepsilon(u_j) \right| \le \left(\sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \gamma_j^2 \right)^{1/2} \left(\sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \varepsilon^2(u_j) \right)^{1/2}$$

by the Cauchy–Schwartz inequality. Now $\sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \gamma_j^2 = e_1'(X'X)^{-1}e_1$, where the matrix $X'X$ has the following form:

$$X'X = \begin{bmatrix} \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \log^2(j/n) & \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \log(j/n) & 2\sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \log(j/n)\cos(2\pi j/n) & \cdots \\ \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \log(j/n) & \lfloor nb \rfloor - \lceil na \rceil & 2\sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \cos(2\pi j/n) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

By the definition of Riemann integration, $X'X/n \to M(\tilde{p}_0)$ as $n \to \infty$, where $M(\tilde{p}_0)$ is given by

$$M(\tilde{p}_0) = \begin{bmatrix} \int_a^b \log^2(u)\, du & \int_a^b \log(u)\, du & 2\int_a^b \log(u)\cos(2\pi u)\, du & \cdots \\ \int_a^b \log(u)\, du & b-a & 2\int_a^b \cos(2\pi u)\, du & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

This matrix is symmetric and invertible, being a Grammian matrix, and thus it follows that

$$|\widehat{v}_0 - v_0| \le \sqrt{[M^{-1}(\tilde{p}_0)]_{11}} \left(n^{-1} \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \varepsilon^2(u_j) \right)^{1/2}.$$

The stochastic error is

$$\varepsilon(j/n) = -\log\left(1 + \frac{\widehat{q}(j/n) - q(j/n)}{q(j/n)} \right),$$

to which we can apply a Taylor series expansion. Since j is chosen such that j/n is bounded away from 0 and 1, $q(j/n)$ is bounded away from zero. Following the notation in Cheng (1995), let $M_g = \sup_{u\in U}|g(u)|$ for any given function g defined on (0,1). Further, let $\tilde{q}_n(u) = \int_0^1 Q(t)K_n'(u,t)\, d\mu_n(t)$; this is a sort of deterministic approximation to $q(u)$ using kernel-smoothing, and is used as an intermediary term in our analysis. Then $d_n = \sup_{u\in U}|\tilde{q}_n(u) - q(u)|$ is the deterministic error of the estimate of $\widehat{q}_n(u)$ in estimating the qdf q. Additionally, let

$$B(q;K_n) = n^{-1/2}[M_q \Lambda_n^* \sqrt{2\delta_n \log\delta_n^{-1}} + M_{q'} + C_0 M_q n^{-1/2} A_\gamma(n)\Lambda_n^*]$$

with $\Lambda_n^* = \sup_{u\in U}\Lambda(u;\, K_n)$, C_0 a universal constant, q' equal to the derivative of q, and $n^{-\delta}A_\gamma(n) = o(1)$ for any $\delta > 0$ (γ is defined in Q_2). Then by Theorem 2.1 of Cheng (1995)

$$\sup_{u\in U} |\widehat{q}(u) - q(u)| = O_P(B(q;K_n) + d_n),$$

and it follows that by the use of Taylor series that $\sup_{\lceil na \rceil \le j \le \lfloor nb \rfloor}|\varepsilon(j/n)| = O_P(B(q;K_n)+d_n)$ as well, so that $\sqrt{n^{-1}\sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \varepsilon^2(u_j)} = O_P(B(q;K_n)+d_n)$. This establishes the consistency of $\widehat{v}_0$ (given that the kernel is selected such that $B(q;K_n)+d_n \to 0$). □

Proof of Theorem 4. First, we note that Propositions 1 and 2 can be extended from $C[0,1]$ to $C(U)$ trivially. The idea of the proof is to adapt the ideas from Theorem 4.15 of Karatzas and Shreve (1997)—merely adapt from $C[0,\infty)$ to $C(U)$ using the same arguments—and verify the conditions of Proposition 2 for the particular process at hand. Now recalling that $U=[a,b]$, the first condition is (A.3), which becomes

$$\sup_{n\geq q1}\mathbb{P}[\sqrt{n}|\widehat{Q}_n(a)-\tilde{Q}_n(a)|>\lambda]\to 0 \tag{B.1}$$

as $\lambda\to\infty$, using the definition of the induced measure P_n. Now Theorem 1.3 of Falk (1985) holds, due to the conditions in our theorem, so

$$\lim_{n\to\infty}\mathbb{P}[\sqrt{n}|\widehat{Q}_n(a)-\tilde{Q}_n(a)|>\lambda]=\mathbb{P}[|W_a|>\lambda].$$

