Statistical Methods and Applications manuscript No.
(will be inserted by the editor)

Modeling Heavy-Tailed, Skewed and Peaked Uncertainty Phenomena with Bounded Support

C. B. García · J. García Pérez · J. R. van Dorp

Received: 12 November 2010 / Accepted: 28 July 2011

Abstract A prevalence of heavy-tailed, peaked and skewed uncertainty phenomena have been cited in literature, dealing with economic, physics, and engineering data. This fact has invigorated the search for continuous distributions of this nature. In this paper we shall generalize the two-sided framework presented in Kotz and van Dorp (2004) for the construction of families of distributions with bounded support via a mixture technique utilizing two generating densities instead of one. The family of Elevated Two-Sided Power (ETSP) distributions is studied as an instance of this generalized framework. Through a moment ratio diagram comparison, we demonstrate that the ETSP family allows for a remarkable flexibility when modeling heavy-tailed and peaked, but skewed, uncertainty phenomena. We shall demonstrate its applicability via an illustrative example utilizing 2008 US income data.

Keywords Uncertainty modeling · Applied Probability · Income Distribution · Lorentz Curve

C. B. García
Departamento de Métodos Cuantitativos para la Economía y la Empresa, Facultad de Ciencias Económicas y Empresariales, University of Granada, Campus de Cartuja s/n, Granada, 18071, Spain
E-mail: cbgarcia@ugr.es

J. García Pérez
Departamento de Economía Aplicada. University of Almería. Ctra. La Cañada de San Urbano, s/n, Almería, 04120, Spain
E-mail: jgarcia@ual.es

J. R. van Dorp
Department of Engineering Management and Systems Engineering, The George Washington University, 1776 G Street, N.W., Washington D.C. 20052, USA; Corresponding Author
Tel.: 202-994-6638
Fax: 202-994-0245
E-mail: dorpjr@gwu.edu

1 Introduction

Further confirmation that financial and insurance data reveal a heavy-tailed, peaked and skewed behavior have appeared in the relevant literature (e.g., Levy and Duchin 2004; McFall Lamm 2003; Kotz et al. 2001; Solomon and Levy 2000; Fernandez and Steel 1998; Embrechts et al. 1997; McCulloch 1996). Other areas where a prevalence of heavy tails have been cited are physics, hydrology, meteorology and engineering (see, e.g., Gomez et al. 2007; Douglas and Barros 2003; Lu and Molz 2001; Barkai et al. 2000; Adler et al. 1998; Fernandez and Steel 1998; Resnick 1997; Samorodnitsky and Taqqu 1994). Naturally, this has sparked the search for distribution models that reflect these properties. Early developments in this domain can be traced back to Lévy (1925), who introduced Lévy stable distributions with support $(-\infty, \infty)$ that rely on a power law behavior to model tail obesity. Other proposed distributions that share the same power law tails and real line support are the Student t (e.g., Zabell 2008), Pearson-type IV (e.g., Nagahara 1999) and the Doubly-Pareto Uniform (DPU) distributions (Singh et al. 2007), amongst others. Contrary to the Student t, Pearson-type IV and Lévy stable distributions, a DPU cumulative distribution function (cdf) can be expressed in closed form using only elementary functions. DPU distributions generalized the Pareto distributions with support $[a, \infty]$, $a \in \mathbb{R}^+$ originally introduced by Vilfredo Pareto (1897) to study wealth and income distributions. Recently, Aban et al. (2006) studied truncated Pareto distributions with bounded support $[a, b]$ and probability density function (pdf)

$$f(x|a,b,c) = \frac{ca^c\,(x)^{-c-1}}{1-\left(\frac{a}{b}\right)^c}, 0 < a < x \leq b,\ c > 1, \tag{1}$$

as an example of a bounded distribution with heavy-tails. Aban et al. (2006) too cited applications in finance, groundwater hydrology and atmospheric science in their motivation for the heavy-tailed but bounded truncated Pareto distribution.

Even today, the multitude of existing unbounded continuous distributions developed during the 20th century contrasts sharply with the relative scarcity of bounded distributions. This especially applies to bounded heavy-tailed distributions. While distributions with unbounded support rely on the power law to model tail obesity, this is not a requirement when dealing with a bounded support. Hahn (2008) constructed a heavy-tailed distribution with bounded support as a mixture of a uniform and the classical beta distribution, denoting them beta-rectangular (BR) distributions. A BR pdf with support $[0, 1]$ is given by:

$$\begin{aligned} &f(x|\alpha,\beta,\delta) = \delta + (1-\delta)b(x|\alpha,\beta), \delta \in [0,1] \text{ where} \\ &b(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ \alpha,\beta > 0. \end{aligned} \tag{2}$$

Hahn (2008) mentioned the pervasiveness of heavy-tailed phenomena in business contexts as his motivating argument for (2). Similar to truncated Pareto

distributions (1), BR densities (2) allow for strictly positive densities at their lower- and upper bounds, albeit requiring them to be of the same value. Contrary to truncated Pareto distributions, BR distributions offer the flexibility of a modal value location over the range of their entire support, whereas truncated Pareto distributions only allow for a mode at their lower bound. However, if heavy-tailed bounded distributions with strictly positive density values at their bounds and a modal value somewhere in between are to be considered useful, one could argue that it could also be desirable to allow for strictly positive but different density values at their lower and upper bounds, respectively.

In this paper we shall consider Elevated Two-Sided Power distributions (ETSP) that allow for that flexibility. The functional form of an ETSP pdf with support $[0, 1]$ will be constructed in this paper and is given by:

$$g\{x|\Theta\} = \mathcal{C}^{-1}(\Theta)\times \tag{3}$$
$$\begin{cases} \{\delta + (1-\delta)n\}\{\lambda + (1-\lambda)m(\frac{x}{\theta})^{m-1}\}, & \text{for } 0 < x \leq \theta \\ \{\lambda + (1-\lambda)m\}\{\delta + (1-\delta)n(\frac{1-x}{1-\theta})^{n-1}\}, & \text{for } \theta < x < 1, \end{cases}$$

where $\Theta = \{\theta, m, n, \lambda, \delta\}$, the threshold parameter $0 \leq \theta \leq 1$, elevation parameters $0 \leq \lambda, \delta \leq 1$ and the power parameters $m, n > 0$. The normalization constant $\mathcal{C}(\Theta)$ is given by:

$$\mathcal{C}(\Theta) = (1-\theta)\{\lambda + (1-\lambda)m\} + \theta\{\delta + (1-\delta)n\}. \tag{4}$$

Figure 1 plots example ETSP and BR distributions fitted utilizing the least squares method to standardized 2008 US Census Bureau annual income data for black females in the income range $[\$2,500; \$250,000)$ covering 78.34% of that population. Of black females, 0.10% have an income larger than $\$250,000$ (with no upper bound specified by the US Census Bureau) and 21.56% fall in the category of $[0; \$2,500)$, which also includes those with no income at all. Hence the income bracket $[\$2,500; \$250,000)$ is thought to be representative of the predominant black female sub-population that have annual income arising from remuneration for employment, and was therefore selected for the analysis in Figure 1. Observe from Figure 1A a larger positive ETSP density value at the lower than at its upper bound. The BR density in Figure 1B also has positive, but equal, density values at its boundaries. Figures 1C and 1D present QQ-plots for the fitted ETSP and BR cdf's, respectively. While both demonstrate a good fit to the income data, one visually observes from Figures 1C and D a better fit for the ETSP cdf at the lower income ranges than the BR cdf. The ETSP distribution in Figures 1A and C also fits "better" than the BR one in Figures 1B and 1D in the least squares sense, and in terms of the log-likelihood and Kolmogorov-Smirnov statistics (see Section 5).

The additional flexibility depicted in Figure 1A naturally comes at a price when compared to the density in Figure 1B. Whereas the distribution in Figure 1B belongs to a three parameter family (two parameters of the beta distribution and one mixture parameter, see (2)), the distribution in Figure 1A has

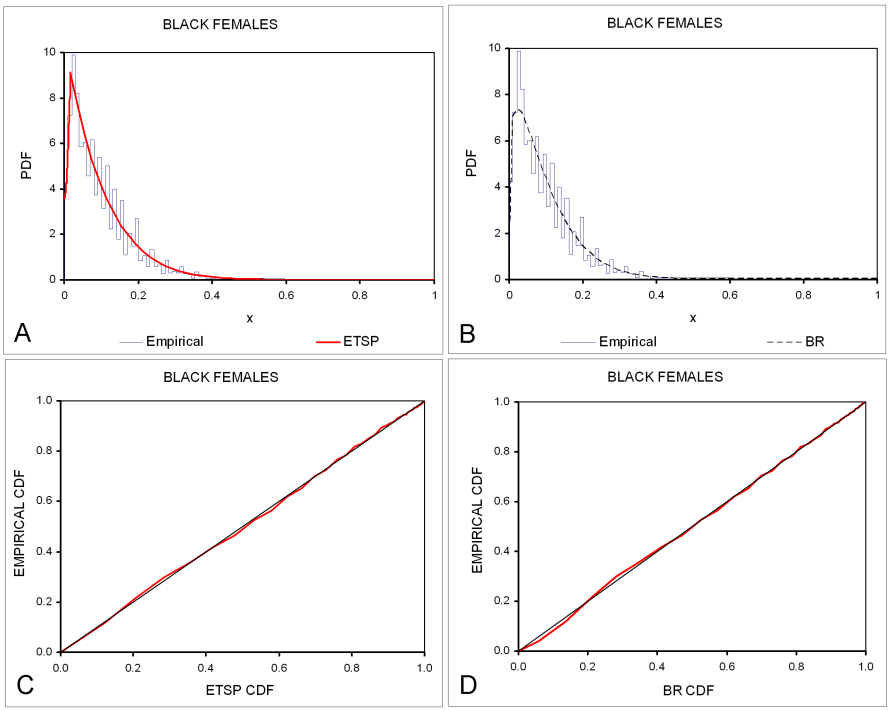

Fig. 1 Comparison of ETSP (3) and BR (2) fitted densities to standardized US 2008 annual income data for black females within the range [\$2, 500; \$250, 000]; A: ETSP density (3) with $\lambda = 0.659$, $\delta = 0.009$, $\theta = 0.0169$, $m = 3.007$, $n = 9.949$; B: BR density (2) with $\delta = 0.020$, $\alpha = 1.229$, $\beta = 11.731$; C: QQ-plot ETSP cdf; D: QQ-plot BR cdf.

five parameters: a threshold parameter, two power parameters, one for each tail, and two elevation parameters for its two branches, see (3). This does not have to be a drawback, however, when an abundance of data is available not uncommon when dealing with, for example, financial data. A similar argument applies to the US 2008 income distribution data used in Figure 1 which is constructed from thousands of households. In that case the task of a distribution modeler becomes to device one that "best" describes the data and essentially parsimony of the fitted distribution is not the issue it is when dealing with a lack of data.

