Chapter 1

MODELING INCOME DISTRIBUTIONS USING ELEVATED DISTRIBUTIONS ON A BOUNDED DOMAIN

J. RENÉ VAN DORP

Engineering Management and Systems Engineering Department
The George Washington University
1776 G street, Suite 110, NW, Washington DC, 20052

SAMUEL KOTZ

Engineering Management and Systems Engineering Department
The George Washington University
1776 G street, Suite 110, NW, Washington DC, 20052

This paper presents a new two parameter family of continuous distribution on a bounded domain which has an elevated but finite density value at its lower bound. Such a characteristic appears to be useful, for example, when representing income distributions at lower income ranges. The family generalizes the one parameter Topp and Leone distribution originated in the 1950's and recently rediscovered. The family of beta distributions has been used for modeling bounded income distribution phenomena, but it only allows for an infinite and zero density values at its lower bound, and a constant density of 1 in case of its uniform member. The proposed family alleviates this apparent jump discontinuity at the lower bound. The U.S. Income distribution data for the year 2001 is used to fit distributions for Caucasian (Non-Hispanic), Hispanic and African-American populations via a maximum likelihood procedure. The results reveal stochastic ordering when comparing the Caucasian (Non-Hispanic) income distribution to that of the Hispanic or African-American population. The latter indicates that although substantial advances have reportedly been made in reducing the income distribution gap amongst different ethnic groups in the U.S. during the last 20 years or so, these differences still exist.

1. Introduction

In a 1955 issue of the Journal of the American Statistical Association an isolated paper on a bounded continuous distribution by Topp and Leone [1] appeared which received little attention. The paper was re-discovered by Nadarajah and Kotz [2] and motivated by investigations of van Dorp and Kotz [3,4] on the Two-Sided Power (TSP) distribution and other alternatives to the

DISTRIBUTION MODELS THEORY

popular and versatile beta distribution which has been used in various applications for over a century. Even in the late nineties of the 20th century the arsenal of bounded univariate distributions contained very few members. Amongst them, the triangular and uniform distribution are the most widely used together with some "curious" distributions appearing as problems or exercises in various Mathematical and Statistical journals. Other, somewhat artificial empirical bounded continuous distributions are based on mathematical transformations of the normal distribution (of an unbounded domain) - the most wide spread amongst them are perhaps the Johnson [5] family of transformations. On the other hand the existence of multitudes of unbounded continuous distributions developed in the 20th century is well known and amply documented.

The construction of the Topp and Leone distribution is quite straightforward and based on the principle that by raising an arbitrary cdf $F(x) \in [0,1]$ to an arbitrary power $\beta > 0$, a new cdf $G(x) = F^{\beta}(x)$ emerges with one additional parameter. This devise was used in 1939 by W. Weibull [6] proposing his Weibull distribution, which has achieved substantial popularity the second part of the 20th century, especially in reliability and biometrical applications. The cdf F(x) in the above construction method may be referred to as the generating cdf. Figure 1 demonstrates the construction of the Topp and Leone distribution. The generating density of the Topp and Leone family is the right triangular density $(2-2x), x \in [0,1]$. It is displayed in Figure 1A. Figure 1B depicts its cdf $(2x-2x^2)$ and Figures 2C and 2D plot the pdf and cdf of a one parameter Topp and Leone distribution for $\beta = 3$. Note, the appearance of a mode in the

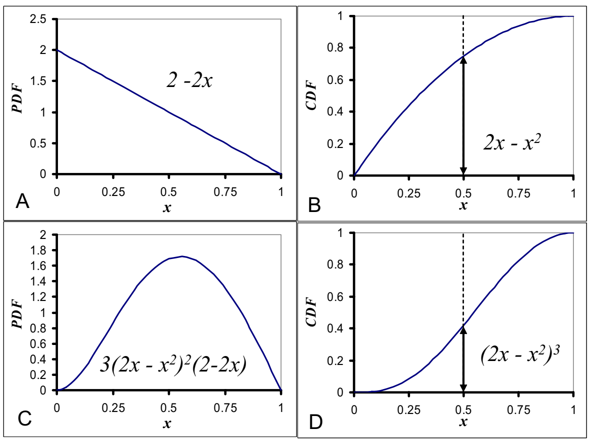

Figure 1. Construction of Topp and Leone distribution from a right triangular distribution

