RESEARCH REPORT SERIES
(Statistics #2007-1)

Semiparametric Tail Index Estimation: A Density Quantile Approach

Scott Holan* and Tucker McElroy

University of Missouri-Columbia*

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: January 24, 2007

Semiparametric Tail Index Estimation: A Density Quantile Approach

Scott Holan* and Tucker McElroy†
University of Missouri-Columbia and U.S. Census Bureau

Abstract

Heavy tail probability distributions are important in many scientific disciplines, such as hydrology, geology, and physics among others. To this end many heavy tail distributions are commonly used in practice. In order to determine an appropriate family of distributions for a specified application it is useful to classify the probability law via its tail behavior. Through the use of Parzen's density-quantile function, this work proposes a semiparametric estimator of the tail index. The method we develop is useful when little or nothing is known about the distribution a priori. Furthermore the approach we develop allows for separate estimates of the left and right tail indices. In the development of the asymptotic theory of the tail index estimator we provide results of independent interest that may be used to establish weak convergence of stochastic processes. Finally, we present theoretical properties for the tail index estimator and explore its finite sample properties through simulation.

Keywords. Density quantile, Extreme-value theory, Quantile density, Semiparametric, Tail exponent.

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. Thus, it is important to characterize the probability law by classifying its tail behavior. To this end many tail index estimators have been proposed.

*Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO, 65211-6100, holans@missouri.edu

†Statistical Research Division, U.S. Census Bureau, 4700 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

Perhaps the most popular estimator is due to Hill (1975). This method provides a robust estimator based on the asymptotics of extreme values. However, when applied to stable data this estimator can give misleading results, see McCulloch (1997) for further discussion. Another widely used index for classifying tail behavior is the Pickands estimator, see Pickands (1975). Although this estimator is easy to compute and invariant to certain shift and scale transformations, it suffers from poor asymptotic efficiency. Several refinements have been suggested for both estimators, see for example Gomes and Martins (2001) and Drees (1995).

In addition to the Hill and Pickands estimators, and their refinements, many other methods for classifying tail behavior have been proposed. One example of an alternative method is provided by Csörgo et al. (1985). In this paper the authors develop an estimate that is expressed as the convolution of a kernel with the logarithm of the quantile function. Additionally, the estimator they suggest includes both the Hill estimators and de Haan estimator as special cases. For a complete discussion of the de Haan estimator see Dekkers et al. (1989). 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 general nonparametric model. The estimator that they consider assumes that the only available information is in the form of asymptotic properties of the distributions tails. Additionally, recent research has provided many other contributions to the area of tail index estimation; for a comprehensive discussion see Embrechts et al. (1997) and the references therein.

In contrast to the methods previously described, Parzen (1979) suggests an alternative approach for determining probability distributions by assessing their tail behavior. Specifically, Parzen introduces the density-quantile function and uses it as a measure of tail orderings. Subsequently, Schuster (1984) refines Parzen's classification scheme and provides a connection with the limit in probability of extreme spacings. Finally, Rojo (1996) develops an approach that relaxes the smoothness conditions required in Schuster (1984).

This paper provides a new semiparametric estimator for classifying tail behavior. The technique we develop uses the logarithm of the density-quantile function and provides a natural framework for separately estimating both the left and right tail index. This paper is organized as follows. In Section Two 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^c and α^p, the classical tail index and the tail index proposed by Parzen respectively. Section Three contains theoretical results that establish the asymptotic behavior of the tail index estimator. Specifically, we provide results for the consistency and asymptotic normality of the tail index estimator. The methodology we develop is tested in Section Four; simulations provide an indication of the mean square error of our estimator for finite sample sizes and different underlying distributions. Section Five contains an empirical study, and Section Six has a discussion. For convenience of exposition all proofs are left to the appendix.

2 Tail Exponents and Indices

Parzen (1979) discusses an approach to classifying tail behavior or probability laws. The method he suggests 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 distribution function and let Q denote the quantile function, then $F(Q(u)) = u$ for all u. Thus, by taking derivatives, it follows that

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

Here $f(Q(u)) = fQ(u)$ is defined to be the density quantile function and $q(u) = Q'(u)$ is defined to be the quantile density. Therefore, (1) implies

$$fQ(u) = \frac{1}{q(u)}. \tag{2}$$

Furthermore, let $J(u)$ denote the score function where

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

Following Parzen (2004) we assume, in practice, the representation near 0 and 1 to be given by regularly varying functions

$$fQ(u) = u^{\alpha_0} L(u) \qquad u \in [0, 1/2) \tag{3}$$

$$fQ(1-u) = u^{\alpha_1} L(u), \qquad u \in (1/2, 1] \tag{4}$$

where $L(u)$ is a slowly varying function. That is, $L(u)$ satisfies the condition that for a fixed $y > 0$

$$\frac{L(yu)}{L(u)} \longrightarrow 1 \quad \text{as } u \longrightarrow 0.$$

Note that in Parzen (2004), the above relations only hold asymptotically as $u \to 0$ and $u \to 1$, respectively. However, by redefining L we can easily obtain the exact relations (3) and (4). In this context we call α_0 and α_1 the left and right tail exponents respectively and they are used as a measure of tail behavior. Parzen (1979) shows that

$$\alpha_0 = \lim_{u \to 0^+} -uJ(u)/fQ(u), \tag{5}$$

$$\alpha_1 = \lim_{u \to 1^-} (1-u)J(u)/fQ(u), \tag{6}$$

and classifies tail behavior as short ($\alpha < 1$), medium ($\alpha = 1$) and large ($\alpha > 1$). Subsequently, Schuster (1984) further divides the medium class to include medium-short, medium-medium, and medium-long.