To construct the ETSP pdf (3) we shall first, in Section 2, present a generalized two-sided framework of distributions where two separate generating densities describe each branch while maintaining continuity at the threshold θ. We shall derive the general expressions for the cdf and moments for this framework. In Section 3, we shall present an instance of this generalized two-sided framework, by providing the two generating densities that yield the pdf (3). We shall also further develop the properties for the ETSP family of distributions utilizing its generating densities and the generalized two-sided framework in Section 2. Skewness and kurtosis behavior of ETSP distributions is studied

in Section 4 using various moment ratio diagram comparisons involving those of Asymmetric Laplace (AL), Two-Sided Power (TSP) and beta distributions (see, e.g., Kotz and van Dorp, 2004). While it should be no surprise that ETSP distributions indeed allow for higher kurtosis values (reflecting "heavier tails") than TSP, AL and beta distributions, it is perhaps surprising that they too allow for a larger skewness coverage in their moment ratio diagrams when considering only unimodal ETSP distributions. In Section 5 we shall present a more detailed analysis of 2008 US income data using fitted ETSP and BR distribution by gender and race demonstrating an overall preference of ETSP distributions over BR fitted ones in more cases than not.

2 Generalized two-sided power distributions

Van Dorp and Kotz (2003) deal with the following family of two-sided probability density functions (pdf's)

$$g\{x|\theta, p(\cdot|\Psi)\} = \begin{cases} p(\frac{x}{\theta}|\Psi), & \text{for } 0 < x \leq \theta \\ p(\frac{1-x}{1-\theta}|\Psi), & \text{for } \theta < x < 1, \end{cases} \tag{5}$$

where $p(\cdot|\Psi)$ is a generating density function with bounded support $(0, 1)$. Substitution in (5) of, e.g., the power density $p(\cdot|n) = nx^{n-1}$ leads to the Two-Sided Power (TSP) family of distributions studied in detail in their first paper on this topic, van Dorp and Kotz (2002). Other related contributions by these authors appearing prior to this paper have been summarized in the monograph by Kotz and van Dorp (2004).

Consider now the following generalization of (5) with pdf's defined by:

$$g\{x|\theta, p(\cdot|\Psi), q(\cdot|\Upsilon)\} = \begin{cases} \frac{\pi}{\theta} p(\frac{x}{\theta}|\Psi), & \text{for } 0 < x \leq \theta \\ \frac{1-\pi}{1-\theta} q(\frac{1-x}{1-\theta}|\Upsilon), & \text{for } \theta < x < 1, \end{cases} \tag{6}$$

where $0 \leq \pi \leq 1$ and $p(\cdot|\Psi)$ and $q(\cdot|\Upsilon)$ are continuous generating densities with support $(0, 1]$. While the pdf (5) is continuous at the threshold θ, the following condition

$$\frac{\pi}{\theta} p(1|\Psi) = \frac{1-\pi}{1-\theta} q(1|\Upsilon) \tag{7}$$

must hold for the pdf (6) to be continuous. From (7) one obtains

$$\pi = \frac{\theta q(1|\Upsilon)}{(1-\theta)p(1|\Psi) + \theta q(1|\Upsilon)} \Rightarrow \begin{cases} \theta = 0 \Rightarrow & \pi = 0, \\ \theta = 1 \Rightarrow & \pi = 1. \end{cases} \tag{8}$$

Substitution of (8) into (6) yields the framework of continuous Generalized Two-Sided (GTS) distributions with support $(0, 1)$ and density functions

$$g\{x|\theta, p(\cdot|\Psi), q(\cdot|\Upsilon)\} = \begin{cases} \frac{q(1|\Upsilon)}{(1-\theta)p(1|\Psi)+\theta q(1|\Upsilon)} p(\frac{x}{\theta}|\Psi), & \text{for } 0 < x \leq \theta, \\ \frac{p(1|\Psi)}{(1-\theta)p(1|\Psi)+\theta q(1|\Upsilon)} q(\frac{1-x}{1-\theta}|\Upsilon), & \text{for } \theta < x < 1. \end{cases} \tag{9}$$

The advantage of (9) over (5) is that it allows for separate branch specification via two generating densities $p(\cdot|\Psi)$ and $q(\cdot|\Upsilon)$ thereby (substantially) enhancing framework's (5) flexibility. Observe from (9) that the members of the GTS family too hinge at θ, which can be interpreted as the pivotal point of the distribution. Note the conceptual difference between the threshold parameter θ which determines the "turning point" of the distribution under consideration and the parameters included in Ψ and Υ, which control the form of the two sides of the distribution (to the left and the right of θ).

Continuing our investigation of the general structure of GTS families of distributions it is at times convenient to rewrite (9) in its mixture form:

$$g\{x|\theta, p(\cdot|\Psi), q(\cdot|\Upsilon)\} = \pi \frac{p(\frac{x}{\theta}|\Psi)}{\theta} + (1-\pi)\frac{q(\frac{1-x}{1-\theta}|\Upsilon)}{1-\theta}, \tag{10}$$

where the mixture weight π is given by (8). Observe from (8) that π solely follows from the threshold θ and the generating density values at 1, $q(1|\Upsilon)$ and $p(1|\Psi)$, and not the shape of the generating densities $p(\cdot|\Psi)$ and $q(\cdot|\Upsilon)$ up to 1 (which is perhaps counterintuitive). If $X \sim g\{x|\theta, p(\cdot|\Psi), q(\cdot|\Upsilon)\}$ we have from (10) and the $(0,1]$ support for both $p(\cdot|\Psi)$ and $q(\cdot|\Upsilon)$ that

$$Pr(X \leq \theta) = \pi. \tag{11}$$

When $q(1|\Upsilon) = p(1|\Psi)$ we have from (8) and (11) that $Pr(X \leq \theta) = \theta$. In the special case that $q(\cdot|\Upsilon) \equiv p(\cdot|\Psi)$, $Pr(X \leq \theta) = \theta$ holds regardless of the generating density's shape and the framework (9) reduces to the framework of two-sided distributions (5).

From (9) we immediately obtain the cdf for X

$$G\{x|\theta, p(\cdot|\Psi), q(\cdot|\Upsilon)\} = \begin{cases} \pi P(\frac{x}{\theta}|\Psi), & \text{for } 0 < x < \theta \\ 1 - (1-\pi)Q(\frac{1-x}{1-\theta}|\Upsilon), & \text{for } \theta \leq x < 1, \end{cases} \tag{12}$$

where $P(\cdot|\Psi)$ and $Q(\cdot|\Upsilon)$ are the cdf's of the generating densities $p(\cdot|\Psi)$ and $q(\cdot|\Upsilon)$, respectively. From (12) we derive for the quantile function of X

$$G^{-1}\{y|\theta, p(\cdot|\Psi), q(\cdot|\Upsilon)\} = \begin{cases} \theta P^{-1}(\frac{y}{\pi}|\Psi), & \text{for } 0 < y < \pi \\ 1 - (1-\theta)Q^{-1}(\frac{1-y}{1-\pi}|\Upsilon), & \text{for } \pi \leq y < 1, \end{cases} \tag{13}$$

where $P^{-1}(\cdot|)\Psi)$ and $Q^{-1}(\cdot|\Upsilon)$ are the quantile functions of the generating cdf's $P(\cdot|\Psi)$ and $Q(\cdot|\Upsilon)$, respectively.

If $X \sim g\{\cdot|\theta, p(\cdot|\Psi), q(\cdot|\Upsilon)\}, Y \sim p(\cdot|\Psi), Z \sim q(\cdot|\Upsilon)$ the following relationship between the moments around zero of X, Y and Z can straightforwardly be derived:

$$E[X^k|\theta, \Psi, \Upsilon] = \pi\theta^k E[Y^k|\Psi] + (1-\pi)\sum_{i=0}^{k}\binom{k}{i}(-1)^i(1-\theta)^{i+1}E[Z^i|\Upsilon]. \tag{14}$$

where π is given by (8). From (14) and (8), utilizing modern computational facilities (especially for large values of k), moments of generalized two-sided

distributions can be calculated at least in the case when closed form expressions for the central moments of Y and Z are available. In particular, we have for the first two moments

$$E[X|\theta,\Psi,\Upsilon] = \pi\theta E[Y|\Psi] + (1-\pi)\{1-(1-\theta)E[Z|\Upsilon]\}, \tag{15}$$

$$E[X^2|\theta,\Psi,\Upsilon] = \pi\theta^2 E[Y^2|\Psi] + (1-\pi)\times$$
$$\{1-2(1-\theta)E[Z|\Upsilon] + (1-\theta)^2 E[Z^2|\Upsilon]\}.$$

Note that utilizing (8) we can consider the following special cases of (15) :

$E[X\|\theta,\Psi,\Upsilon] = \theta^2 E[Y\|\Psi] - (1-\theta)^2 E[Z\|\Upsilon] + (1-\theta)$,	$q(1\|\Upsilon) = p(1\|\Psi)$,
$E[X\|1,\Psi] = E[Y\|\Psi]$,	$\theta = 1$,
$E[X\|0,\Upsilon] = 1 - E[Z\|\Upsilon]$,	$\theta = 0$,
$E[X\|\theta,\Psi] = (2\theta-1)E[Y\|\Psi] + (1-\theta)$,	$q(\cdot\|\Upsilon) \equiv p(\cdot\|\Psi)$,
$E[X\|\frac{1}{2},\Psi] = \frac{1}{2}$,	for $q(\cdot\|\Upsilon) \equiv p(\cdot\|\Psi)$, $\theta = \frac{1}{2}$.