Now pick any $\varepsilon>0$, and find M large enough such that $\mathbb{P}[|W_a|>\lambda]<\varepsilon$ for all $|\lambda|>M$ (this is accomplished, because W_a is Gaussian with finite variance). Then find N such that

$$|\mathbb{P}[\sqrt{n}|\widehat{Q}_n(a)-\tilde{Q}_n(a)|>\lambda]-\mathbb{P}[|W_a|>\lambda]|<\varepsilon$$

for all $n\geq N$ and $|\lambda|>M$. Then for $|\lambda|>M$,

$$\sup_{n\geq 1}\mathbb{P}[\sqrt{n}|\widehat{Q}_n(a)-\tilde{Q}_n(a)|>\lambda]=\max_{1\leq n<N}\mathbb{P}[\sqrt{n}|\widehat{Q}_n(a)-\tilde{Q}_n(a)|>\lambda]+\sup_{n\geq N}\mathbb{P}[\sqrt{n}|\widehat{Q}_n(a)-\tilde{Q}_n(a)|>\lambda].$$

The second term is bounded by 2ε, and by taking $|\lambda|$ still larger, the first term can be bounded by ε. This demonstrates (B.1). Next, we consider the condition that for any $\varepsilon>0$ we have

$$\sup_{n\geq 1}\mathbb{P}\left[\max_{|s-t|\leq\delta}|\sqrt{n}(\widehat{Q}_n(s)-\tilde{Q}_n(s))-\sqrt{n}(\widehat{Q}_n(t)-\tilde{Q}_n(t))|>\varepsilon\right] \tag{B.2}$$

tends to zero as $\delta\to 0$; this formulation is equivalent to (A.4) using the definition of induced measure. Now assuming K_8, take any $\varepsilon>0$ and $\delta>0$, it follows that

$$\begin{aligned}
&\sup_{n\geq 1}\mathbb{P}\left[\max_{|s-t|\leq\delta}|\sqrt{n}(\widehat{Q}_n(s)-\tilde{Q}_n(s))-\sqrt{n}(\widehat{Q}_n(t)-\tilde{Q}_n(t))|>\varepsilon\right]\\
&\quad=\sup_{n\geq 1}\mathbb{P}\left[\max_{|s-t|\leq\delta}\sqrt{n}\left|\int_U(F_n^{-1}(u)-F^{-1}(u))\left(h_n^{-1}K\left(\frac{u-s}{h_n}\right)-h_n^{-1}K\left(\frac{u-t}{h_n}\right)\right)du\right|>\varepsilon\right]\\
&\quad\leq\sup_{n\geq 1}\mathbb{P}\left[\max_{|s-t|\leq\delta}\sqrt{n}\int_U|F_n^{-1}(u)-F^{-1}(u)|C_n|t-s|^{\beta}\,du>\varepsilon\right]\\
&\quad=\sup_{n\geq 1}\mathbb{P}\left[\sqrt{n}\int_U|F_n^{-1}(u)-F^{-1}(u)|\,du>\varepsilon\delta^{-\beta}/C\right].
\end{aligned}$$

Now along the lines of the proof of (B.1), we can make δ smaller if needed, in order to replace the supremum by a limit superior. Hence we have the bound of

$$\limsup_{n\to\infty}\mathbb{P}\left[\sqrt{n}\int_U|F_n^{-1}(u)-F^{-1}(u)|\,du>\varepsilon\delta^{-\beta}/C\right]=\mathbb{P}\left[\int_U|q(u)W(u)|\,du>\varepsilon\delta^{-\beta}/C\right],$$

which uses the known weak convergence result $\sqrt{n}(F_n^{-1}(u)-F^{-1}(u))\overset{\mathcal{L}}{\Longrightarrow}q(u)W(u)$ (Gihman and Skorohod, 1980, p. 437). We have applied the continuous functional of absolute integration to this weak convergence result. Now we can let $\delta\to 0$, and obtain

$$\lim_{\delta\to 0}\mathbb{P}\left[\int_U|q(u)W(u)|\,du>\varepsilon\delta^{-\beta}/C\right]=0.$$