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 $\alpha_0 = \alpha_1 = 0$. 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 index is $\alpha_1 = 1$. Finally, the Cauchy distribution has $F(x) = (\arctan x)/\pi$, and the density quantile is $fQ(u) = \frac{1}{\pi}\cos^2(\pi u)$. Using a Taylor series expansion, we find that $\alpha_0 = \alpha_1 = 2$.

The estimator we develop begins with the representations in (3) and (4). More specifically, we define $L(u)$ by

$$L(u) = \exp\left\{\theta_0 + 2\sum_{k=1}^{\infty}\theta_k \cos(2\pi k u)\right\},$$

and using obvious notation we write

$$L(u, p) = \exp\left\{\theta_0 + 2\sum_{k=1}^{p}\theta_k \cos(2\pi k u)\right\}. \tag{7}$$

It then follows easily that

$$\lim_{u\to 0}\frac{L(yu, p)}{L(u, p)} = 1,$$

and thus $L(u, p)$ constitutes a slowly varying function. Additionally, since the system $C = \{1, 2\cos(2\pi u), 2\cos(2\pi 2u), \ldots\}$ (the Fourier representation) is complete for the class of functions on C$[0, 1]$, the continuous functions on $[0, 1]$, $L(u, p)$ converges to $L(u)$ in mean square as $p \longrightarrow \infty$, see Hart (1997). That is, the system C forms an orthogonal basis for C$[0, 1]$. The significance of defining $\log L(u)$ in terms of its Fourier representation is that it provides a method of estimating the tail index without having to specify a functional form for $L(u)$ a priori.

In order to expand the utility of our estimator we provide a mapping between the "classical" and "Parzen" tail index estimators. The equivalence formula we derive presupposes that the distribution under consideration is symmetric. To this end, consider a heavy-tailed random variable X of index $\alpha > 0$. This is defined as follows. First, let $F(x)$ denote the cdf and $G(x) = 1 - F(x)$. Then

$$F(-x) = \frac{c_1 + o(1)}{x^\alpha}L(x) \qquad G(x) = \frac{c_2 + o(1)}{x^\alpha}L(x)$$

as $x \to \infty$, where c_1 and c_2 are non-negative (and not both zero) and L is a slowly varying function; compare with Embrechts et al. (1997, p.75). Further, suppose that the probability density function f is ultimately monotone, i.e., it is monotone on (z_1, ∞) and $(-\infty, z_2)$ for some z_1 and z_2. Then by Theorem A3.7 of Embrechts et al. (1997, p.568), we have

$$f(x) \sim c_1\alpha x^{-(\alpha+1)}L(x) \quad \text{as } x \to \infty,$$
$$f(x) \sim c_2\alpha(-x)^{-(\alpha+1)}L(-x) \quad \text{as } x \to -\infty.$$

Now, Parzen's left and right tail exponents are given by the limits

$$\alpha_0 = \lim_{u \to 0^+} \frac{\log fQ(u)}{\log u},$$
$$\alpha_1 = \lim_{u \to 1^-} \frac{\log fQ(1-u)}{\log u}.$$

This can be seen by considering the definition of the exponents α_0 and α_1 given in Parzen (2004):

$$fQ(u) \sim u^{\alpha_0} L(u), \tag{8}$$
$$fQ(1-u) \sim u^{\alpha_1} L(u), \tag{9}$$

where the asymptotics are as u tends to zero and one, respectively. Note that (8) and (9) slightly differ from the exact representations in (3) and (4) in that here we use the asymptotic formulation of Parzen (2004). Furthermore, by taking logarithms of (8) and (9) and dividing by $\log u$, we obtain the above expression by noting that

$$\log L(u)/\log u \to 0$$

as $u \to 0$, for any function L that is slowly-varying at zero. 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.

In what follows we focus on deriving the equivalence relation for the left index. The limit can be achieved by using the sequence $1/n$ as follows:

$$\alpha_0 = \lim_{u \to 0} \frac{\log fQ(u)}{\log u} = -\lim_{n \to \infty} \frac{\log fQ(1/n)}{\log n}$$

Let $a_n = Q(1/n)$, with $Q(x)$ equal to the quantile function. Then we can write $a_n = n^{1/\alpha} K(n)$ for some slowly-varying function $K(x)$ (see Embrechts et al. (1997, p. 78) for a similar statement). It then follows that

$$\begin{aligned} fQ(1/n) &\sim c_1 \alpha a_n^{-(1+\alpha)} L(a_n) \\ &= \alpha n^{-(1+1/\alpha)} K(n)^{-1} \end{aligned}$$

so that

$$-\frac{\log fQ(1/n)}{\log n} = (1+1/\alpha) - \frac{\log \alpha}{\log n} + \frac{\log K(n)}{\log n}.$$

Thus $\alpha_0 = 1 + 1/\alpha$. For the right tail index, we use the sequence $u = 1 - 1/n$ and utilize the expression for G instead of F. Therefore, for the given heavy-tailed variable with equal right and left tail index, we have

$$\alpha_L = 1 + 1/\alpha = \alpha_1. \tag{10}$$

That is, for symmetric distributions, the relationship in (10) provides an equivalence between the classical approach and approaches using Parzen's density quantile function. The mapping between α^c and α^p is illustrated in the following examples.