3 Construction and properties of ETSP distributions

Consider the following two generating densities and their cdf's:

$$\begin{cases} p(x|\lambda,m) = \lambda + (1-\lambda)mx^{m-1}, & P(x|\lambda,m) = \lambda x + (1-\lambda)x^m, \\ q(x|\delta,n) = \delta + (1-\delta)nx^{n-1}, & Q(x|\delta,n) = \delta x + (1-\delta)x^n, \end{cases} \tag{16}$$

where $x, \lambda, \delta \in [0,1]$ and $m, n > 0$. Observe from (16) that these generating distributions themselves are mixtures of a uniform and power distributions both with support $[0,1]$ and thus are members within Hahn's (2008) class of distributions. Substitution of $p(1|\lambda,m)$, $q(1|\delta,n)$ into (8) yields

$$\pi(\Theta) = \frac{\theta\{\delta + (1-\delta)n\}}{(1-\theta)\{\lambda + (1-\lambda)m\} + \theta\{\delta + (1-\delta)n\}}, \tag{17}$$

where $\Theta = \{\theta, m, n, \lambda, \delta\}$. Substitution of the pdf's (16) and $\pi(\Theta)$ (17) into (9) yields the pdf (3). From (3) and (4) we have

$$\frac{g(0|\Theta)}{g(1|\Theta)} = \frac{\{\delta + (1-\delta)n\}\lambda}{\{\lambda + (1-\lambda)m\}\delta},$$

which does not depend on the threshold parameter θ, is not necessarily equal to 1, and thus the pdf (3) allows for different strictly positive values of its density at the lower and upper bounds. Observe from (17) that $\pi \to 0$ as $m \to \infty$ keeping n fixed and $0 < \theta, \lambda < 1$. Hence, in that case the pdf (3) converges to the density

$$g(x|\theta,\delta,n) = \frac{1}{1-\theta}\left\{\delta + (1-\delta)n\Big(\frac{1-x}{1-\theta}\Big)^{n-1}\right\}$$

with support $[\theta, 1]$. Vice versa, $\pi \to 1$ as $n \to \infty$ keeping m fixed and $0 < \theta, \delta < 1$. Hence, in that case the pdf (3) converges to the density

$$g(x|\theta, \lambda, m) = \frac{1}{\theta}\left\{\lambda + (1-\lambda)m\Big(\frac{x}{\theta}\Big)^{m-1}\right\}$$

with support $[0, \theta]$. Substitution of the cdf's $P(x|\lambda, m)$ and $Q(x|\delta, n)$ (see (16)) into (12) yields the cdf of (3):

$$G(x|\Theta) = \tag{18}$$
$$\begin{cases} x\Big\{\frac{\pi(\Theta)}{\theta}\Big\}\{\lambda + (1-\lambda)(\frac{x}{\theta})^{m-1}\}, & \text{for } 0 < x < \theta \\ 1 - (1-x)\Big\{\frac{1-\pi(\Theta)}{1-\theta}\Big\}\{\delta + (1-\delta)(\frac{1-x}{1-\theta})^{n-1}\}, & \text{for } \theta \le x < 1, \end{cases}$$

where $\pi(\Theta)$ is given by (17). Figure 2 plots an ETSP pdf with $g(0|\Theta) = 1$ and $g(1|\Theta) = 0.25$. From its parameters settings $\theta = 0.25$, $m = 3.059$, $n = 4$, $\lambda = 0.661$, $\delta = 0.301$, we obtain from (17) that $\pi(\Theta) = 0.378$. Hence, from (11) we have that the median of this pdf is located within its second branch.

Various subclasses may be of interest within the ETSP class. These can be constructed with additional parameter restrictions. We have

$$\begin{aligned} m = n,\ \lambda = \delta = 0 &: ETSP\,(5\,\text{parameters}) \to TSP\,(\,2\,\text{parameters}) \\ \lambda = \delta = 0 &: ETSP\,(5\,\text{parameters}) \to GTSP\,(\,3\,\text{parameters}) \\ 0 \le \lambda = \delta \le 1 &: ETSP\,(5\,\text{parameters}) \to UTSP\,(\,4\,\text{parameters}) \end{aligned}$$

Hence, by forcing the same power parameters in both branches and no elevation in its tails, the ETSP distribution reduces to the TSP distribution (see van Dorp and Kotz, 2002). Allowing for separate power parameters, but no tail elevation, the ETSP distribution simplifies to the GTSP distribution only mentioned in passing in Kotz and van Dorp (2004) and investigated recently in more detail in Herrerías-Velasco et al. (2008). If one next allows for elevation in both tails but requires this elevation to be the same in both branches, we obtain Uniform TSP (UTSP) distributions that are reminiscent of BR distributions (2) introduced by Hahn (2008).

Unfortunately, the quantile functions $P^{-1}(y|\lambda, m)$ and $Q^{-1}(y|\delta, n)$ are not available in a closed form, and we devised the numerical algorithm (19) below to evaluate them. In Steps 1 and 2 we take advantage of the mixture structure of $P(y|\lambda, m)$ when constructing the starting interval for the search algorithm. Note, that E_1 and E_2 in Step 1 are the quantile functions of the mixture components of $P(x|\lambda, m)$ and by design the interval $[LB, UB]$ constructed in Step 2 contains the solution to the equation $y = P(x|\lambda, m)$. Please observe that Steps 3 through 6 in (19) follow the structure of a standard bisection algorithm. Obvious modifications to the algorithm (19) can be made to solve

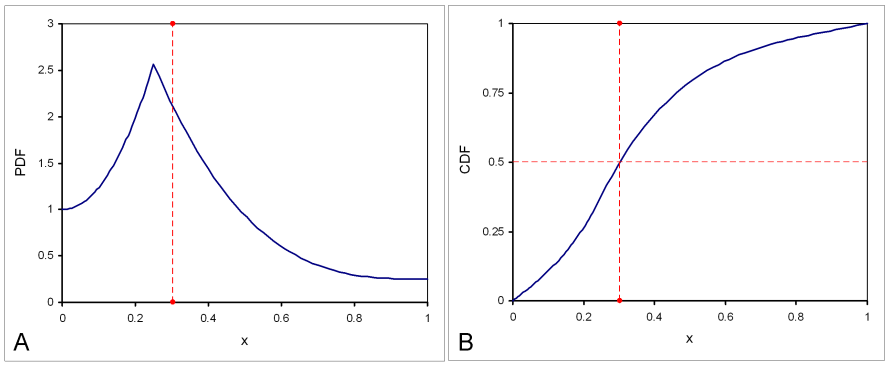

Fig. 2 An example ETSP distribution with $\theta = 0.25$, $m = 3.059$, $n = 4$, $\lambda = 0.661$, $\delta = 0.301$ and a median at 0.302; A: ETSP pdf (3); B: ETSP cdf (18).

for the equation $y = Q(x|\delta, n)$.

$$\begin{aligned}
&\text{STEP 1: } E_1 = y,\ E_2 = y^{\frac{1}{m}}. \\
&\text{STEP 2: If } E_1 < E_2 \text{ then } LB = E_1,\ UB = E_2, \\
&\qquad \text{Else } LB = E_2,\ UB = E_1. \\
&\text{STEP 3: } x = (LB + UB)/2. \\
&\text{STEP 4: If } |P(x|\lambda, m) - y| < \epsilon \text{ then Stop.} \\
&\text{STEP 5: If } P(x|\lambda, m) < y \text{ then UB} = x,\ x = (LB + UB)/2, \\
&\qquad \text{Go to STEP 4.} \\
&\text{STEP 6: If } P(x|\lambda, m) > y \text{ then LB} = x, x = (LB + UB)/2, \\
&\qquad \text{Go to STEP 4.}
\end{aligned} \tag{19}$$

Combination of both algorithms for evaluation of $P^{-1}(\frac{y}{\pi}|\lambda, m)$ and $Q^{-1}(\frac{1-y}{1-\pi}|\delta, n)$ with expression (13) yields the following master algorithm for the evaluation of $G^{-1}(y|m, n, \theta, \lambda, \delta)$:

$$\begin{aligned}
&\text{If } (y < \pi) \text{ then } y = y/\pi,\ x = P^{-1}(y|\lambda, m),\ x = \theta x \\
&\text{Else } y = (1 - y)/(1 - \pi),\ x = Q^{-1}(y|\delta, n),\ x = 1 - (1 - \theta)x,
\end{aligned} \tag{20}$$

where π in (20) is given by (17). The algorithm (20) was used to solve for the median value 0.302 of the ETSP distribution displayed in Figure 2. The algorithm (20) used 6 iterations, with a setting of $\epsilon = 0.001$ in the algorithm (19). The algorithm (20) can be considered somewhat efficient since the generating cdf's $P(x|\lambda, m)$, $Q(x|\delta, n)$ used in their versions of the sub-algorithm (20) are available in a closed form. In contrast, the cdf's of beta distributions (and thus the cdf's of BR distributions (2)) are not available in a closed form, and a separate numerical routine is required to evaluate the incomplete beta function (i.e. the beta cdf).