pdf presented in Figure 1C due to S-shapedness of the corresponding cdf in Figure 1D obtained by using a cdf transformation with $\beta > 1$. Topp and Leone's [1] original interest focused on the construction of J-shaped distributions utilizing similar cdf transformations with $0 < \beta < 1$; They have fitted their distribution to transmitter tubes failure data. Nadarajah and Kotz [2] showed that the J-shaped Topp and Leone distributions exhibit a bath tub failure rate functions with natural applications in reliability.

Our generalization of the Topp and Leone distribution (GTL) utilizes a slightly more general slope distribution with pdf $\alpha \mathrm{x} - 2(\alpha - 1)\mathrm{x}, 0 \le \alpha \le 2$, as the generating density (see Figure 2A with $\alpha = 1.5$), where $\mathrm{x} \in (0,1)$. Slope distributions possess linear pdf's and play a central role in deriving a generalization of the trapezoidal distribution (see, e.g., Van Dorp and Kotz [7]). From the restriction that $\alpha \mathrm{x} - 2(\alpha - 1)\mathrm{x} \ge 0$ for all $\mathrm{x} \in (0,1)$, it follows that $0 \le \alpha \le 2$. For $\alpha \in [0,1)$ $\left(\alpha \in (1,2]\right)$, the slope of the pdf is increasing (decreasing). For $\alpha = 1$, the slope distribution (1) simplifies to a uniform distribution on $(0,1)$. Figure 2B plots the cdf of the linear pdf in Figure 2A.

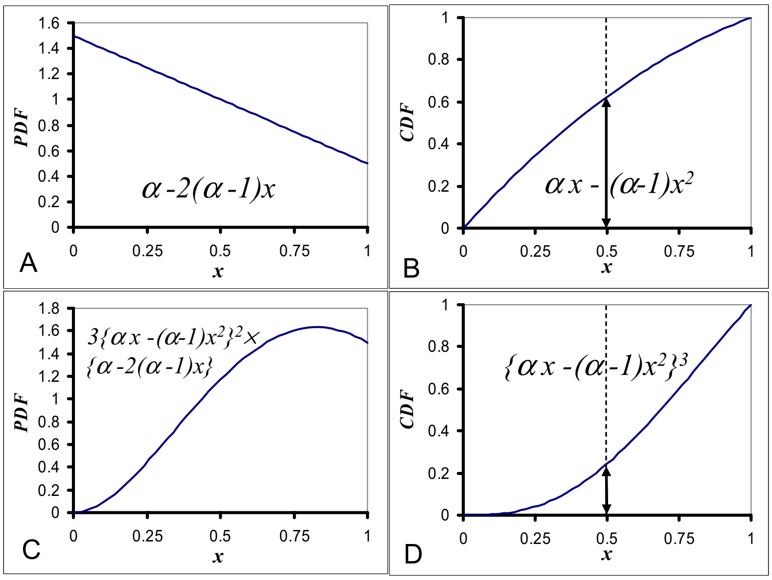

Figure 2. Construction of generalized Topp and Leone distribution from a slope distribution

Now the Generalized Topp and Leone (GTL) distribution that follows from Figure 2B (utilizing the above construction method with $\beta = 3$) is depicted in Figure 2D. The density associated with this cdf is displayed in Figure 2C. Note that, while a mode in $(0,1)$ is present in Figure 2C, it has been shifted to the right when compared to the situation in Figure 1C. More importantly, the density at the upper bound is strictly positive in Figure 2C while being zero in Figure 1C (representing the original Topp and Leone density).