This establishes (B.2). Hence the induced measures P_n are tight, and the weak convergence is proved. □

Proof of Theorem 5. The proof proceeds in three major steps. First, we apply a Taylor series expansion to the logarithm. Second, we analyze the linearization of (A.5) and compute a Riemann sum approximation. Third, we apply continuous functionals to the resulting expression, utilizing Theorem 4 to obtain the stated convergence. For the first step, we expand in Taylor series as follows:

$$\log\left(\frac{\widehat{q}_n(k/n)}{q(k/n)}\right)G(k/n)=\left(\frac{\widehat{q}_n(k/n)-q(k/n)}{q(k/n)}\right)G(k/n)+R_{k,n},$$

where $R_{k,n}$ is the quadratic remainder, which depends on k and n. Now by Theorem 2.1 of Cheng (1995), which applies by our stated assumptions, there exists $0<\delta<\frac{2}{5}$ such that

$$\sup_{u\in U}|\widehat{q}_n(u)-q(u)|=O_P(n^{-\delta}).$$

Since G and $\tilde{q}_n$ are bounded away from infinity and zero, respectively, on U, the error satisfies $\sup_{k/n\in U} R_{k,n} = O_P(n^{-2\delta})$. Hence, multiplying by $\sqrt{n}$, the error still tends to zero, so that

$$n^{-1/2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\log\left(\frac{\widehat{q}_n(k/n)}{q(k/n)}\right)G(k/n) = O_P(n^{1/2-2\delta}) + n^{-1/2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\left(\frac{\widehat{q}_n(k/n)-q(k/n)}{q(k/n)}\right)G(k/n)$$

as $n\to\infty$. For the second step, it will be more convenient to work with

$$n^{-1/2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}(\widehat{q}_n(k/n)-\tilde{q}_n(k/n))\frac{G(k/n)}{q(k/n)},$$

the difference is given by

$$n^{-1/2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}(\tilde{q}_n(k/n)-q(k/n))\frac{G(k/n)}{q(k/n)}.$$

Now $\sqrt{n}(\tilde{q}_n(u)-q(u))$ will tend to zero uniformly for $u\in U$ since $nh_n^4\to 0$. This is because, as in Cheng (1995), the deterministic error d_n is $O(h_n^2)$ under Q_1 when K is symmetric on $[-1,1]$. Hence we require $\sqrt{n}h_n^2 = \sqrt{nh_n^4}\to 0$. Next, we have

$$n^{-1}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}(\widehat{q}_n(k/n)-\tilde{q}_n(k/n))\frac{G(k/n)}{q(k/n)} = \int_U (F_n^{-1}(x)-F^{-1}(x))n^{-1}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}h_n^{-2}K'\left(\frac{k/n-x}{h_n}\right)\frac{G(k/n)}{q(k/n)}\,dx,$$

with the inner sum being recognized as a deterministic Riemann sum. For each fixed x, we have

$$\begin{aligned}&\int_U h_n^{-2}K'\left(\frac{u-x}{h_n}\right)\frac{G(u)}{q(u)}\,du - n^{-1}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}h_n^{-2}K'\left(\frac{k/n-x}{h_n}\right)\frac{G(k/n)}{q(k/n)}\\ &\quad= h_n^{-2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\int_{k/n}^{(k+1)/n}\left(K'\left(\frac{u-x}{h_n}\right)\frac{G(u)}{q(u)} - K'\left(\frac{k/n-x}{h_n}\right)\frac{G(k/n)}{q(k/n)}\right)du\\ &\qquad+\int_a^{\lceil na\rceil/n}h_n^{-2}K'\left(\frac{u-x}{h_n}\right)\frac{G(u)}{q(u)}\,du - \int_b^{(\lfloor nb\rfloor+1)/n}h_n^{-2}K'\left(\frac{u-x}{h_n}\right)\frac{G(u)}{q(u)}\,du.\end{aligned}$$