If $X \sim g\{\cdot|\theta, p(\cdot|\Psi), q(\cdot|\Upsilon)\}$ given by (5), $\Psi = \{\lambda, m\}$, $\Upsilon = \{\delta, n\}$ and $Y \sim p(\cdot|\Psi), Z \sim q(\cdot|\Upsilon)$ are given by (16), we derive the following expressions for the moments of Y and Z

$$\begin{aligned} E[Y^k|\lambda, m] &= \tfrac{m(k+1)+\lambda(1-m)k}{(k+1)(m+k)}, \\ E[Z^k|\delta, n] &= \tfrac{n(k+1)+\delta(1-n)k}{(k+1)(n+k)}, \end{aligned} \tag{21}$$

and substitution of (21) into (14) yields the following moments expression for X

$$E[X^k|\Theta] = \pi(\Theta) \times \theta^k \Big\{ \frac{m(k+1)+\lambda(1-m)k}{(k+1)(m+k)} \Big\} + \tag{22}$$

$$\{1-\pi(\Theta)\} \times \sum_{i=0}^{k} \binom{k}{i} (-1)^i (1-\theta)^i \Big\{ \frac{n(i+1)+\delta(1-n)i}{(i+1)(n+i)} \Big\},$$

where $\Theta = \{\theta, m, n, \lambda, \delta\}$ and $\pi(\Theta)$ is given by (17). For example, substitution of $k = 1$ in (22) yields the following expression for the mean $E[X]$:

$$E[X|\Theta] = \pi(\Theta) \times \theta \Big\{ \frac{2m+\lambda(1-m)}{2(m+1)} \Big\} + \tag{23}$$

$$\{1-\pi(\Theta)\} \times \Big[1 - (1-\theta) \Big\{ \frac{2n+\delta(1-n)}{2(n+1)} \Big\} \Big].$$

Explicit forms for the variance $\text{var}[X] = E[X^2] - E^2[X]$, skewness $\sqrt{\beta_1}$ and kurtosis β_2 for ETSP distributions result in cumbersome expressions and are omitted. However, values for the variance, skewness and kurtosis can be calculated directly using the general expression for the moments around the origin $\mu'_k = E[Y^k]$, $k = 1, \ldots, 4$, (23) and their relationship with the central moments $\mu_k = E[(Y - E[Y])^k]$, $k = 2, 3, 4$, (e.g., Stuart and Ord 1994). Upon request, a Microsoft Excel spreadsheet is available from the authors which evaluates the measures $\text{var}[X]$, $\sqrt{\beta_1}$ and β_2 for ETSP distributions. In addition, this spreadsheet facility allows one to graph the pdf and cdf of an ETSP distribution, and contains a macro implementation of the algorithm (20) to evaluate its quantiles at arbitrary quantile levels.

4 A moment ratio diagram comparison

Moment ratio diagrams, popularized for Pearson-type distributions by Elderton and Johnson (1969), provide a convenient means to study skewness and kurtosis behavior of a particular family of distributions. A moment ratio diagram is a plot with the skewness $\sqrt{\beta_1}$ on the horizontal axis and the kurtosis β_2 on the vertical axis, with the convention that $\sqrt{\beta_1}$ retains the sign of μ_3 (e.g., Kotz and Johnson, 1985). Values of $\sqrt{\beta_1}$ and β_2 for ETSP distributions can be straightforwardly evaluated as outlined in the previous section. A comparison of the moment ratio diagram's coverage of two separate families may provide a direct visual assessment of a contrast between each family's flexibility

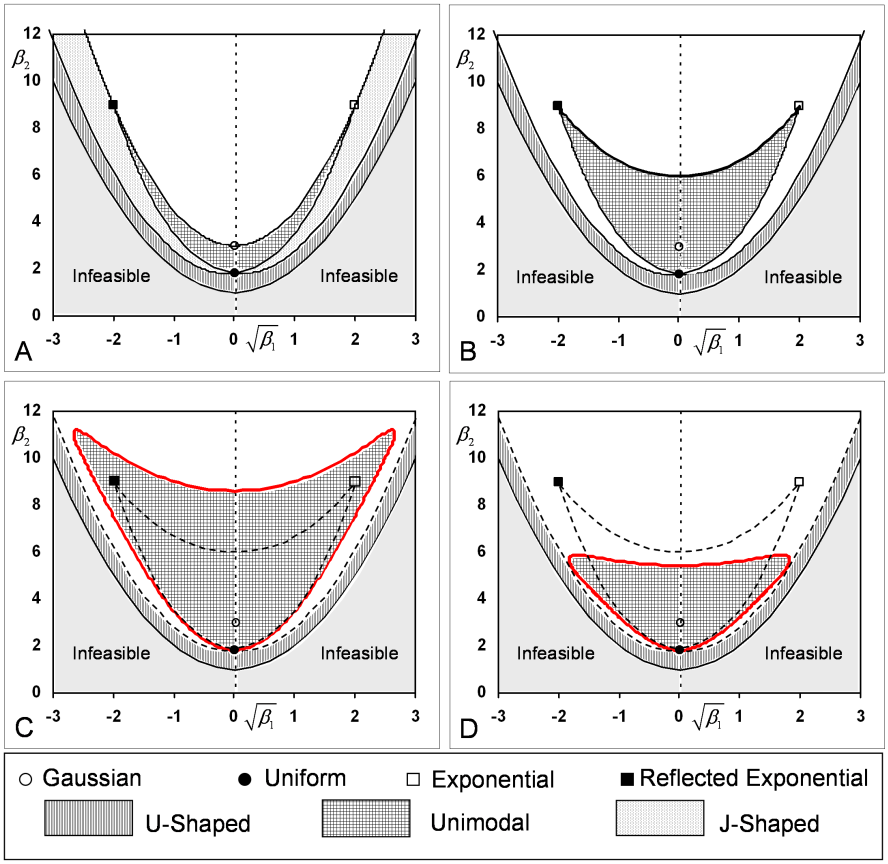

Fig. 3 Moment ratio diagrams; A: Beta distributions; B: TSP distributions; C: U-Shaped and unimodal ETSP distributions $\lambda = \delta = 0.1$, $1 \leq m \leq 9$, $m = n$; D: U-Shaped and unimodal ETSP distributions $\lambda = \delta = 0.25$, $1 \leq m \leq 9$, $m = n$.

to model peaked, heavy-tailed and skewed phenomena. Consider for example the moment ratio diagrams in Figure 3A and 3B for beta distributions and TSP distributions respectively. A synopsis of the moment ratio comparison between beta and TSP distribution is helpful for the appreciation of Figures 3C-D.

The central (checkered) part of both diagrams in Figures 3A and 3B represent the coverage of the unimodal forms of both beta and TSP distributions and we immediately observe a larger coverage for the TSP distributions in this domain. The coverage of the TSP distributions towards the feasibility boundary (vertically hatched) represents the coverage of the U-Shaped TSP distribution and we observe the same coverage for the U-shaped beta distributions. The feasibility boundary arises from Bernoulli distributions which are a limiting distribution for both the TSP and beta family of distributions. The coverage area between the central unimodal coverage and U-shaped cov-

erage (dotted) for beta distributions in Figure 3A arises from the J-shaped beta distributions. J-shapes cannot be modeled by TSP distributions (due to a requirement of equal power parameters in both branches) and hence this J-shaped coverage is lacking in Figure 3B. Figures 3A and 3B also indicate the (skewness, kurtosis) locations for the Gaussian $(0, 3)$, Uniform $(0, 1.8)$, Exponential $(2, 9)$, Reflected Exponential $(-2, 9)$. The dark solid line in Figure 3B represents the skewness-kurtosis combination of the Asymmetric Laplace (AL) distribution. Van Dorp and Kotz (2002) studied a version of the diagrams in Figures 3A and 3B with restricted parameter ranges which did not include the AL boundary. Kotz and van Dorp (2005) showed that the AL distribution is a limiting distribution of the TSP family of distributions. Hence, both Figure 3A and 3B provide a complete coverage for beta and TSP distributions.

Figure 3C shows the coverage (checkered) of unimodal ETSP distributions with the parameter restrictions $\lambda = \delta = 0.1$, $1 \leq m \leq 9$, $m = n$. The dotted lines in Figure 3C indicate the contours of the TSP boundaries in Figure 3B. From these dotted boundaries, however, we can draw the perhaps surprising conclusion that the coverage of unimodal ETSP distributions transitions into the J-shaped domain of the beta distributions in Figure 3A. This indicates an additional skewness flexibility of unimodal ETSP distributions over beta and TSP distributions. Moreover, we observe a larger coverage of the unimodal domain, even with the limited parameter settings $\lambda = \delta = 0.1$, $1 \leq m \leq 9$, $m = n$, than the complete unimodal coverage of TSP distributions in Figure 3B (even beyond the limiting boundary represented by the AL distribution).

Figure 3D shows the coverage (checkered) of unimodal ETSP distributions with the parameter restrictions $\lambda = \delta = 0.25$, $1 \leq m \leq 9$, $m = n$. We continue to observe a "filling in" of the J-shaped beta domain by unimodal ETSP distributions. However, even though these distributions have a heavier tail ($\lambda = \delta = 0.25$ in Figure 3D as opposed to $\lambda = \delta = 0.1$ in Figure 3C), we now observe smaller values for the kurtosis range in Figure 3D than in Figure 3C. This demonstrates once more that kurtosis is both a measure of peakedness and heavy-tailedness of distributions and that those distributions in Figure 3D are considered less peaked (i.e. more uniform due to the higher λ and δ weight). Indeed, when $\lambda, \delta \to 1$ the coverage area for the unimodal domain of ETSP distributions with parameter restriction $1 \leq m \leq 9$, $m = n$ converges to the single point skewness kurtosis combination $(0, 1.8)$ of the uniform distribution displayed in Figure 3.

TSP distributions are members within the ETSP class and hence Figure 3C and 3D also includes the U-shaped coverage of Figure 3B. Finally, for clarity of presentation we have omitted from Figures 3C and 3D the coverage of J-shaped ETSP distributions. J-shaped ETSP distribution arise when $\lambda = \delta = 0$ and either m or n having a value strictly less than 1. With this additional flexibility J-shaped ETSP distributions cover the same area as that of the J-shaped beta distributions displayed in Figure 3A. Summarizing, ETSP distribution provide a complete coverage of skewness and kurtosis combinations of the moment ratio diagrams in Figure 3, and even higher kurtosis values than the maximum value 12 indicated in Figure 3 can be attained. Said differently, any

skewness-kurtosis combination in the feasible moment-ratio Figure 3 can be represented by a member within the ETSP family of distributions which is a testament to its flexibility.