Our main interest in this paper is to represent income distributions. We shall therefore consider the reflected version of the Generalized Topp and Leone (GTL) distribution utilizing the cdf transformation $H(x) = 1 - G(1-x)$, where G is a GTL cdf on $[0,1]$. The latter transformation typically assigns the mode towards the left hand side of its support and allows for strictly positive density values at the lower bound. This form seems to be appropriate when representing income distributions at lower income ranges. (Compare, e.g., with Figure 2 of Barsky et al. [8], p. 668). The U.S. Income distribution data for the year 2001 is used to fit Reflected GTL (RGTL) distributions for Caucasian (Non-Hispanic), Hispanic and African-American populations via a maximum likelihood procedure. The results reveal stochastic ordering when comparing the Caucasian (Non-Hispanic) income distribution to that of the Hispanic or African-American populations. In particular when comparing Americans of Caucasian Origin, African-Americans appear to be approximately 1.9 times as likely and the Hispanics 1.5 times as likely to have inadequate or no income at all. The latter indicates that although substantial advances have indeed occurred in reducing the income distribution gap amongst different ethnic groups in the U.S. during the last 20 years or so (see, e.g., Couch and Daly [9]), these differences still exist.

Another reason to consider reflected GTL distribution rather than GTL distributions is that a drift of the mode towards the left hand side mimics the behavior of the classical unbounded continuous distributions such as the Gamma, Weibull and Lognormal. (We note, in passing, that these three distributions are in a strong competition amongst themselves as to which is the best one for fitting numerous phenomena in economics, engineering and medical applications). One can therefore conjecture that application of Reflected GTL (RGTL) distributions may not be limited to the area of income distributions.

In Section 2, we shall present the cdf and pdf of a four parameter RGTL distribution and investigate its various forms. In Section 3, we will elaborate on some properties of RGTL distributions. Moment expressions for RGTL distributions, to the best of our knowledge, cannot be derived in closed form (except for certain special cases). The cdf of the beta distribution while not available in a closed form (whereas that of an RGTL distribution is) is, however, useful for calculating moments of RGTL distributions for $1 < \alpha \leq 2$. In Section 4, we shall discuss a Maximum Likelihood Estimation (MLE) procedure utilizing standard root finding algorithms that are readily available in various software packages such as e.g. Microsoft Excel. In Section 5, we shall fit RGTL distributions to the U.S. 2001 income distribution data with seemingly satisfactory results. Some brief concluding remarks are presented in Section 6.

2. Cumulative distribution function and density function

The four parameter RGTL distribution with support $[a, b]$ the cdf

$$F(x|\, a, b, \alpha, \beta) = 1 - \left(\frac{b-x}{b-a}\right)^{\beta} \left\{\alpha - (\alpha-1)\left(\frac{b-x}{b-a}\right)\right\}^{\beta} \quad (1)$$

where $a \le x \le b$, $0 < \alpha \le 2$ and $\beta > 0$ Evidently, $F(a) = 0$ and $F(b) = 1$. The probability density function (pdf) follows from (1) to be

$$f(x|\, a, b, \alpha, \beta) = \frac{\beta}{b-a}\left(\frac{b-x}{b-a}\right)^{\beta-1} \times \left\{\alpha - (\alpha-1)\left(\frac{b-x}{b-a}\right)\right\}^{\beta-1} \left\{\alpha - 2(\alpha-1)\left(\frac{b-x}{b-a}\right)\right\} \quad (2)$$

with the same constraint on x, α and β as in (1) . From (2) it follows that in particular

$$f(a|\, a, b, \alpha, \beta) = \frac{\beta(2-\alpha)}{b-a} \quad (3)$$

and

$$f(b|\, a, b, \alpha, \beta) = \begin{cases} 0 & \beta > 1 \\ \dfrac{\beta\alpha}{b-a} & \beta = 1 \\ \to \infty \text{ as } x \uparrow b & \beta < 1 \end{cases} \quad (4)$$