Example 1 A stable variable has characteristic exponent $\nu \in (0, 2]$, with $\nu = 2$ corresponding to the Gaussian distribution. When $\nu < 2$, the stable variable is heavy-tailed with classical $\alpha = \nu$. Note that $\nu = 1$ corresponds to the Cauchy distribution. Thus, for the Cauchy the Parzen tail index 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 $x > 0$. The corresponding Parzen tail indexes are $\alpha_1 = 1 + 1/\alpha$ and $\alpha_0 = 1$. 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 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. Roughly speaking, they correspond to $\alpha = \infty$, since their tails decay faster than any polynomial power of x. This maps into a Parzen tail index of 1.

3 Tail Exponent Estimators

In order to estimate α_0 and α_1 we consider $\log fQ(u)$ for $u \in (0, u_L]$ and $\log fQ(1-u)$ for $u \in [u_R, 1)$ respectively. Now using (7) we write

$$\log fQ(u) = \alpha_0 \log(u) + \theta_0 + 2\sum_{k=1}^{p} \theta_k \cos(2\pi k u),$$

and

$$\log fQ(1-u) = \alpha_1 \log(u) + \theta_0 + 2\sum_{k=1}^{p} \theta_k \cos(2\pi k u).$$

Further, let $\widetilde{q}(u_j)$ denote an estimator of $q(u)$ obtained from the data, where $u_j = (j - .5)/n$. Then, using (2), $\widetilde{f}Q(u_j) = 1/\widetilde{q}(u_j)$ and the form for $\log \widetilde{f}Q(u_j)$ follows. For the remainder of this section we consider only left tail index estimation (i.e. α_0) and note that right tail index estimation follows analogously.

Let $y = \log \widetilde{f}Q(\underline{\boldsymbol{u}})$ be given, where $\underline{\boldsymbol{u}} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_L)'$ and $G_k = \cos(2\pi k \underline{\boldsymbol{u}})$. Then for any positive integer p, define $X = [G^*, G_0, 2G_1, \ldots, 2G_p]$ where $G^* = \log(\underline{\boldsymbol{u}})$. Finally, define

$\widehat{\beta}_p = (\widehat{\alpha}_0, \widehat{\theta}_0, \widehat{\theta}_1, \ldots, \widehat{\theta}_p)' = (X'X)^{-1}X'y$ which can be viewed as the ordinary least squares estimator of $\beta_p = (\alpha_0, \theta_0, \theta_1, \ldots, \theta_p)'$.

Given this framework, one question that naturally arises in practice is how to choose $\widetilde{q}(u_j)$, the estimated quantile density function (qdf). There have been many research efforts aimed at estimating the qdf, see Cheng and Parzen (1997), Xiang (1994) and Falk (1986) for a more detailed discussion. In what follows we provide some examples of possible qdf estimators.

3.1 Sample spacings

Consider a sample $X_1, X_2, \ldots, X_n$ with order statistics $X_{(1;n)} < X_{(2;n)} < \cdots < X_{(n;n)}$ then one estimator of $q(u)$ is expressed in terms of the sample spacings

$$\widetilde{q}(u_j) = n\{X_{(j+1;n)} - X_{(j;n)}\}, \tag{11}$$

for $u_j = (j - 0.5)/n$ ($j = 1, 2, \ldots, n-1$), see Parzen (1982). Using (11) as a starting point in the estimation of α_0 provides an asymptotically unbiased estimate. However, the variability in this estimate is unsatisfactory contributing to a large mean squared error. Therefore, we find this estimator unsatisfactory for use in finite sample sizes.

3.2 Kernel quantile density estimators

Kernel quantile density estimators provide a rich class of estimators for the qdf and were first introduced by Parzen (1979). Specifically, let $Q(u) = \inf\{x : F(x) \geq u\}$, $0 < u < 1$, be the quantile function associated with $F(x)$ and, as before, $q(u) = Q'(u)$ the quantile density function. Then, one expression for the kernel quantile density estimator is

$$\widehat{q}_n(t) = \frac{1}{h_n^2} \int_0^1 F_n^{-1}(x) K'\left(\frac{x-t}{h_n}\right) dx \tag{12}$$

for some kernel K where $F_n^{-1}(x) = \inf\{u : F_n(u) \geq x\}$, see Xiang (1994). Moreover, Xiang (1994) suggests the quantile density estimator

$$\widehat{q}_n(t) = \frac{1}{nh_n^2} \sum_{i=1}^{n} K'\left(\frac{i/n - t}{h_n}\right) X_{(i)}$$

as an easier to calculate alternative to (12).

In the sequel we consider a kernel smoothed estimator $\widehat{q}(u)$, as a starting point for tail index estimation, that satisfies assumptions $K_1 - K_7$ of Cheng (1995). Specifically, one estimator we consider is the boundary-modified Bernstein polynomial. Let ϵ be such that $U \subset [\epsilon, 1-\epsilon] \subset (0, 1)$,

$L_\epsilon = 1-2\epsilon$ and $t_j = \epsilon + (j/k)L_\epsilon$, $j = 0, 1, \ldots, k$. Then the k^{th} degree boundary-modified Bernstein polynomial qdf estimator on U can be expressed as

$$\widehat{q}_n^B = \frac{1}{L_\epsilon^k} \sum_{j=0}^{k-1} \frac{\widetilde{Q}_n(t_{j+1}) - \widetilde{Q}_n(t_j)}{1/k} \binom{k-1}{j} (u-\epsilon)^j (1-\epsilon-u)^{k-1-j}, \tag{13}$$

where $\widetilde{Q}_n(t_j)$ is the sample quantile. Letting $k = k_n \uparrow \infty$ as $n \uparrow \infty$, Cheng (1995) shows assumptions $K_1 - K_7$ are satisfied. Finally, with $y = \log \widehat{f}Q(u) = -\log \widehat{q}_n^B(u)$ we can form an estimate of α_0 via $(X'X)^{-1}X'Y$ as before.