5 Modeling US 2008 income data by gender and ethnicity

A leading article of the 459 issue of the Journal of the American Statistical Association (2002, Vol. 97, pp. 663-673) by Barsky et al. (2002) presents an illuminating and comprehensive analysis of the African-American and Caucasian (Non-Hispanic) wealth gap based on a longitudinal survey of approximately over 6000 households over the period 1968-1992. Barsky et al. (2002) conclude that the role of earnings differences is largest at the lower tails of the wealth distribution, and decreases dramatically at higher wealth levels. In fact, their results indicate that differences in household income account for all of the racial wealth difference in the first quartile of the wealth distribution. This latter observation emphasizes the relevance of modeling the lower parts of income distributions, whereas over the course of many years, starting with Viflredo Pareto (1897), a larger emphasis has been put on modeling the upper parts, in particular the upper tails (see also, e.g., Arnold 1983; Coelho et al. 2008; Miyazima and Yamamoto 2006; Clementi and Gallegatib 2005).

Herein we shall present an illustrative analysis of modeling income distributions by gender and race, less focused on upper tail behavior, using BR (2) and ETSP (3) fitted distributions to 2008 US income data provided in Table 1. The data in Table 1 is extracted from the U.S. Census Bureau 2009 current population survey. We have from Table 1 that for each class by gender or by gender and ethnicity, the percentage of people that have an income larger than $\$250,000$ ranges from 0.03% (Hispanic females) to 1.69% (Asian males). No upper bound is specified by the US Census Bureau for this latter class and thus we shall focus our analysis on the subpopulation that have annual income less than $\$250,000$. Moreover, as per the US Census Bureau, the income class [0,$\$2,500$) in Table 1 also includes those that have no income at all. Since no means is provided for separating the population in the bracket [0,$\$2,500$) into those that have no income from the ones that do, this class is also omitted from our income distribution analysis. Summarizing, we shall model the US income distributions for the range from $[\$2,500, \$250,000)$ focused on a US sub-population that have income and by inference have some level of remuneration for employment. No doubt this income range may be thought of as being representative for the mainstream of the US population that have income and covers the lower parts of their income distributions (recall the discussion in the first paragraph).

Figures 4 and 5 plot fitted ETSP (3) and BR (2) densities to standardized annual income data in Table 1 for the range $[\$2,500, \$250,000)$. Hence, the origins in Figures 4 and 5 coincide with $\$2,500$ annual income and the upper boundary $x = 1$ coincides with the value $\$250,000$. These distributions were fitted by means of a least squares method. Specifically the distribution

Table 1 US 2008 male and female annual income data (Numbers in thousands. Income category under $2,500 includes people with no income.) Source: U.S. Census Bureau, Current Population Survey, 2009 Annual Social and Economic Supplement, http://www.census.gov/hhes/www/cpstables/032009/perinc/new11_000.htm.

		All Races	White	Black	Asian	Hispanic
Bracket	Male ; Female	Number	Number	Number	Number	Number
	Total	116,720 ; 123,424	79,542 ; 83,832	13,103 ; 15,816	5,082 ; 5,704	17,217 ; 16,231
1	Under $2,500	15,759 ; 25,498	8,824 ; 14,643	2,892 ; 3,410	784 ; 1,611	2,897 ; 5,408
2	$2,500 to $4,999	2,715 ; 4,693	1,653 ; 3,112	392 ; 529	129 ; 213	469 ; 742
3	$5,000 to $7,499	3,457 ; 6,523	1,944 ; 4,301	592 ; 904	177 ; 248	677 ; 951
4	$7,500 to $9,999	3,988 ; 7,508	2,253 ; 4,840	785 ; 1,236	127 ; 267	748 ; 1,036
5	$10,000 to $12,499	5,045 ; 7,829	2,963 ; 5,306	742 ; 1,029	234 ; 315	1,049 ; 1,080
6	$12,500 to $14,999	3,551 ; 5,556	2,280 ; 3,958	450 ; 732	107 ; 188	663 ; 602
7	$15,000 to $17,499	4,841 ; 6,211	3,078 ; 4,291	485 ; 755	148 ; 192	1,043 ; 881
8	$17,500 to $19,999	3,310 ; 4,439	2,118 ; 3,098	375 ; 573	114 ; 134	654 ; 563
9	$20,000 to $22,499	5,273 ; 5,671	3,130 ; 3,746	608 ; 773	199 ; 239	1,264 ; 842
10	$22,500 to $24,999	3,097 ; 3,701	2,066 ; 2,621	336 ; 468	104 ; 124	542 ; 422
11	$25,000 to $27,499	4,399 ; 4,793	2,818 ; 3,363	568 ; 678	156 ; 165	800 ; 510
12	$27,500 to $29,999	2,821 ; 2,929	1,907 ; 2,129	294 ; 395	116 ; 84	467 ; 283
13	$30,000 to $32,499	5,183 ; 4,652	3,356 ; 3,253	553 ; 630	212 ; 202	1,003 ; 511
14	$32,500 to $34,999	2,144 ; 2,099	1,504 ; 1,547	197 ; 281	104 ; 59	304 ; 184
15	$35,000 to $37,499	4,236 ; 3,680	2,879 ; 2,673	457 ; 502	183 ; 131	677 ; 338
16	$37,500 to $39,999	2,075 ; 1,903	1,614 ; 1,451	142 ; 224	56 ; 50	224 ; 146
17	$40,000 to $42,499	4,240 ; 3,328	2,942 ; 2,440	497 ; 440	167 ; 150	571 ; 270
18	$42,500 to $44,999	1,529 ; 1,349	1,138 ; 1,040	146 ; 137	43 ; 56	191 ; 95
19	$45,000 to $47,499	2,832 ; 2,274	2,140 ; 1,704	255 ; 257	92 ; 121	314 ; 169
20	$47,500 to $49,999	1,585 ; 1,303	1,194 ; 948	151 ; 184	58 ; 47	154 ; 107
21	$50,000 to $52,499	3,849 ; 2,720	2,802 ; 1,996	374 ; 339	192 ; 139	434 ; 226
22	$52,500 to $54,999	1,388 ; 1,006	1,105 ; 767	106 ; 105	46 ; 46	119 ; 63
23	$55,000 to $57,499	2,005 ; 1,355	1,545 ; 1,027	159 ; 132	84 ; 76	190 ; 93
24	$57,500 to $59,999	1,061 ; 715	850 ; 560	89 ; 69	29 ; 26	70 ; 54
25	$60,000 to $62,499	2,752 ; 1,593	2,087 ; 1,206	236 ; 168	130 ; 97	271 ; 108
26	$62,500 to $64,999	937 ; 640	747 ; 517	73 ; 73	33 ; 18	82 ; 23
27	$65,000 to $67,499	1,565 ; 1,037	1,223 ; 805	117 ; 83	65 ; 60	140 ; 74
28	$67,500 to $69,999	882 ; 425	737 ; 336	52 ; 34	22 ; 18	59 ; 30
29	$70,000 to $72,499	1,912 ; 945	1,512 ; 685	114 ; 108	106 ; 71	149 ; 65
30	$72,500 to $74,999	717 ; 388	608 ; 304	37 ; 36	32 ; 18	31 ; 22
31	$75,000 to $77,499	1,495 ; 671	1,176 ; 511	109 ; 49	77 ; 62	103 ; 39
32	$77,500 to $79,999	685 ; 288	558 ; 213	55 ; 39	27 ; 14	31 ; 17
33	$80,000 to $82,499	1,490 ; 644	1,226 ; 466	80 ; 70	75 ; 66	91 ; 33
34	$82,500 to $84,999	483 ; 268	385 ; 197	32 ; 34	24 ; 24	36 ; 12
35	$85,000 to $87,499	797 ; 396	637 ; 308	40 ; 30	57 ; 29	44 ; 23
36	$87,500 to $89,999	387 ; 226	326 ; 185	11 ; 10	10 ; 13	32 ; 14
37	$90,000 to $92,499	954 ; 431	756 ; 319	59 ; 42	66 ; 45	62 ; 16
38	$92,500 to $94,999	351 ; 174	286 ; 136	25 ; 19	19 ; 9	15 ; 8
39	$95,000 to $97,499	577 ; 253	486 ; 193	31 ; 17	33 ; 19	24 ; 19
40	$97,500 to $99,999	326 ; 122	273 ; 104	21 ; 4	11 ; 10	17 ; 1
41	$100,000 to $149,999	5,885 ; 2,264	4,843 ; 1,779	251 ; 165	415 ; 182	305 ; 113
42	$150,000 to $199,999	2,054 ; 483	1,765 ; 390	72 ; 28	96 ; 33	101 ; 20
43	$200,000 to $249,999	760 ; 214	683 ; 173	7 ; 9	41 ; 16	30 ; 14
44	$250,000 and above	1,329 ; 227	1,125 ; 187	36 ; 16	86 ; 18	69 ; 5

parameters, say Θ, in Table 2 for the ETSP (3) and BR (2) distributions were estimated by minimizing for a subpopulation by gender, or gender and ethnicity, the objective function

$$\sum_{i=2}^{43}\{F(x_i|\Theta) - \widehat{F}(x_i)\}^2, \tag{24}$$

where $\widehat{F}(x_i) = \frac{1}{n}\sum_{j=2}^{i} n_i, n_i =$#households in bracket $[x_{i-1}, x_i), i = 2, \ldots, 43$ and $n = \sum_{j=2}^{43} n_i$ is the total number of households in that class within the income range $[\$2,500, \$250,000)$ specified in Table 1. Starting values for the parameters Θ in the least squares procedure (24) were selected by visually matching an ETSP or BR density to the empirical density functions also displayed in Figures 4 and 5 for overall shape. Comparable QQ-plots to the ones displayed in Figures 1C and Figure 1D were obtained for all income distributions that were fitted in Figures 4 and 5, but are omitted here.