Relation (3) shows that the RGTL family allows for arbitrary density values at its lower bound a. Expressions (1) and (2) are reduced to the Topp and Leone distribution (see Topp and Leone [1]):

$$f(x|\, a, b, \alpha, \beta) = \frac{2\beta}{b}\left(\frac{x}{b}\right)^{\beta-1} \left\{2 - \left(\frac{x}{b}\right)\right\}^{\beta-1} \left\{1 - \left(\frac{x}{b}\right)\right\} \quad (5)$$

by setting $a = 0$, $b = 2$ and utilizing the reflection transformation $Y = b + a - X$. Figure 1C depicts a graph of the Topp and Leone distribution with parameters $b = 1$ and $\beta = 3$ in (5). Figure 3A displays its reflected version. Note the transition in the form of graphical representations of the pdf's from Figure 3B to Figure 3D which all have the same value of α with decreasing values of β.

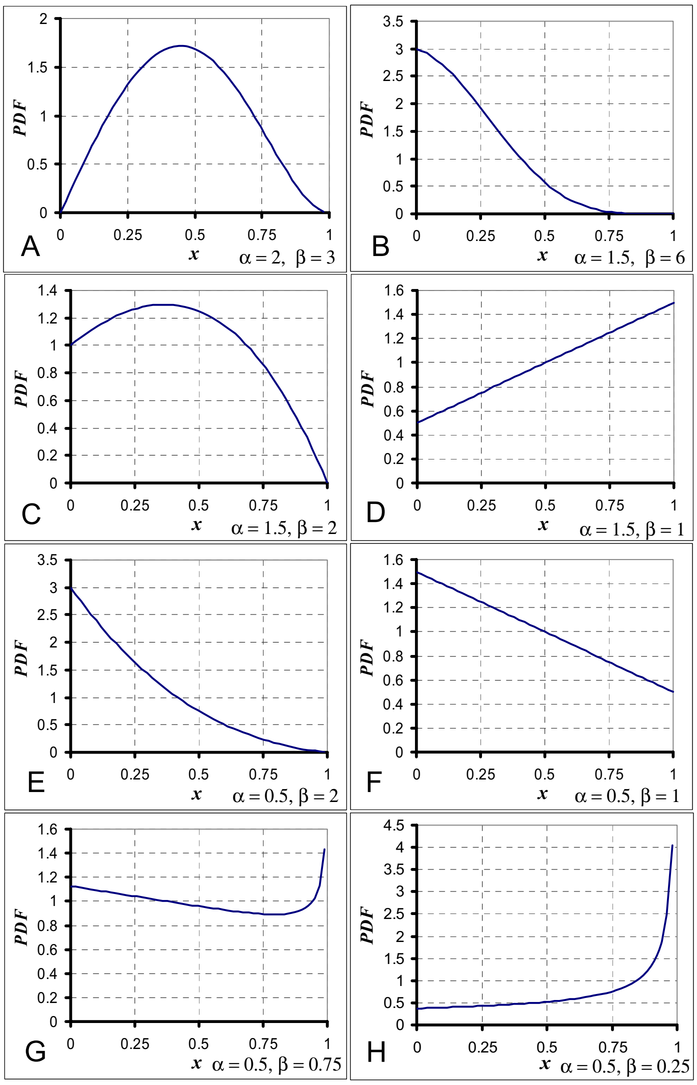

Figure 3. Examples of Standard RGTL distributions ($a = 0$, $b = 1$):
A: $\alpha = 2, \beta = 3$; B: $\alpha = 1.5, \beta = 6$; C: $\alpha = 1.5, \beta = 2$; D: $\alpha = 1.5, \beta = 1$
E: $\alpha = 0.5, \beta = 2$; F: $\alpha = 0.5, \beta = 1$; G: $\alpha = 0.5, \beta = 0.75$; H: $\alpha = 0.5, \beta = 0.25$

Note that in case of Figure 3B the pdf assumes a similar form to that of a reliability function whereas Figure 3C displays a mode at a value greater than 0. Similarly in Figures 3E to 3H the pdf's with the same value of α (= 0.5) with progressively decreasing β from 2 to 0.25, indicate the change in form of the pdf from a monotonically decreasing concave form, a linear function with decreasing slope, a mild U-shaped function, up to a monotonically increasing convex curve.