Remark 1 *Using the boundary-modified Bernstein polynomial requires the choice of user-selected parameters. Additionally, the values $k = \lceil \delta n \rceil$, for $\delta = .95$ and $.99$, performed well in simulation and satisfy the necessary assumptions.*

4 Asymptotic Results

When estimating the qdf, in the context of tail index estimation, the issue of how to choose the percentiles $\underline{\boldsymbol{u}}$ arises. For the asymptotic results below, we will suppose them to be of the form $\boldsymbol{u_k} = k/n$, with k ranging between 1 and $\lfloor nx \rfloor$, where $x \in (0, 1/2]$ is a user-selected parameter. For many situations, $x = 1/2$ is appropriate; however, lower values may be selected.

For the following consistency result, we suppose that the quantile-density function $q(u)$ is estimated with a kernel-smoothed estimator $\widehat{q}(u)$, as in Cheng (1995). The kernel that such an estimator relies upon must satisfy some basic assumptions, such as K_1 through K_7 of Cheng (1995). One example of such an estimator is given in (13). Additionally, some regularity conditions on the quantile-density are also necessary: assumptions Q_1 through 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 $\sup_{u \in (0,1)} u(1-u)|J(u)|/fQ(u) \leq \gamma$.

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 $(u^*, 1)$.

Taking the lower percentiles, we have $q(u) = u^{-\alpha_0}/L(u)$ for $u < 1/2$, so that Q_1 is satisfied if L is twice differentiable in $(0, 1/2)$. Q_2 is automatically satisfied, given that the limits (5) and (6) exist. Q_3 may or may not be satisfied in general, depending on the form of $L(u)$. Certainly, the assumption of Q_1 and Q_3 places no burdensome restriction on the slowly varying function $L(u)$.

Again, since $\log fQ(u) = -\log q(u)$, we let $\log \widehat{f}Q(u) = -\log \widehat{q}(u)$, and consider percentiles of the form $\boldsymbol{u}_k = k/n$, with $1 \leq k \leq \lfloor nx \rfloor$. Thus we can write regression equations for both the left-hand and right-hand tail indexes, i.e., α_0 and α_1:

$$\log \widehat{f}Q(\boldsymbol{u}_k) = \alpha_0 \log(\boldsymbol{u}_k) + \theta_0 + 2\sum_{t=1}^{p} \theta_t \cos(2\pi t \boldsymbol{u}_k) + \epsilon(\boldsymbol{u}_k)$$

$$\log \widehat{f}Q(1-\boldsymbol{u}_k) = \alpha_1 \log(\boldsymbol{u}_k) + \theta_0 + 2\sum_{t=1}^{p} \theta_t \cos(2\pi t \boldsymbol{u}_k) + \epsilon(1-\boldsymbol{u}_k),$$

where $\epsilon(u) = -\log\{\widehat{q}(u)/q(u)\}$. Note that in this formulation, we have substituted $L(u,p)$ for $L(u)$, which involves some approximation error. The asymptotic result ignores this substitution since the deterministic approximation error can be made as small as desired by choosing p sufficiently large.

Theorem 1 *Suppose that the density-quantile function $q(u)$ satisfies Q_1 through Q_3, and we construct a kernel-smoothed estimator $\widehat{q}(u)$ with kernel satisfying K_1 through K_7 of Cheng (1995). Moreover, suppose that we consider each regression with the percentiles restricted to some closed subset $U = [a,b]$, with regressor functions chosen such that $(X'X)/n$ converges to an invertible matrix. Then the estimates $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ are consistent at rate $B(q;K_n) + d_n$, where these are defined in Theorem* 2.1 *of Cheng (1995). In particular, a simple convolution kernel yields a best rate of $n^{-2/5}\log n^{2/5}$.*

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 (12). Furthermore, to establish the result we need the following additional notation: let $G^*(u) = \log(u)$ and $G_k(u) = \cos(2\pi k u)$ for $k = 0, \cdots, p$. Also let the $(w^*, w_0, \cdots, w_p)$ denote the first row of the limiting inverse matrix of $(X'X)/n$. Then we define

$$G(u) = w^* G^*(u) + w_0 G_0(u) + \cdots + w_p G_p(u). \tag{14}$$

For convenience, we formulate the result in terms of a general estimate $\widehat{\alpha}$, which is either $\widehat{\alpha}_0$ or $\widehat{\alpha}_1$ depending on the choice of a and b.

Theorem 2 *(Asymptotic Normality) Suppose the same assumptions as in Theorem 1, and in addition that the kernel satisfies assumption K_8 and (20). Let $G(u)$ be given by (14), and assume that its derivative $g(u) = G'(u)$ exists, with both g and G uniformly bounded on U. Let h_n be chosen such that $nh_n^2 \to \infty$ but $h_n \to 0$ as $n \to \infty$. Then*

$$\sqrt{n}(\widehat{\alpha} - \alpha) \stackrel{\mathcal{L}}{\Longrightarrow} G(b)W(b) - G(a)W(a) - \int_a^b W(u)\left(g(u) - G(u)\frac{q'(u)}{q(u)}\right) du,$$

where $W(u)$ is a Brownian Bridge.

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). First, let us define the concept of the *modulus of continuity* on $[0,1]$:

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

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

$$\sup_{\omega\in A} |\omega(0)| < \infty \tag{15}$$

$$\lim_{\delta\downarrow 0} \sup_{\omega\in A} m^T(\omega,\delta) = 0 \quad \textit{for every} T \in (0,1). \tag{16}$$

In condition (15), the time point 0 can be replaced by the time point 1.