Table 3 provides the density values by gender and by gender and ethnicity at the lower ($x = 0$) and upper bounds ($x = 1$) in Figures 4 and 5. One observes from Table 3, with the exception of the density values at the lower bound for Hispanic and white males, that all density values are strictly positive. By design, the strictly positive density values at the lower and upper bound for the BR fitted densities are the same, whereas the density values at the bounds in Table 3 for the fitted ETSP densities differ without exception (recall the discussion in the introduction). This demonstrates the usefulness of allowing for such a flexibility if indeed those ETSP densities provide a "better" fit than the BR fitted ones. While QQ- plots for the fitted distributions in Figures 4 and 5 demonstrate a goodness-of-sit similar to that demonstrated in Figures 1C and D, observed differences in them between ETSP and BR do not offer a means for formal discrimination between the two. Hence, to that end Tables 4 and 5 provides the sum-of-squares statistic (SS) (24), the Log-Likelihood statistic (Ln-L)

$$\sum_{i=2}^{43} n_i \times Ln\{F(x_i|\Theta) - F(x_{i-1}|\Theta)\} \tag{25}$$

and the Kolmogorov-Smirnov statistic (KS)

$$\max_{i=2,\ldots,43} |F(x_i|\Theta) - \widehat{F}(x_i)| \tag{26}$$

for the fitted income distribution to the data in Table 2 within the range $[\$2,500, \$250,000)$. In Tables 4 and 5 values appearing in a bold font indicate that density that outperforms the other one in that particular statistic. Next, favoring that distribution in Tables 4 and 5 that outperforms the other one in at least 2 out of the three statistics (24), (25) and (26) leads to six out of the ten cases preferring the fitted ETSP density (3), with different strictly positive density values at their bounds, over the BR (2) fitted one. In particular in the case of the female subpopulation (Table 5), the ETSP density is selected in four out of the five fitted cases. Certainly, it is no stretch to observe that no evidence in the empirical densities in Figure 4 and 5 calls for equality of density values at their boundaries, nor is their any evidence within these figures that suggest that zero density values at the boundaries are a requirement. Thus, in a broader sense, the analysis above supports both the usefulness of the ETSP (3) and BR (2) densities.

In Table 6 we present some statistics for the selected fitted income distributions by gender and by gender and ethnicity. Some interesting observations

Table 2 Parameter estimates of LSQ (24) fitted distributions in Figures 4 and 5 by gender.

		ETSP	BR
Male	All	$\theta = 0.085$, $m = 1.143$, $n = 6.391$, $\lambda = 0.250$, $\delta = 0.086$	$\alpha = 1.339$, $\beta = 7.807$, $\delta = 0.083$
	White	$\theta = 0.110$, $m = 1.098$, $n = 5.873$, $\lambda = 0.000$, $\delta = 0.094$	$\alpha = 1.394$, $\beta = 7.176$, $\delta = 0.084$
	Black	$\theta = 0.020$, $m = 2.97$, $n = 7.588$, $\lambda = 0.750$, $\delta = 0.001$	$\alpha = 1.204$, $\beta = 9.021$, $\delta = 0.021$
	Asian	$\theta = 0.088$, $m = 1.974$, $n = 5.315$, $\lambda = 1.000$, $\delta = 0.123$	$\alpha = 1.164$, $\beta = 5.849$, $\delta = 0.102$
	Hispanic	$\theta = 0.072$, $m = 1.206$, $n = 10.669$, $\lambda = 0$, $\delta = 0.062$	$\alpha = 1.599$, $\beta = 13.841$, $\delta = 0.051$
Female	All	$\theta = 0.021$, $m = 1.947$, $n = 9.034$, $\lambda = 0.624$, $\delta = 0.031$	$\alpha = 1.172$, $\beta = 10.265$, $\delta = 0.038$
	White	$\theta = 0.024$, $m = 1.954$, $n = 8.569$, $\lambda = 0.658$, $\delta = 0.030$	$\alpha = 1.189$, $\beta = 9.882$, $\delta = 0.039$
	Black	$\theta = 0.017$, $m = 3.007$, $n = 9.494$, $\lambda = 0.659$, $\delta = 0.009$	$\alpha = 1.229$, $\beta = 11.731$, $\delta = 0.020$
	Asian	$\theta = 0.000$, $m = 1.959$, $n = 6.960$, $\lambda = 1.000$, $\delta = 0.051$	$\alpha = 0.959$, $\beta = 6.489$, $\delta = 0.037$
	Hispanic	$\theta = 0.024$, $m = 1.992$, $n = 12.448$, $\lambda = 0.745$, $\delta = 0.027$	$\alpha = 1.218$, $\beta = 14.272$, $\delta = 0.028$

Table 3 Density values at lower and upper bounds for ETSP (3) and BR (2) fitted densities displayed in Figures 4 and 5 and by gender only.

		ETSP		BR	
		LB	UB	LB	UB
Male	All	0.979	0.063	0.083	0.083
	White	0	0.066	0.084	0.084
	Black	3.508	0.001	0.021	0.021
	Asian	3.586	0.092	0.102	0.102
	Hispanic	0	0.041	0.051	0.051
Female	All	3.629	0.028	0.038	0.038
	White	3.668	0.026	0.039	0.039
	Black	3.569	0.008	0.020	0.020
	Asian	6.658	0.051	∞	0.037
	Hispanic	5.973	0.022	0.028	0.028

Table 4 Sum of Squared (SS) error, Log-Likelihood (Ln-L) and Kolmogorov- Smirnov (KS) statistics for ETSP (3) and BR (2) densities displayed in Figure 4.

Male		All	White	Black	Asian	Hispanic
ETSP	SS	0.00095	0.00097	0.00294	***0.00201***	***0.00153***
	Ln-L	***-354,579***	***-249,999***	-34,692	***-15,024***	***-47,479***
	KS	0.01088	0.00989	***0.01913***	***0.01522***	***0.01601***
BR	SS	***0.00087***	***0.00081***	***0.00210***	0.00241	0.00203
	Ln-L	-354,678	*-250,033*	***-34,667***	-15,030	-47,540
	KS	***0.00994***	***0.00919***	0.02222	0.01685	0.01791

Table 5 Sum of Squared (SS) error, Log-Likelihood (Ln-L) and Kolmogorov- Smirnov (KS) statistics for ETSP (3) and BR (2) densities displayed in Figure 5.

Female		All	White	Black	Asian	Hispanic
ETSP	SS	***0.00051***	***0.00060***	***0.00143***	0.00254	***0.00101***
	Ln-L	***-325,079***	***-231,865***	***-40,029***	*-14,150*	***-33,561***
	KS	***0.00779***	***0.00858***	***0.01437***	***0.01696***	***0.01341***
BR	SS	0.00098	0.00085	0.00173	***0.00225***	0.00160
	Ln-L	-325,538	-232,146	-40,121	***-14,143***	-33,599
	KS	0.01397	0.01186	0.01880	0.01957	0.01598

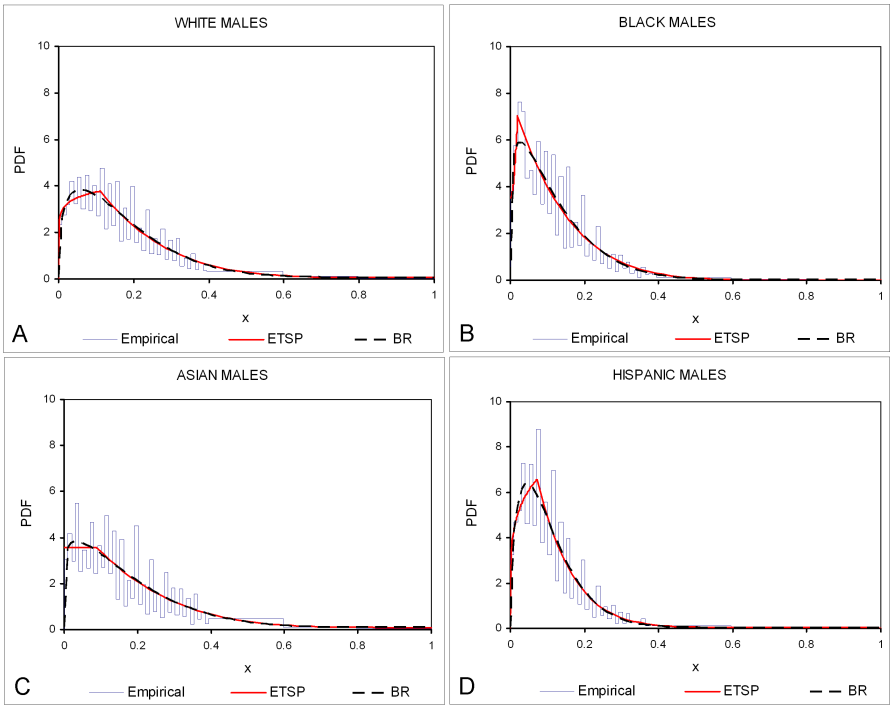

Fig. 4 A Comparison of ETSP (3) and BR (2) fitted densities to standardized US 2008 annual income data for males by ethnicity within the range $[\$2,500; \$250,000]$; A: White males; B: Black males; C: Asian males; D: Hispanic males.

follow immediately from Table 6. Firstly, the average annual income distribution for all females ($\$31,248$) is less than the minimum average annual income by male and ethnicity, i.e. that of Hispanic males ($\$32,712$). Secondly, from the mean income analysis by gender, one immediately observes that the mean income for the Asian subpopulations exceeds those of the others regardless of gender ($\$51,793$ for Asian males and $\$37,769$ for Asian females), whereas the mean income of the Hispanic subpopulations are the lowest ($\$32,712$ for Hispanic males and $\$24,636$ for Hispanic females). Of male to female comparisons, the difference in mean annual income is least for the black subpopulation ($\$33,625 - \$27,510 = \$6,115$). Similar conclusions follow by comparing other locations statistics provided in Table 6, including the 1st quartile of the income distributions. Observe from the skewness and kurtosis values in Table 6 that the fitted distributions may be considered quite skewed, peaked and heavy-tailed when plotting their relative locations in the moment-ratio diagram in Figure 3.