The J-shaped form of the pdf in Figure 3E ($a = 0$, $b = 1$, $\alpha = 0.5$, $\beta = 2$) resembles that of a Weibull distribution with the shape parameter less than one (but on a bounded domain). Note that the structure of (1) is reminiscent to that of the Weibull cdf. Figures 3G and 3H depict a U-shaped pdf form ($a = 0$, $b = 1$, $\alpha = 0.5$, $\beta = 0.75$) and a J-shaped pdf form ($a = 0$, $b = 1$, $\alpha = 0.5$, $\beta = 0.25$) respectively, and are similar to those appearing in the beta family, but with a bounded density value at its lower bound (cf. (3)). Setting $\alpha = 1$, $\beta = 1$ in (2) yields a uniform distribution on $\left[a,b\right]$. Hence, analogously to the four parameter beta distribution with the pdf

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)\,(b-a)}\left(\frac{x-a}{b-a}\right)^{\alpha-1}\left(\frac{b-x}{b-a}\right)^{\beta-1} \tag{6}$$

where $a \le x \le b$, $\alpha > 0$ and $\beta > 0$ and the Two-Sided Power family (see van Dorp and Kotz [3,4]) with the pdf

$$\begin{cases} \dfrac{n}{b-a}\left(\dfrac{x-a}{m-a}\right)^{n-1} & a < x \le m \\ \dfrac{n}{b-a}\left(\dfrac{b-x}{b-m}\right)^{n-1} & m \le x < b \end{cases} \tag{7}$$

where $n > 0$, the RGTL family has the uniform distribution on $\left[a,b\right]$ as one of its members. Another common member amongst these 3 families (Beta, TSP and RGTL) is the reflected power (RP) distribution on $\left[a,b\right]$ the pdf

$$f(x|\,a,b,\alpha,\beta) = \frac{\beta}{b-a}\left(\frac{b-x}{b-a}\right)^{\beta-1} \tag{8}$$

obtained by substituting $\alpha = 1$ in (2). Substituting $\alpha = 0$ in (2) also yields the reflected power distribution but with parameter 2β. The reader is encouraged to construct diagrams connecting the above cited distributions.

A distinguishing feature amongst RGTL distributions, compared with distributions (6) and (9), is the existence of additional pdf forms with a positive density value at its lower bound (see Figures 3B-3H) allowing representation of uncertain phenomena with such a property. Another feature of RGTL distribution (indicating a lesser flexibility within the same family) is that the pdf's of a GTL distributions and its reflections possess different functional forms, whereas the reflection of a TSP pdf as well as a beta pdf belong to the same functional family.

3. Properties of Standard RGTL distributions

We shall provide some properties of the Standard RGTL (SRGTL) distributions (setting a = 0 and b = 1 in (1) and (2) with the cdf

$$F(x|\,\alpha,\beta) = 1-(1-x)^{\beta}\{\alpha-(\alpha-1)(1-x)\}^{\beta} \tag{9}$$

and the pdf

$$f(x|\,\alpha,\beta) = \beta(1-x)^{\beta-1} \times \{\alpha-(\alpha-1)(1-x)\}^{\beta-1}\{\alpha-2(\alpha-1)(1-x)\} \tag{10}$$

where $0 < \alpha \le 2$ and $\beta > 0$ .Results may be extended to the general forms of (1) and (2) by means of a simple linear transformation.

Limiting Distributions

It immediately follows from (9) that the pdf (10) converges to a degenerate distribution with a probability mass of 1 at a (b) when $\beta \to \infty$ ($\beta \downarrow 0$) regardless of the value of α.