Using the boundedness of K' and G and $1/q$, the latter two terms are $O(n^{-1}\,h_n^{-2})$. For the first term, we have an absolute bound of

$$h_n^{-2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\int_{k/n}^{(k+1)/n}\left|K'\left(\frac{u-x}{h_n}\right)-K'\left(\frac{k/n-x}{h_n}\right)\right|\left|\frac{G(u)}{q(u)}\right|du + h_n^{-2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\int_{k/n}^{(k+1)/n}\left|K'\left(\frac{k/n-x}{h_n}\right)\right|\left|\frac{G(k/n)}{q(k/n)}-\frac{G(u)}{q(u)}\right|du.$$

Now since g is uniformly bounded on U, we can use the Mean Value Theorem to bound the second term by $O(n^{-1}h_n^{-2})$. For the first term, we can use K_8 on the following:

$$\left|K'\left(\frac{u-x}{h_n}\right)-K'\left(\frac{k/n-x}{h_n}\right)\right| = |K''(z^*)|\left|\frac{u-k/n}{h_n}\right|,$$

where z^* is between $(u-x)/h_n$ and $(k/n-x)/h_n$. Since $u\in[(k-1)/n,k/n]$, we obtain a bound of $h_n\cdot O(n^{-1}h_n^{-1})$. Hence the overall bound for the Riemann sum approximation is $O(n^{-1}\,h_n^{-2})$, uniformly in x. Therefore,

$$\begin{aligned}&\sqrt{n}\int_U (F_n^{-1}(x)-F^{-1}(x))n^{-1}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}h_n^{-2}K'\left(\frac{k/n-x}{h_n}\right)\frac{G(k/n)}{q(k/n)}\,dx - \sqrt{n}\int_U (F_n^{-1}(x)-F^{-1}(x))\int_U h_n^{-2}K'\left(\frac{u-x}{h_n}\right)\frac{G(u)}{q(u)}\,du\,dx\\ &\quad\le \sqrt{n}\int_U |F_n^{-1}(x)-F^{-1}(x)|\,dx\cdot O(n^{-1}h_n^{-2})\end{aligned}$$

and the random quantity converges weakly (again by Gihman and Skorohod, 1980, p. 437), hence the total error is $O_P(n^{-1}\,h_n^{-2})$, which tends to zero. This concludes the second step of the proof. Next, we re-express the inner integral, using integration by parts:

$$\int_a^b h_n^{-2}K'\left(\frac{u-x}{h_n}\right)\frac{G(u)}{q(u)}\,du = h_n^{-1}K'\left(\frac{b-x}{h_n}\right)\frac{G(b)}{q(b)} - h_n^{-1}K'\left(\frac{a-x}{h_n}\right)\frac{G(a)}{q(a)} - \int_a^b h_n^{-1}K\left(\frac{u-x}{h_n}\right)\left(\frac{g(u)}{q(u)}-\frac{G(u)q'(u)}{q^2(u)}\right)du.$$

Integrating against $\sqrt{n}(F_n^{-1}(x)-F^{-1}(x))$ over $x\in U$ yields

$$\sqrt{n}(\widehat{Q}_n(b)-\tilde{Q}_n(b))\frac{G(b)}{q(b)} - \sqrt{n}(\widehat{Q}_n(a)-\tilde{Q}_n(a))\frac{G(a)}{q(a)} - \sqrt{n}\int_U(\widehat{Q}_n(u)-\tilde{Q}_n(u))\left[\frac{g(u)}{q(u)}-\frac{G(u)q'(u)}{q^2(u)}\right]du.$$

At this point, we utilize Theorem 4 and apply integration against $b(u)$ over U to the convergence result, where

$$b(u) = \Delta_b(u)\frac{G(u)}{q(u)} - \Delta_a(u)\frac{G(u)}{q(u)} - \left[\frac{g(u)}{q(u)} - \frac{G(u)q'(u)}{q^2(u)}\right]$$

and $\Delta_x(u)$ denotes the Dirac delta function at x. (Observe that evaluation at a point is a continuous functional, which amounts to integration against the Dirac delta function at that point.) Writing out $b(u)q(u)W(u)$, we obtain the stated result. This limiting stochastic integral can also be rewritten as