Next we consider an adaption of Theorem 4.10 of Karatzas and Shreve (1997). By $\mathcal{B}(C[0,1])$ we denote the σ-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 $\epsilon > 0$ there exists a compact set K in $C[0,1]$ such that $P_n(K) \ge 1-\epsilon$ 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*

$$\lim_{\lambda\uparrow\infty} \sup_{n\ge 1} P_n[\omega : |\omega(0)| > \lambda] = 0, \tag{17}$$

$$\lim_{\delta\downarrow 0} \sup_{n\ge 1} P_n[\omega : m^T(\omega,\delta) > \epsilon] = 0 \quad \forall T \in (0,1),\, \epsilon > 0. \tag{18}$$

In condition (17), the time point 0 can be replaced by the time point 1.

The preceding Propositions 1 and 2 are of general interest, and may be used to establish the weak convergence of stochastic processes. In what follows, we consider the kernel quantile estimator related to (12)

$$\widehat{Q}_n(t) = \int_0^1 F_n^{-1}(x) h_n^{-1} K\left(\frac{t-x}{h_n}\right) dx,$$

which is introduced in Falk (1985). In a like manner, an approximation to the true $F_n^{-1}(t)$ is given by

$$\widetilde{Q}_n(t) = \int_0^1 F^{-1}(x) h_n^{-1} K\left(\frac{t-x}{h_n}\right) dx.$$

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

$$\sqrt{n}\left\{\left(\widehat{Q}_n(u_1) - \widetilde{Q}_n(u_1)\right), \cdots, \left(\widehat{Q}_n(u_d) - \widetilde{Q}_n(u_d)\right)\right\} \stackrel{\mathcal{L}}{\Longrightarrow} \{W_{u_1}, \cdots, W_{u_d}\}$$

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 $\sqrt{n}(\widehat{Q}_n(u) - \widetilde{Q}_n(u))$ 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 the following additional condition on the kernel K:

$$(K_8) \qquad \sup_{u \in U}\left|h_n^{-1} K\left(\frac{s-u}{h_n}\right) - h_n^{-1} K\left(\frac{t-u}{h_n}\right)\right| \leq C_n |t-s|^\beta$$

Here the C_n's are positive constants with $\sup_{n \geq 1} C_n = C$, a positive constant as well. The rate β can be any positive number. We also require an additional technical concept: let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which the random variables $X_1, X_2, \cdots$ are defined, and let P_n be the measure induced by $\sqrt{n}(\widehat{Q}_n(u) - \widetilde{Q}_n(u))$ on the space $(C(U), \mathcal{B}(C(U)))$.

Theorem 3 *Suppose that F^{-1} has bounded derivative on U, and suppose that $K(x)$ has bounded support on U, integrates to one, and satisfies condition (K_8). Then*

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

i.e., the induced measures P_n corresponding to $\sqrt{n}(\widehat{Q}_n(u) - \widetilde{Q}_n(u))$ 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

$$\frac{1}{n} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \log\left(\frac{\widehat{q}_n(k/n)}{q(k/n)}\right) G(k/n). \tag{19}$$

Later we will need the following deterministic approximation to $q(u)$:

$$\widetilde{q}_n(u) = \int_U F^{-1}(x) h_n^{-2} K'\left(\frac{u-x}{h_n}\right) dx.$$

The function $G(u)$ is a fairly arbitrary regressor function. We formulate a general theory for the asymptotics of expressions (19), which may then be applied to obtain the asymptotic of the tail index estimators. We require an additional assumption on the kernel $K(x)$:

$$|K''(x)| \leq C/|x| \tag{20}$$

for some constant $C > 0$, and $|x|$ sufficiently large. Our main theorem is stated below:

Theorem 4 *Suppose that the density quantile function $q(u)$ satisfies Q_1 through Q_3, and we construct a kernel-smoothed $\widehat{q}_n(u)$ with kernel satisfying K_1 through K_7 of Cheng (1995), as well as K_8 and (20) 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 $nh_n^2 \to \infty$ but $h_n \to 0$ as $n \to \infty$. Then*

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

$$\xrightarrow{\mathcal{L}} G(b)W(b) - G(a)W(a) - \int_a^b W(u)\left(g(u) - G(u)\frac{q^{'}(u)}{q(u)}\right) du,$$

where $W(u)$ is a Brownian Bridge.

5 Empirical Study

In contrast to *classical* tail index estimation, the theory we propose applies to distributions having both symmetric and asymmetric tails. To this end, the benefit of our method is fully realized when estimating a left or right tail index from a distribution having asymmetric tails. However, if one wishes to estimate the tail index of a symmetric distribution our method may still be implemented.

In order to evaluate the utility and finite sample performance of our estimators we present the results of a simulation study. The study we conducted, however, does not compare our estimator with other popular estimators such as the Hill or Pickands since those methods apply only to symmetric distributions. Such a comparison would not be entirely meaningful since the two methods are not strictly compatible. Further, we acknowledge that if one has prior knowledge that the distribution under consideration has symmetric tails that other methods slightly out-perform ours and thus might be advantageous.

Even though there are a few other methods for estimating left and right tail indices within the density quantile framework, see for example Rojo (1996), to our knowledge none of them presents the asymptotic distribution of the proposed estimator. Additionally, previous efforts at estimating left and right tail indices lack simulation studies evaluating their effectiveness in finite sample sizes. Therefore, we restrict our attention to the tail index estimators that we propose.