An income distribution analysis would not be complete without an evaluation of what is called the Gini-index also provided in Table 6 for the fitted distributions therein. The Gini-index is a measure of income inequality and is derived from the Lorentz curve. The Lorentz curve plots the cumulative %

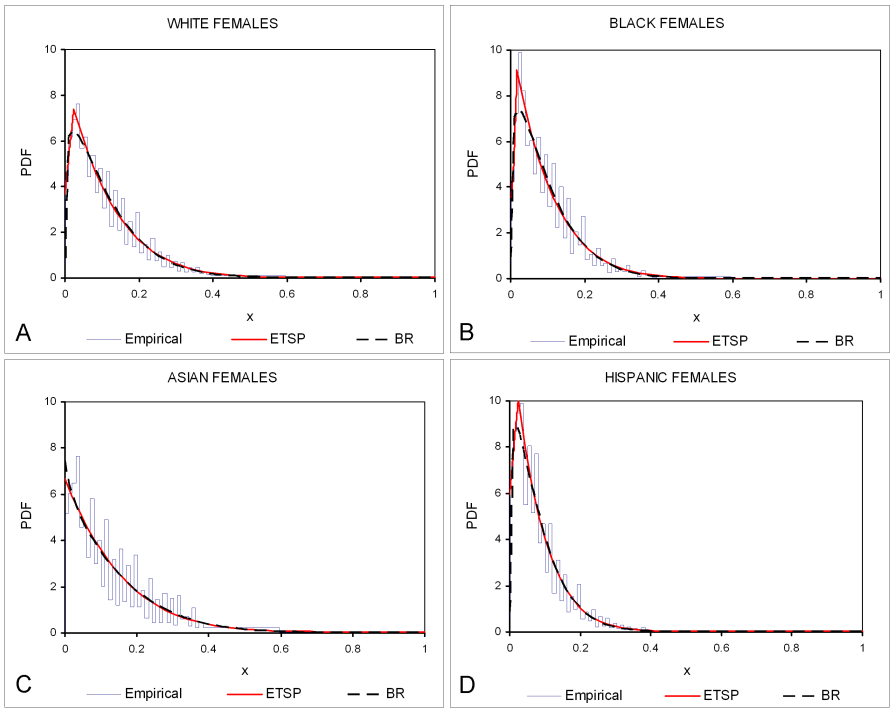

Fig. 5 A Comparison of ETSP (3) and BR (2) fitted densities to standardized US 2008 annual income data for females by ethnicity within the range $[\$2,500; \$250,000]$; A: White females; B: Black females; C: Asian females; D: Hispanic females.

Table 6 US 2008 income statistics derived from distributions in Figures 4 and 5. Rows with (E) use ETSP (3) fitted pdf's, rows with (B) use BR(2) fitted pdf's (see Table 3).

		Mean	Median	Mode	$x_{0.25}$	St. Dev.	CV	β_1	β_2	Gini
All	Males (B)	$46,020	$34,666	$14,239	$18,440	$41,249	0.896	2.160	8.847	46.2%
All	Females (E)	$31,248	$22,842	$7,616	$11,989	$29,701	0.950	2.908	16.009	48.1%
Male	Asian (E)	$51,793	$38,827	$24,341	$19,763	$45,230	0.873	1.721	6.393	46.7%
	White (B)	$49,804	$38,668	$17,351	$20,837	$42,034	0.844	1.944	7.775	44.4%
	Black (B)	$33,625	$26,025	$8,636	$13,635	$29,082	0.865	2.450	13.213	45.8%
	Hispanic (E)	$32,712	$24,784	$20,376	$14,769	$30,876	0.944	3.304	18.281	45.4%
Female	Asian (B)	$37,769	$27,307	$2,500	$12,770	$35,702	0.945	2.164	9.674	49.0%
	White (E)	$32,402	$23,946	$8,462	$12,583	$29,924	0.924	2.721	14.609	47.5%
	Black (E)	$27,510	$21,004	$6,680	$11,366	$23,236	0.845	2.556	15.661	45.6%
	Hispanic (E)	$24,636	$17,846	$8,437	$9,907	$25,112	1.019	3.986	27.192	48.8%

income made by a population against the cumulative % of that population, and may be conveniently plotted using the following relationship given a cdf $F(\cdot)$ and pdf $f(\cdot)$

$$L\{F(x)\} = \frac{\int_0^{F(x)} F^{-1}(v)dv}{E[X]} = \frac{\int_0^x uf(u)du}{E[X]}, \tag{27}$$

see, e.g., Sarabia (2008). For the ETSP distributions (3) we have

$$\int_0^x ug\{u|\Theta\}du =$$
$$\begin{cases} \pi(\Theta)\theta\Big\{\frac{\lambda}{2}(\frac{u}{\theta})^2 + \frac{(1-\lambda)m}{m+1}(\frac{u}{\theta})^{m+1}\Big\}, & \text{for } 0 < x \le \theta, \\ \pi(\Theta)\theta\Big\{\frac{\lambda}{2} + \frac{(1-\lambda)m}{m+1}\Big\} + \frac{1-\pi(\Theta)}{1-\theta}\int_\theta^x x\big\{\delta + (1-\delta)n(\frac{1-x}{1-\theta})^{n-1}\big\}dx, & \text{for } \theta < x < 1, \end{cases}$$

where

$$\int_\theta^x x\{\delta + (1-\delta)n\big(\frac{1-x}{1-\theta}\big)^{n-1}\}dx = (1-\theta)\Big\{1 - \delta\big(\frac{1-u}{1-\theta}\big) - (1-\delta)\big(\frac{1-u}{1-\theta}\big)^n\Big\}$$
$$-(1-\theta)^2\Big[\frac{\delta}{2}\Big\{1 - \big(\frac{1-u}{1-\theta}\big)^2\Big\} + \frac{(1-\delta)n}{n+1}\Big\{1 - \big(\frac{1-u}{1-\theta}\big)^{n+1}\Big\}\Big].$$

For the BR distributions (2), plotting of the Lorentz curves require the use of the incomplete beta $B(x|\alpha,\beta)$ function in both evaluations of the left and right hand sides of (27), where

$$B(x|\alpha,\beta) = \int_0^x u^{\alpha-1}(1-u)^{\beta-1},\ \alpha,\beta > 0.$$

Figure 6 plots the Lorentz curves for the male and female fitted BR and ETSP fitted distributions in the first two rows of Table 6, whereas Figure 7 plots them by gender and ethnicity for the other rows. The $y = x$ line in Figure 6 represents complete equality of income amongst all within a population in question. Complete inequality would be represented by one individual making all the income and a Lorentz curve equal to 0%, up to but not including 100% and jumping to $y = 100\%$ at $x = 100\%$. Please observe from Figure 6 that approximately the lower 80% of the female or male population accounts for about 50% of the cumulative annual income in both of these subpopulations.

The Gini-index for a particular Lorentz curve is given by $a/(a+b)$ or $1-2b$ since $a + b = 0.5$, where the areas a and b are indicated in Figure 6. Thus the Gini-index ranges from 0 to 1, but is often expressed as a percentage (see, e.g., Kleiber and Kotz, 2003). One concludes that the larger the area a (or the lesser the area b), the larger the income inequality within the corresponding population. Observe from Figure 6 that, perhaps surprisingly, the income inequality as measured by the fitted distributions in Table 6 for the male subpopulation is less than the income inequality for the female subpopulation, given the higher average income for males in Table 6 ($\$46,020$) compared to that of the females in Table 6 ($\$31,248$). Hence, one concludes that the Gini-index measures income inequality within a given subpopulation, but does not provide for a means for comparing income levels across subpopulations. One observes from Figure 7A no particular ordering in terms of income inequality amongst the fitted Lorentz curves for males, whereas in Figure 7B one observes more of an ordering with the highest income inequality attributable to the Asian female subpopulation (49.2%). The lowest value for the Gini-index in Table 6 follows for the white male subpopulation (44.2%). Along the same

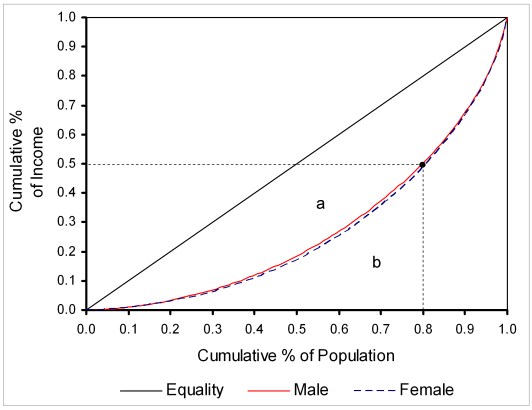

Fig. 6 Lorentz Curves (27) for ETSP (3) and BR (2) fitted densities by gender indicated in Table 6; Vertical and horizontal dotted lines show that top 20% within the male and female sub-groups make about 50% of total cumulative income; Gini-indeces in Table 6 follow from ratios $a/(a+b) = 1 - 2b$, where areas a and b are indicated above.