$$\int_a^b \frac{G(u)}{q(u)}\, d[q(u)W(u)]. \tag{B.3}$$

Using the fact that $W(u)=B(u)-uB(1)$ for Brownian Motion $B(u)$, the integrating measure can also be written as

$$q'(u)B(u)\,du + q(u)dB(u) - B(1)(uq'(u)+q(u))\,du.$$

Now computing the variance of (B.3) yields (A.6). □

Proof of Theorem 2. Following on from the proof of Theorem 1, we have

$$\sqrt{n}(\hat{v}_0 - v_0) = o_P(1) + e_1' M^{-1}(\tilde{p}_0)\left[n^{-1/2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor} \varepsilon(k/n)G^*(k/n), \ldots, n^{-1/2}2\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\varepsilon(k/n)G_{\tilde{p}_0}(k/n)\right]',$$

since $X'X/n = o(1)+M(\tilde{p}_0)$ (in the sense that each entry converges). By (16), the expression on the right is simply

$$n^{-1/2}\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor}\varepsilon(k/n)G(k/n).$$

Now since g and G are uniformly bounded on U, our assumptions validate the hypotheses of Theorem 5, and hence applying that result completes the proof. □

Proof of Theorem 3. We focus on the v_0 case, v_1 being similar. Let $B(u) = \log L_0(u) - \log L_0^{(\tilde{p}_0)}(u)$ be the deterministic bias, and $\underline{B}$ the vector of $B(u_j)$ values. Then it follows that

$$\hat{v}_0 - v_0 = e_1'(X'X)^{-1}X'(\underline{\varepsilon} + \underline{B}).$$

As in the proof of Theorem 1, we have the matrix $M_n(\tilde{p}_0) := n^{-1}X'X$, which can be partitioned into

$$M_n(\tilde{p}_0) = \begin{bmatrix} \gamma_n & \omega_n'(\tilde{p}_0) \\ \omega_n(\tilde{p}_0) & D_n(\tilde{p}_0)\end{bmatrix},$$

where the jth component of $\omega_n(\tilde{p}_0)$ is $\langle \log, G_{j-1}\rangle_n$ for $1 \le j \le \tilde{p}_0$, and $D_n(\tilde{p}_0)$ has jkth entry $\langle G_{j-1}, G_{k-1}\rangle_n$. Also $\gamma_n = \langle \log,\log\rangle_n$. Let the Schur complement be $S_n(\tilde{p}_0) = \gamma_n - \omega_n'(\tilde{p}_0)D_n^{-1}(\tilde{p}_0)\omega_n(\tilde{p}_0)$, so that

$$e_1'M_n^{-1}(\tilde{p}_0)X' = S_n^{-1}(\tilde{p}_0)(G^* - \omega_n'(\tilde{p}_0)D_n^{-1}(\tilde{p}_0)G^{(\tilde{p}_0)}),$$

using Axelsson (1996, p. 93) to compute the inverse matrix. Here $G^{(\tilde{p}_0)}$ is the matrix consisting of the bottom $\tilde{p}_0+1$ rows of X'. Now $\langle \log, G_j\rangle_n = \sum_{k\ge 0}\alpha_k\langle G_k, G_j\rangle_n$, from which it follows that

$$\omega_n(\tilde{p}_0) = D_n(\tilde{p}_0)\alpha(\tilde{p}_0) + \tilde{D}_n(\tilde{p}_0)\alpha_{-\tilde{p}_0}(\infty),$$

where $\alpha(\tilde{p}_0) = (\alpha_0, \ldots, \alpha_{\tilde{p}_0})'$, $\alpha_{-\tilde{p}_0}(\infty) = (\alpha_{\tilde{p}_0+1}, \ldots)'$, and $\tilde{D}_n(\tilde{p}_0)$ is a matrix of dimension $\tilde{p}_0+1$ by infinity, with jkth element equal to $\langle G_{j-1}, G_{k+\tilde{p}_0}\rangle_n$. Letting $\nu_n(\tilde{p}_0) = \tilde{D}_n(\tilde{p}_0)\alpha_{-\tilde{p}_0}(\infty)$, we see that its jth component is $O(n^{-1}\log\tilde{p}_0)$ using the Riemann approximation and the orthogonality of $\{G_k\}$, and the fact that $|\alpha_k| = O(k^{-1})$. Then we obtain

$$G^* - \omega_n'(\tilde{p}_0)D_n^{-1}(\tilde{p}_0)G^{(\tilde{p}_0)} = \sum_{k>\tilde{p}_0}\alpha_k G_k - \nu_n'(\tilde{p}_0)D_n^{-1}(\tilde{p}_0)G^{(\tilde{p}_0)}.$$