The simulation study we undertake uses (13) as an estimate of $q(u)$. Specifically, we choose the number of grid points (i.e. u_j) equal to the sample size, $\delta = .975$, $\epsilon = .01$, and $p = 2$. As noted previously, this choice of parameters does not constitute an *optimal* choice of "tuning" parameters. However, this choice performs well in practice and is thus used here as an illustration. Furthermore, in order to simulate data from a distribution of a specified tail index we utilize an α-stable distribution (symmetric tails) and calculate the tail index using our formulation from (10).

This procedure was carried out for several values of the tail index α, using 1000 repetitions of sample sizes 200 and 1000. Finally, we only consider the left tail index and note that the right tail index yielded similar results.

Table (1) shows the results of the empirical study for both sample size 200 and 1000. Even though the estimator produced a slight upward systematic bias the mean square errors were relatively small. One thing to notice is that the variance (of α^p) decreases as α^c increases. Thus, it seems that tail index estimation in the density quantile framework performs better when estimating lighter tailed distributions. The superior estimation is due to the fact that the low tail thickness corresponds to values of α^p close to unity, whereas α^c is tending to infinity. Specifically, it will be easier to estimate values close to one than around infinity. Thus, intuitively, the variance will tend to be lower. In summary our estimator improves for larger values of α^c.

Additionally, Figure (1) presents the histogram of the distribution of the studentized estimators for $\alpha^p = 2$ ($n = 1000$); for convenience the Normal(0,1) pdf is superimposed. This plot illustrates that the asymptotic distribution is achieved even in finite samples. Similar plots were constructed for the distribution of each studentized tail index in Table (1) (for $n = 1000$) with all results being similar and thus are not displayed.

6 Discussion

In this paper we developed a new method of tail index estimation. The approach we propose is semiparametric and 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 index estimation when little or nothing is known about the distribution a priori. The only implicit requirement we impose is that the data don't exhibit any strong dependence. If one wishes to estimate the tail index for dependent data, there are several methods; for example, see McElroy and Politis (2006) and the references therein.

In order to increase the utility of our estimator we provide a mapping between our estimator and the *classical* estimator. Additionally, under fairly mild conditions, we show that our tail index estimator is both consistent and asymptotically normal. 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.

To illustrate the finite sample performance of our estimator we provide the results of an empirical study. This study is facilitated by simulating from an α-stable distribution, making use of equivalence formula (10), and includes the mean, standard error and mean square error for several different values of α^p. The results indicate good performance for the estimator we develop. Additionally, our estimator improves (i.e. has both smaller variance and mean square error) for larger values of α^c and thus is better for getting at lighter tailed distributions. Furthermore, we display a histogram of the distribution of the studentized estimator with the pdf of a Normal(0,1) superimposed. This plot shows convergence to the asymptotic distribution even in finite sample sizes.

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 choices we do not provide optimal selection criteria.

Appendix

Proof of Theorem 1. We focus on the α_0 case, since the α_1 case is similar. It follows from basic linear regression that

$$\widehat{\alpha}_0 - \alpha_0 = e_1^{'}(X^{'}X)^{-1}X^{'}\underline{\epsilon}$$

with $e_1^{'} = (1, 0, \cdots, 0)$ and $\underline{\epsilon}$ the vector of $\epsilon(\boldsymbol{u}_k)$ such that the percentiles all lie in the set U. This amounts to considering $\boldsymbol{u}_k$ with $\lceil na \rceil \leq k \leq \lfloor nb \rfloor$. Let $\gamma = X(X^{'}X)^{-1}e_1$, so that

$$|\widehat{\alpha}_0 - \alpha_0| = \left| \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \gamma_j \epsilon(\boldsymbol{u}_j) \right| \leq \left(\sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \gamma_j^2 \right)^{1/2} \left(\sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \epsilon^2(\boldsymbol{u}_j) \right)^{1/2}$$

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:

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

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

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

This matrix is symmetric and is invertible by assumption. Thus it follows that

$$|\widehat{\alpha}_0 - \alpha_0| \leq C \left(n^{-1} \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \epsilon^2(\boldsymbol{u}_j) \right)^{1/2}$$

for some constant $C > 0$. Now

$$\epsilon(k/n) = -\log\left(1 + \frac{\widehat{q}(k/n) - q(k/n)}{q(k/n)}\right)$$

to which we can apply a Taylor Series expansion. Since k is chosen such that k/n is bounded away from 0 and 1, $q(k/n)$ is bounded away from zero. Now 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 $\sup_{\lceil na \rceil \leq k \leq \lfloor nb \rfloor} |\epsilon(k/n)| = O_P(B(q; K_n) + d_n)$ as well. Finally, it follows that

$$\left(n^{-1} \sum_{j=\lceil na \rceil}^{\lfloor nb \rfloor} \epsilon^2(\boldsymbol{u}_j) \right)^{1/2} = O_P(B(q; K_n) + d_n).$$

This proves the consistency of $\widehat{\alpha}$, given that the kernel is selected such that $B(q; K_n) + d_n \to 0$. The last assertion of the theorem follows from the discussion following Theorem 2.1 of Cheng (1995). □