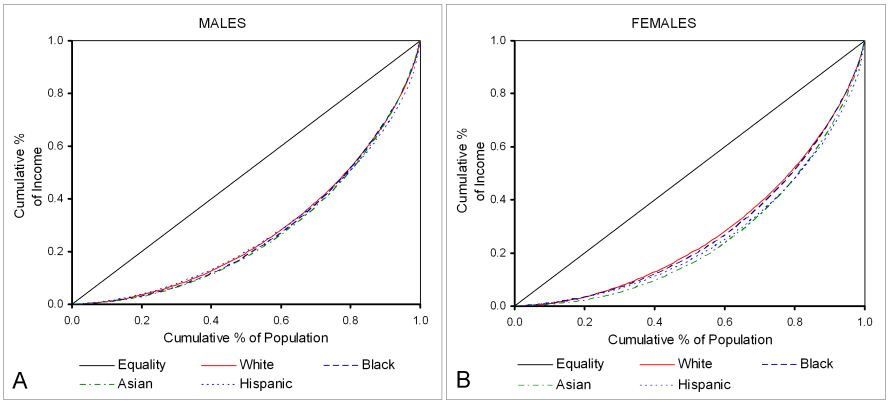

Fig. 7 Lorentz Curves (27) for ETSP (3) and BR (2) fitted densities indicated in Table 6 by ethnicity; A: Males; B: Females.

lines it is interesting to note that the Gini-index for the fitted distribution for the Asian female subpopulation (49.2%) ranks highest amongst all, whereas amongst the female subpopulations they too rank highest in terms of average annual income. Finally, please observe a similar income inequality measured by the fitted distributions for the black male (45.8%) and female (45.6%) populations, whereas, as mentioned above, the average annual income for black males is $\$6,115$ higher than that of black females.

Given than that the Gini-index only measures income inequality within a subpopulation, a natural question would be which of the characteristics: mean, median, mode, 1st quartile, standard deviation, coefficient of variation (CV), skewness and kurtosis listed in Table 6 (which can be used to described

Table 7 Best subset regression results using the Gini-index in Table 6 as the dependent variable and the mean, median, mode, 1st quartile ($x_{0.25}$), standard deviation, coefficient of variation, skewness (β_1), kurtosis (β_2) as independent variables.

# Var	R-Sq	R-Sq (adj)	Mean	Median	Mode	$x_{0.25}$	St. Dev.	CV	β_1	β_2
1	61.1	56.2						x		
1	35.7	27.7				x				
2	94.2	92.5				x	x			
2	90.2	87.4	x	x						
3	98.8	98.2	x			x			x	
3	98.3	97.4	x			x		x		
4	99.1	98.4	x	x		x			x	
4	99.0	98.2	x			x			x	x
5	99.2	98.3	x	x		x	x	x		
5	99.2	98.2	x	x	x	x			x	
6	99.6	98.7	x	x		x	x		x	x
6	99.3	97.9	x	x		x		x	x	x
7	99.9	99.4	x	x	x	x		x	x	x
7	99.7	98.7	x	x	x	x	x		x	x
8	99.9	99.5	x	x	x	x	x	x	x	x

differences in income levels across subpopulations) are also determinants for the Gini-index values provided in Table 6. Hence to that end, a best subset regression was performed utilizing the off-the-shelf statistical software Minitab and the results thereof are displayed in Table 7. Please observe from Table 7 that the first quartile $x_{0.25}$ of the fitted income distributions appears in all but two of the subset regressions, whereas the first quartile $x_{0.25}$ and the standard deviation combined account for 92.5% of the sample variance of the Gini-indeces values in Table 6. In fact, if one regresses the Gini-indeces in Table 6 against the first quartile and the standard deviation one obtains

$$\text{Gini} = 0.477 - 0.00973 \times x_{0.25} + 0.00399 \times \text{St.Dev.}, \tag{28}$$

where in (28) $x_{0.25}$ and St.Dev. are measured in terms of thousands of dollars. Hence, from (28) one may cautiously infer that an increase of $\$1,000$ in the first quartile $x_{0.25}$ of one of the income distributions in Table 6 would result in about a decrease of 1% in the Gini-index value, given that an increase in $x_{0.25}$ would not result in an increase of the standard deviation (which is unlikely since a decrease could be expected). These analysis results therefore reinforce the relevance of the first quartile of an income distribution in describing racial and gender income gaps across subpopulations as earlier concluded by Barsky et al. (2002). Please recall the introduction of this illustrative US 2008 income distribution analysis example.

Acknowledgements The authors are thankful to editor and the referees of earlier versions of this paper whose valuable comments improved the contents and presentation.

References

1. Aban, I.B., Meerschaert, M.M., Panorska, A.K., Parameter estimation for the truncated Pareto distribution, Journal of the American Statistical Association, 101(473), 270 - 277 (2006).
2. Adler, R., Feldman, R., Taqqu, M., A Practical Guide to Heavy Tails. Birkhäuser, Boston (1998).
3. Arnold, B. C., Pareto Distributions. International Cooperative Publishing House, Fairland, MD, USA (1983).
4. Barkai, E., Metzler, R., Klafter, J., From continuous time random walks to the fractional Fokker-Planck equation, Physics Review, E 61, 132-138 (2000).
5. Barsky, R., Bound, J., Kerwin, K.C. and Lupton, J.P., Accounting for the black-white wealth gap: a nonparametric approach, Journal of the American Statistical Association, 97 (459), 663-673, (2002).
6. Clementi, F. , Gallegatib, M., Power law tails in the Italian personal income distribution, Physica A: Statistical Mechanics and its Applications, 350 (2-4), 427-438, (2005).
7. Coelho, R., Richmond, P., Barrya, J., Hutzlera, S., Double power laws in income and wealth distributions, Physica A: Statistical Mechanics and its Applications, 387 (15), 3847-3851 (2008).
8. Douglas, E.M., Barros, A.P., Probable Maximum Precipitation Estimation Using Multifractals: Application in the Eastern United States, Journal of Hydrometeorology 4(6), 1012-1024 (2003).
9. Elderton, W.P., Johnson, N.L., Systems of frequency curves. Cambridge University Press, London (1969).
10. Embrechts, P., Klüppelberg, C., Mikosch, T., Modelling Extremal Events for Insurance and Finance, Springer-Verlag, Berlin (1997).
11. Fernandez, C., Steel, M.F.J., On Bayesian modeling of fat tails and skewness, Journal of the American Statistical Association, 93, 359-371, (1998).
12. Gomez, H.W., Torres, F.J., Bolfarine, H., Large-sample inference for the epsilon-skew-distribution, Communications in Statistics - Theory and Methods, 36, 73-81, (2007).
13. Hahn, E.D., Mixture densities for project management activity times: A robust approach to PERT, European Journal of Operational Research 188, 450-459, (2008).
14. Herrerías-Velasco, J.M., Herrerías-Pleguezuelo, R., van Dorp, J.R., The generalized two-sided power distribution, Journal of Applied Statistics 36 (5), pp. 573-587, (2008).
15. Kleiber, C., Kotz, S., Statistical Size Distributions in Economics and Actuarial Sciences. John Wiley, Hoboken, NJ, (2003).
16. Kotz, S., Johnson, N.L., Moment ratio diagrams, In: Encyclopedia of statistical sciences, Vol.5. Wiley, New York (1985).
17. Kotz, S., Kozubowski, T.J., Podgórski K, The Laplace distribution and generalizations. Birkhäuser, Boston, MA (2001).
18. Kotz, S., van Dorp, J.R., Beyond beta: other continuous families of distributions with bounded support and applications. World Scientific Press, Singapore (2004).
19. Kotz, S., van Dorp, J.R., A link between Two-Sided Power and Asymmetric Laplace Distributions: with Applications to Mean and Variance Approximations, Statistics and Probability Letters 71, 382-394, (2005).
20. Lévy, P., Calcul des probabilitiés, 2nd part, Chapter VI. Gauthier-Villars, Paris (1925).
21. Levy, H., Duchin, R., Asset return distribution and the Investment Horizon, explaining contradictions, The Journal of Portfolio management 30(3), 47-62 (2004).
22. Lu, S.L., Molz, F.J., How well are hydraulic conductivity variations approximated by additive stable processes?, Advances in Environmental Research 5(1), 39-45 (2001).
23. McFall Lamm, Jr. R., Asymmetric returns and optimal hedge fund portfolios, The Journal of Alternative Investments, 9-21, (2003).
24. McCulloch, J., Financial Applications of Stable Distributions. In Statistical Methods in Finance, eds. G. Madfala and C. R. Rao, pp. 393-425. Elsevier, Amsterdam (1996).
25. Miyazima, S., Yamamoto K., Power-law behaviors in high income distribution, Practical Fruits of Econophysics, 5, 344-348, (2006).
26. Nagahara, Y., The PDF and CDF of Pearson Type IV distributions and the ML estimation of the parameters, Statistics & Probability Letters 43, 251-264 (1999).

27. Pareto, V., Cours d'Économie Politique. Rouge, Lausanne et Paris (1897).
28. Resnick, S., Heavy tail modeling and teletraffic data, Annals of Statistics, 25, 1805–1869, (1997).
29. Samorodnitsky, G., Taqqu, M., Stable non-Gaussian Random Processes. Chapman & Hall, New York (2004).
30. Sarabia, J.M., Parametric Lorenz curves: models and applications, Modeling Income Distributions and Lorenz Curves. Series: Economic Studies in Inequality, Social Exclusion and Well-Being 4, Chotikapanich, D. (Ed.), pp. 167-190. Springer-Verlag, Berlin (2008).
31. Singh, A., van Dorp, J.R., Mazzuchi, T.A., A novel assymetric distribution with power tails, Communications in Statistics, Theory and Methods, 36(2), 235-252 (2007).
32. Solomon, S., Levy, M., Market ecology, Pareto wealth distribution and leptokurtic returns in microscopic simulation of the LLS stock market model. Proceedings of complex behavior in economics, Aix en Provence (Marseille), France, May 4-6, (2000).
33. Stuart, A., Ord, J.K., Kendall's advanced theory of statistics. Wiley, New York (1994).
34. Van Dorp, J.R., Kotz, S., The standard two sided power distribution and its properties: with applications in financial engineering, The American Statistician 56(2), 90-99, (2002).
35. Van Dorp, J.R., Kotz, S., Generalizations of two-sided power distributions and their convolution, Communications in Statistics: Theory and Methods 32(9), 1703-1723 (2003).
36. Zabell, S.L., On Student's 1908 Article "The Probable Error of a Mean", Journal of the American Statistical Association 103 (481),1 - 7 (2008).