Stochastic Dominance Properties

Note that for $\beta = 1$ (9) simplifies to a slope distribution with the cdf

$$F(x|\,\alpha,\beta=1) = 1-\{\alpha(1-x)-(\alpha-1)(1-x)^2\} \tag{11}$$

which is stochastically decreasing in α, i.e.,

$$\alpha_1 \le \alpha_2,\ x \in (0,1) \Rightarrow F(x|\,\alpha_1,\beta=1) \ge F(x|\,\alpha_2,\beta=1) \tag{12}$$

Let now $\beta_1 \ge \beta_2 > 0$. From (12) it follows that for all $x \in (0,1)$ and for any β_1

$$1-\{1-F(x|\,\alpha_1,\beta=1)\}^{\beta_1} \ge 1-\{1-F(x|\,\alpha_2,\beta=1)\}^{\beta_1} \tag{13}$$

From the fact that the function z^a is a decreasing function in a for $z \in (0,1)$ it follows from $\beta_1 \ge \beta_2 > 0$ that

$$1-\{1-F(x|\,\alpha_2,\beta=1)\}^{\beta_1} \ge 1-\{1-F(x|\,\alpha_2,\beta=1)\}^{\beta_2} \tag{14}$$

However, simple algebra shows that

$$F(x|\,\alpha,\beta) = 1-\{1-F(x|\,\alpha,\beta=1)\}^{\beta} \tag{15}$$

where $F(x|\,\alpha,\beta)$, $F(x|\,\alpha,\beta=1)$ are given by (9) and (11), respectively, which together with (13) and (14) implies

$$\alpha_1 \le \alpha_2,\ \beta_1 \ge \beta_2,\ x \in (0,1) \Rightarrow F(x|\,\alpha_1,\beta_1) \ge F(x|\,\alpha_2,\beta_2) \tag{16}$$

Hence, RGTL distributions are stochastically increasing in α and stochastically decreasing in β. This seems to be an interesting property shedding an additional light on the meaning of the parameters α and β in (9) and (10), especially in applications. Note that, relation (16) could be verbally expressed as connecting the generating cdf $F(x|\,\alpha,\beta=1)$ with the generated one, i.e. $F(x|\,\alpha,\beta)$.

Mode Analysis

As it was already mentioned for $\beta=1$ and $\alpha=1$ the pdf (10) simplifies to a uniform $[0,1]$ density. For $\alpha=1$, $\beta\neq 1$ the pdf (10) becomes a RP distribution (cf. (8)) with a finite mode at 0 with value $\beta>1$ and an infinite mode at 1 for $\beta<1$. Taking the derivative of (10) with respect to x we have

$$\frac{df(x|\,\alpha,\beta)}{dx}=C(x|\,\alpha,\beta)\,f(x|\,\alpha,\beta) \tag{17}$$

where the multiplier

$$C(x|\,\alpha,\beta)= \\ (\alpha-1)\frac{2}{\alpha-2(\alpha-1)(1-x)}-(\beta-1)\frac{\{\alpha-2(\alpha-1)(1-x)\}}{(1-x)\{\alpha-(\alpha-1)(1-x)\}} \tag{18}$$

is a linear function in β. From the relations

$$\begin{aligned} &f(x|\,a,b,\alpha,\beta)>0\\ &\{\alpha-2(\alpha-1)(1-x)\}>0\\ &\{\alpha-(\alpha-1)(1-x)\}>0 \end{aligned} \tag{19}$$

for $\alpha\in[0,2]$ and $\beta>1$ it follows from (17) and (18) that the following four additional cases should be considered: Case 1 : $0<\alpha<1$, $\beta\geq 1$; Case 2 : $1<\alpha\leq 2$, $\beta\leq 1$; Case 3 : $1<\alpha\leq 2$, $\beta>1$; Case 4 : $0<\alpha<1$, $\beta<1$

Case 1 : $0<\alpha<1$, $\beta\geq 1$: see figures 3E and 3F :

From (17), (18) and (19) it follows that the SRGTL pdf (10) is strictly decreasing on $[0,1]$ and hence possesses a mode at 0 with the value $\beta(2-\alpha)$ (cf. (3)). For example, setting $\alpha=0.5$ and $\beta=2$ (as in Figure 3E) yields a mode at 0 with value 3. Setting $\alpha=0.5$ and $\beta=1$ (as in Figure 3F) yields a mode at 0 with value 1.5.