Proof of Proposition 2. Assume (17) and (18). Fix $\eta > 0$, and consider any positive integer L, which defines T via $T = 1 - 1/(L+1)$. Additionally, choose $\lambda > 0$ such that

$$\sup_{n \geq 1} P_n[\omega : |\omega(0)| > \lambda] \leq \eta 2^{-L},$$

which is guaranteed by property (17). Next, for each positive integer k, choose $\delta_k > 0$ such that

$$\sup_{n \geq 1} P_n[\omega : m^T(\omega, \delta_k) > 1/k] \leq \eta 2^{-(L+k)},$$

which is guaranteed by property (18). Define the sets

$$A_L = \{\omega : |\omega(0)| \leq \lambda, m^T(\omega, \delta_k) \leq 1/k,\ k = 1, 2, \cdots\}$$

for $L = 1, 2, \cdots$. Also let $A = \bigcap_{L=1}^{\infty} A_L$. Now if $\omega \in A$, then $\sup_{\omega \in A} |\omega(0)| \leq \lambda$ and condition (15) is satisfied. Now since $m^T(\omega, \delta) \leq m^{T'}(\omega, \delta)$ if $T \leq T'$, it follows that $m^{1-1/(L+1)}(\omega, \delta_k) \leq 1/k$ implies $m^T(\omega, \delta_k) \leq 1/k$ for every $T \leq 1 - 1/(L+1)$. So if $\omega \in A$, then $m^T(\omega, \delta_k) \leq 1/k$ for $k = 1, 2, \cdots$ and for all $T \in (0, 1)$. Thus, for every $T \in (0, 1)$, $\sup_{\omega \in A} m^T(\omega, \delta_k) \leq 1/k$ for $k = 1, 2, \cdots$, which implies condition (16). Hence by Proposition 1, the set A has compact closure. But by the continuity of $m^T(\cdot, \delta)$ (see Karatzas and Shreve (1997, p.62)), each A_L is closed, and hence so is A.

In order to show tightness of $\{P_n\}_{n=1}^{\infty}$, we must demonstrate that $P_n(A) \geq 1 - \eta$ for all $n \geq 1$. Now $P_n(A_L) \geq 1 - \eta 2^{-L+1}$ is easily shown. Finally,

$$P_n(A) \geq 1 - \sum_{L=1}^{\infty} P_n(A_L^c) \geq 1 - \eta$$

where A_L^c denotes the complement of A_L. This proves the proposition. $\square$

Proof of Theorem 3. 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 (17), which becomes

$$\sup_{n \geq 1} \mathbb{P}\left[\sqrt{n}\left|\widehat{Q}_n(a) - \widetilde{Q}_n(a)\right| > \lambda\right] \to 0 \tag{A.1}$$

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

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

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

$$\left|\mathbb{P}[\sqrt{n}|\widehat{Q}_n(a) - \widetilde{Q}_n(a)| > \lambda] - \mathbb{P}[|W_a| > \lambda]\right| < \epsilon$$

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

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

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

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

tends to zero as $\delta \to 0$; this formulation follows (18) using the definition of induced measure. Now assuming (K_8), take any $\epsilon > 0$ and $\delta > 0$, it follows that

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

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

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

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

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

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

Proof of Theorem 4. The proof proceeds in three major steps. First, we apply a Taylor Series expansion to the logarithm. Second, we analyze the linearization of (19) and compute a Riemann sum approximation. Third, we apply continuous functionals to the resulting expression, utilizing Theorem 3 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 $0 < \delta < 2/5$ such that

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

Since G and $\widetilde{q}_n$ 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, i.e.,

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

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

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

the difference is given by

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

Now $\sqrt{n}(\widetilde{q}_n(u) - q(u))$ will tend to zero uniformly for $u \in U$ given a kernel $K^{'}$ with rapidly decaying tails, as discussed in Falk (1986), so we make the replacement as indicated. Next, we have

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

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

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

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

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

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 (20) 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,

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

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 simplify 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

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

At this point, we utilize Theorem 3 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, and the proof is complete. □

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

$$\sqrt{n}(\widehat{\alpha} - \alpha) = o_P(1) +$$
$$e_1' M^{-1}(a,b) \left[n^{-1} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \epsilon(k/n) G^*(k/n), \cdots, n^{-1} \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor} \epsilon(k/n) G_p(k/n)\right]',$$

since $X^{'}X/n = o(1) + M(a,b)$ (in the sense that each entry converges). The expression on the right is simply

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

Now, our assumptions validate the hypotheses of Theorem 4, and hence applying that result completes the proof. $\square$

Acknowledgements 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. The authors would like to thank Cheng Cheng for his useful comments leading to the improvement of an earlier draft.

References

[1] Cheng, C.(1995) Uniform consistency of generalized kernel estimators of quantile density. *The Annals of Statistics*, **23** 2285-2291.

[2] Cheng, C., and Parzen, E. (1997) Unified estimators of smooth quantile and quantile density functions. *Journal of Statistical Planning and Inference*, **59** 291–307.

[3] Csrgő, S., Deheuvels, P., and Mason, D. (1985) Kernel Estimates of the Tail Index of a Distribution. *The Annals of Statistics*, **13** 1050 – 1077.

[4] de Haan, L., and Resnick, S. (1980) A simple asymptotic estimate for the index of a stable distribution. *Journal of the Royal Statistical Society, Series B*, **42** 83 – 87.

[5] Dekkers, A., Einmahl, J., and de Haan, L. (1989) A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, **17** 1833 – 1855.

[6] Drees, H. (1995) Refined Pickands Estimators of the Extreme Value Index. *The Annals of Statistics*, **23** 2059 – 2080.

[7] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling Extremal Events.* New York: Springer.

[8] Falk, M., (1985) Asymptotic normality of the kernel quantile estimator. *The Annals of Statistics*, **13** 428–433.

[9] Falk, M., (1986) On the estimation of the quantile density function. *Statistics & Probability Letters*, **4** 69–73.

[10] Gihman, I., and Skorohod, A. (1980) *The Theory of Stochastic Processes.* Berlin: Springer.