Assuming that $\tilde{p}_0 = o(n)$, we approximate $D_n(\tilde{p}_0)$ by $2(b-a)$ times the identity matrix (id_n), plus a matrix $E_n(\tilde{p}_0)$ with every entry $O(n^{-1})$. Since $D_n^{-1}(\tilde{p}_0) = .5(b-a)^{-1}id_n + \sum_{l\ge 1}(-.5E_n(\tilde{p}_0)/(b-a))^l$, we obtain $\nu_n'(\tilde{p}_0)D_n^{-1}(\tilde{p}_0)G^{(\tilde{p}_0)} = O(\tilde{p}_0\log\tilde{p}_0/n)$. Next turning to $S_n(\tilde{p}_0)$, we have $\gamma_n = O(n^{-1}) + 2(b-a)\sum_{j\ge 0}\alpha_j^2$. Also

$$\omega_n'(\tilde{p}_0)D_n^{-1}(\tilde{p}_0)\omega_n(\tilde{p}_0) = \omega_n'(\tilde{p}_0)\alpha(\tilde{p}_0) + \omega_n'(\tilde{p}_0)D_n^{-1}(\tilde{p}_0)\nu_n(\tilde{p}_0),$$

and the second term on the right hand side is $O(\tilde{p}_0\log\tilde{p}_0/n)$. The first term is $O(n^{-1})$ plus $2(b-a)\sum_{j=0}^{\tilde{p}_0}\alpha_j^2$, so that

$$S_n(\tilde{p}_0) = O(\tilde{p}_0\log\tilde{p}_0/n) + 2(b-a)\sum_{j>\tilde{p}_0}\alpha_j^2.$$

In summary, we have

$$e_1' M_n^{-1}(\tilde{p}_0) X' = \frac{\mathrm{O}(\tilde{p}_0 \log \tilde{p}_0/n) + \sum_{k>\tilde{p}_0} \alpha_k G_k}{\mathrm{O}(\tilde{p}_0 \log \tilde{p}_0/n) + 2(b-a)\sum_{j>\tilde{p}_0} \alpha_j^2},$$

which is averaged over u_j against $\underline{\varepsilon}$ and $\underline{B}$. For the stochastic term, we have

$$\left\langle \sum_{k>\tilde{p}_0} G_k \alpha_k, \varepsilon \right\rangle_n \le \sqrt{\langle \varepsilon, \varepsilon \rangle_n} \sqrt{\left\langle \sum_{k>\tilde{p}_0} G_k \alpha_k, \sum_{k>\tilde{p}_0} G_k \alpha_k \right\rangle_n}$$

by the Cauchy–Schwarz inequality. The first term is $\mathrm{O_P}(B(q;K_n)+d_n)$ as in the proof of Theorem 1, and the second term is $\mathrm{O}(n^{-1})$ plus a term tending to zero as $\tilde{p}_0 \to \infty$. For the deterministic bias term, we have

$$\left\langle \sum_{k>\tilde{p}_0} G_k \alpha_k, B \right\rangle_n = \mathrm{O}(n^{-1}) + 2(b-a) \sum_{k>\tilde{p}_0} \alpha_k \theta_{k,0}.$$

This means that the asymptotic bias is $\sum_{k>\tilde{p}_0} \alpha_k \theta_{k,0} / \sum_{k>\tilde{p}_0} \alpha_k^2$, so long as all the error terms vanish; but this is true since $\tilde{p}_0$ is chosen such that $\tilde{p}_0 \log \tilde{p}_0/n \to 0$ as $n \to \infty$. Now the Cauchy–Schwarz inequality shows that the numerator of the bias is a finite sum for any $\tilde{p}_0$, and thus both numerator and denominator tend to zero as $\tilde{p}_0$ tends to infinity. Applying L'Hopital's rule yields the asymptote of $\theta_{\tilde{p}_0,0}/\alpha_{\tilde{p}_0}$. Finally, since $\alpha_k = \mathrm{O}(1/k)$ and $\log L_0(u)$ is continuously differentiable, it follows that $\theta_{k,0} = \mathrm{o}(1/k)$ and the bias tends to zero. This completes the proof of consistency. □