Case 2 : $1<\alpha\leq 2$, $\beta\leq 1$: See Figure 3D:

From (17), (18) and (19) it follows that the SRGTL pdf (10) is strictly increasing on $[0,1]$.

From (4) it follows that the pdf (10) has an infinite mode at 1for $\beta<1$ and a finite mode at 1 for $\beta=1$. Setting $\alpha=1.5$ and $\beta=1$ (as in Figure 3D) yields a finite mode at 1 with value 1.5.

Case 3 : $1<\alpha\leq 2$, $\beta>1$: See Figures 3A, 3B and 3C:

This seems to be the most interesting case. From (17), (18) and (19) it follows that the SRGTL pdf (12) may possess a mode in $(0,1)$. Defining y = 1 – x and setting the derivative (17) to zero yields the following quadratic equation in y

$$2\ (\alpha-1)^2\,y^2 - 2\alpha(\alpha-1)\,y + \frac{\alpha^2(\beta-1)}{2\beta-1} = 0 \tag{20}$$

(The left hand side of (20) is a parabolic function in y). Noting that the symmetry axis of the parabola associated with the l.h.s. of (20) has the value

$$\frac{\alpha}{2(\alpha-1)} \tag{21}$$

which is strictly greater than 1 for $\alpha > 1$, and that $y = 1 - x \in [\,0,1\,]$ $\Leftrightarrow$ $x \in [\,0,1\,]$, it follows that out of the two possible solutions of (20) only the solution

$$y^* = \frac{\alpha}{2(\alpha-1)}\left\{1 - \sqrt{\frac{1}{2\beta-1}}\right\} \tag{22}$$

can yield a mode $x^* \in (\,0,1\,)$. Moreover, from $1 < \alpha \leq 2$, $\beta > 1$ it follows that $y^* > 0$. Also, from (22) we have that $y^* \to \dfrac{\alpha}{2(\alpha-1)} > 1$ for $1 < \alpha \leq 2$ when $\beta \to \infty$. Hence, from (22) we conclude that the mode $x^* = 1 - y^*$ is

$$x^* = \text{Max}\left[0, \frac{1}{2(\alpha-1)}\left\{\alpha\left(1 + \sqrt{\frac{1}{2\beta-1}}\right) - 2\right\}\right] \tag{23}$$

Setting $\alpha = 1.5$ and $\beta = 2$ (as in Figure 3C) yields $x^* = \text{Max}\left[0, -\tfrac{1}{2} + \tfrac{1}{2}\sqrt{3}\right] \approx 0.366$. Setting $\alpha = 1.5$ and $\beta = 6$ (as in Figure 3B) yields $x^* = \text{Max}\left[0, -\tfrac{1}{2} + \tfrac{3}{22}\sqrt{11}\right] = 0$ and hence a mode is located at the lower bound 0 with value $\beta(\alpha-2) = 3$ (cf. (3) with a = 0, b= 1). Utilizing (23) it follows that a Standard Reflected Topp and Leone distribution ($\alpha = 2$) has a mode at

$$\sqrt{\frac{1}{2\beta-1}}$$

for $\beta \geq 1$. Setting $\beta = 3$ (as in Figure 3A) yields a mode at $\dfrac{1}{5}\sqrt{5} \approx 0.447$

Case 4 : $0 < \alpha < 1$, $\beta < 1$: See Figures 3G and 3H:

Similarly to Case 2 it follows that the pdf (10) has an infinite mode at 1for $0 < \alpha < 1$, $\beta < 1$. However, from (17), (18) and (19) it follows that the pdf (10) may also have an anti-mode $x^* \in (0,1)$ (resulting in a U-shaped form) in this case. The formula for the anti-mode is also given by (23) provided $\beta > \frac{1}{2}$. For example, setting $\alpha = 0.5$, $\beta = 0.75$ (as in Figure 3G) yields $x^* = \text{Max}\left[0, \frac{3}{2} - \frac{1}{2}\sqrt{2}\right]$ and hence an anti-mode at approximately 0.793. For $\beta \le \frac{1}{2}$ (as in Figure 3H) the anti-mode of an RGTL distribution occurs at $x^* = 0$, with value $\beta(2-\alpha)$ (cf. (3) with a = 0 , b = 1).