[11] Gomes, M., and Martins, M. (2001) Generalizations of the Hill Estimator – Asymptotic Versus Finite Sample Behaviour. *Journal of Statistical Planning and Inference*, **93** 161–180.

[12] Hall, P., and Welsh, A.H. (1985) Limit theorems for the median deviation. *Annals of the Institute of Statistical Mathematics*, **37**, 27–36.

[13] Hart, J. (1997) *Nonparametric Smoothing and Lack-of-Fit Tests.* New York: Springer.

[14] Hill, B.M. (1975) A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, **3** 1163 – 1174.

[15] Karatzas, I., and Shreve, S. (1997) *Brownian Motion and Stochastic Calculus.* New York: Springer.

[16] McCulloch, J. (1997) Measuring Tail Thickness to Estimate the Stable Index α: A Critique. *Journal of Business and Economic Statistics*, **15** 74 – 81.

[17] McElroy, T., and Politis, D. (In Press) Moment Based Tail Index Estimation. *Journal of Statistical Planning and Inference.*

[18] Parzen, E. (1979) Nonparametric Statistical Data Modeling *Journal of the American Statistical Association*, **74** 105 – 121.

[19] Parzen, E. (1982) Data Modeling Using Quantile and Density-Quantile Functions *Some Recent Advances in Statistics, ed. J. Tiago de Oliveira and B. Epstein,* Academic Press: New York. 23-52.

[20] Parzen, E. (2004) Quantile probability and statistical data modeling *Statistical Science*, **19** 652 – 662.

[21] Pickands, J. III (1975) Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, **3** 119 – 131.

[22] Rojo, J. (1996) On tail categorization of probability laws. *Journal of the American Statistical Association*, **91** 378–384.

[23] Schuster, E. (1984) Classification of probability laws by tail behavior. *Journal of the American Statistical Association*, **79** 936–939.

[24] Teugels, J. (1981) Limit theorems on order statistics. *The Annals of Statistics*, **9** 868 – 880.

[25] Xiang, X. (1994) A law of the logarithm for kernel quantile density estimators. *The Annals of Probability*, **22** 1078–1091

		$n = 1000$			$n = 200$		
α^p	α^c	Mean	Stdev	MSE	Mean	Stdev	MSE
3	.5	3.008	.339	.115	3.279	.954	.988
2.333	.75	2.274	.279	.081	2.445	.681	.477
2	1	1.921	.236	.062	2.056	.587	.348
1.667	1.5	1.546	.208	.058	1.65	.507	.257

Table 1: Left tail index estimation (of α^p) using qdf estimator (13). The simulations consisted of 1000 repetitions of sample sizes 1000 and 200.

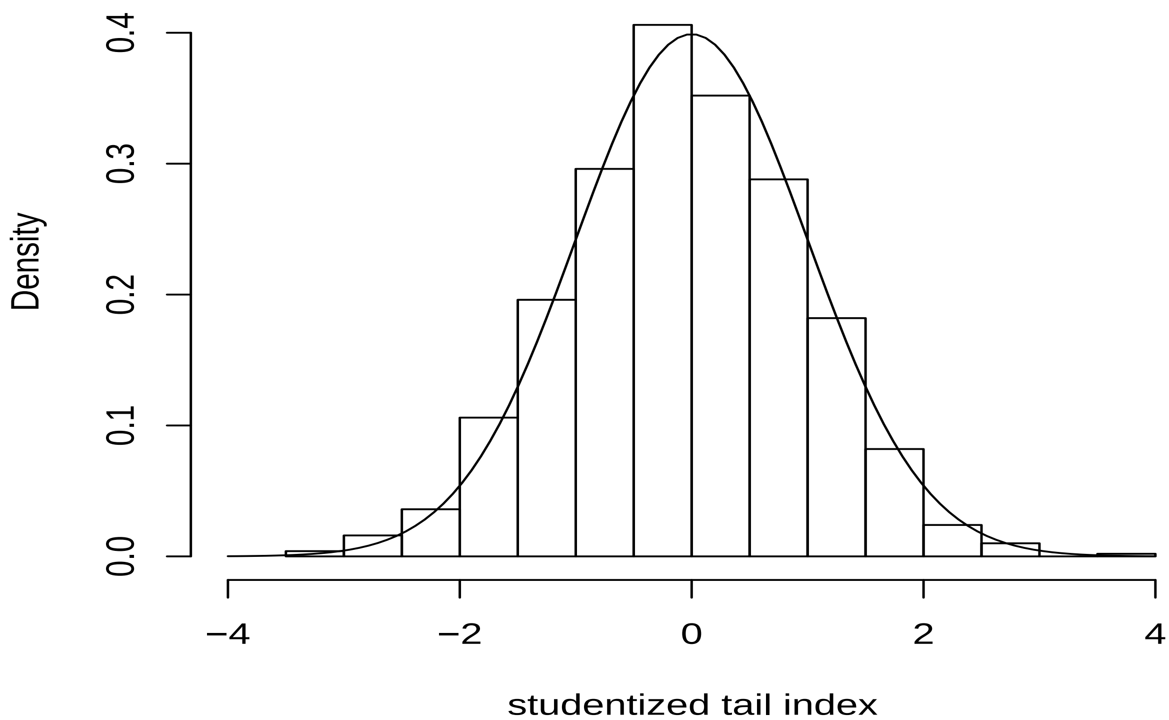

Figure 1: This figure contains a histogram for the studentized distribution of the left tail index estimator for $\alpha^p = 2$ from a simulation with 1000 repetitions and of sample size 1000.