References

Axelsson, O., 1996. Iterative Solution Methods. Cambridge University Press, New York.
Cheng, C., 1995. Uniform consistency of generalized kernel estimators of quantile density. Ann. Statist. 23, 2285–2291.
Cheng, C., Parzen, E., 1997. Unified estimators of smooth quantile and quantile density functions. J. Statist. Plann. Inference 59, 291–307.
Csörgo, S., Deheuvels, P., Mason, D., 1985. Kernel estimates of the tail index of a distribution. Ann. Statist. 13, 1050–1077.
de Haan, L., 1981. Estimating the minimum of a function using order statistics. J. Amer. Statist. Assoc. 76, 467–469.
de Haan, L., Resnick, S., 1980. A simple asymptotic estimate for the index of a stable distribution. J. Roy. Statist. Soc. Ser. B 42, 83–87.
Dekkers, A., Einmahl, J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. Ann. Statist. 17, 1833–1855.
Drees, H., 1995. Refined Pickands estimators of the extreme value index. Ann. Statist. 23, 2059–2080.
Drees, H., 1998. Optimal rates of convergence for estimates of the extreme value index. Ann. Statist. 26, 434–448.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.
Falk, M., 1985. Asymptotic normality of the kernel quantile estimator. Ann. Statist. 13, 428–433.
Falk, M., 1986. On the estimation of the quantile density function. Statist. Probab. Lett. 4, 69–73.
Gihman, I., Skorohod, A., 1980. The Theory of Stochastic Processes. Springer, Berlin.
Gomes, M., Martins, M., 2001. Generalizations of the Hill estimator—asymptotic versus finite sample behaviour. J. Statist. Plann. Inference 93, 161–180.
Hall, P., Welsh, A.H., 1985. Limit theorems for the median deviation. Ann. Inst. Statist. Math. 37, 27–36.
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., 2009. Tail exponent estimation via broadband log density-quantile regression. SRD Research Report No. 2009/2, U.S. Census Bureau, available at 〈http://www.census.gov/srd/www/byname.html〉.
Hurvich, C., Brodsky, J., 2001. Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. J. Time Ser. Anal. 22, 221–249.
Karatzas, I., Shreve, S., 1997. Brownian Motion and Stochastic Calculus. Springer, New York.
Mallat, S., 2001. A Wavelet Tour of Signal Processing. Academic Press, New York.
McCulloch, J., 1997. Measuring tail thickness to estimate the stable index α: a critique. J. Bus. Econom. Statist. 15, 74–81.
McElroy, T., Politis, D., 2007. Moment based tail index estimation. J. Statist. Plann. Inference 137, 1389–1406.
McElroy, T., Politis, D., 2008. A fine-tuned estimator of a general convergence rate. Austal. New Zealand J. Statist. 50, 1–11.
Parzen, E., 1979. Nonparametric statistical data modeling. J. Amer. Statist. Assoc. 74, 105–121.
Parzen, E., 1982. Data modeling using quantile and density-quantile functions. In: Tiago, J., Epstein, B. (Eds.), Some Recent Advances in Statistics. Academic Press, New York, pp. 23–52.
Parzen, E., 2004. Quantile probability and statistical data modeling. Statist. Sci. 19, 652–662.
Pickands III, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131.
Politis, D., Romano, J., Wolf, M., 1999. Subsampling. Springer-Verlag, New York.
Rojo, J., 1996. On tail categorization of probability laws. J. Amer. Statist. Assoc. 91, 378–384.
Schuster, E., 1984. Classification of probability laws by tail behavior. J. Amer. Statist. Assoc. 79, 936–939.
Shao, J., Tu, D., 1995. The Jackknife and the Bootstrap. Springer-Verlag, New York.
Teugels, J., 1981. Limit theorems on order statistics. Ann. Statist. 9, 868–880.
Xiang, X., 1994. A law of the logarithm for kernel quantile density estimators. Ann. Probab. 22, 1078–1091.