Failure Rate

The failure rate function $r(t) = f(t)/\{1-F(t)\}$ for an SRGTL density follows from (9) and (10) to be

$$D(\alpha, x)\frac{\beta}{1-x} \tag{24}$$

where

$$D(\alpha, x) = \frac{\alpha - 2(\alpha-1)(1-x)}{\alpha - (\alpha-1)(1-x)} \tag{25}$$

and it is straightforward to check that $\beta/(1-x)$ is the failure rate of a standard reflected power (SRP) distribution ((10) with $\alpha = 1$). From (24) it follows that $D(\alpha, x)$ may be interpreted as the relative increase (or decrease) in the failure rate of an SRGTL distribution as compared to a SRP distribution. Taking the derivative of (25) with respect to x yields

$$\frac{\partial D(\alpha, x)}{\partial x} = \frac{\alpha(1-\alpha)}{\{\alpha - (\alpha-1)(1-x)\}^2} \tag{26}$$

Hence, $D(1, x) = 1$ for all $x \in [0,1]$ and it follows from (26) that $D(\alpha, x) < 1$ (>1) for all $x \in [0,1]$ when $1 < \alpha \le 2$ ($0 \le \alpha < 1$). Thus, α may be interpreted as a failure deceleration parameter (relative to the reflected standard power distribution) when $1 < \alpha \le 2$ and a failure acceleration parameter when $0 \le \alpha < 1$. On the other hand, (24) shows that β is a failure acceleration parameter for all $\beta > 0$.

Cumulative Moments

Due to the functional form of the cdf (9) calculations of cumulative moments

$$M_k = \int_0^1 x^k (1-F(x))dx \tag{27}$$

for SRGTL distributions have a slight advantage over that of central moments about the mean. The mean μ_1' and the central moments about the mean μ_2 (variance), μ_3 (skewness) and μ_4 (kurtosis) are connected with the cumulative moments M_k, k = 1,…,4, via

$$\begin{aligned} \mu_1' &= M_0 \\ \mu_2 &= 2M_1 - {M_0}^2 \\ \mu_3 &= 3M_2 - 6M_1M_0 + 2{M_0}^3 \\ \mu_4 &= 4M_3 - 12M_2M_0 + 12M_1{M_0}^2 - 3{M_0}^4 \end{aligned} \tag{28}$$

(see, e.g., Stuart and Ord [10]). The cumulative moments M_k for SRGTL distributions follow from (9) and (27) to be

$$\begin{aligned} &\int_0^1 x^k (1-x)^{\beta} \{\alpha - (\alpha-1)(1-x)\}^{\beta}\, dx = \\ &= \sum_{i=0}^{k} \binom{k}{i} (-1)^i \alpha^{\beta} \int_0^1 x^{\beta+i} \left\{ 1 - \frac{(\alpha-1)x}{\alpha} \right\}^{\beta} dx \end{aligned} \tag{29}$$

For $\alpha = 1$, expression (29) simplifies to that of the cumulative moments of an SRP distribution (cf. (10) with $\alpha = 1$). For $\alpha \in (1,2]$, the cumulative moments can be expressed utilizing the incomplete Beta function

$$B(x \mid \alpha, \beta) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b) \int_0^x p^{a-1} (1-p)^{b-1}\, dp} \tag{30}$$

as

$$M_k = \sum_{i=0}^{k} \binom{k}{i} (-1)^i \alpha^{\beta} \left\{ \frac{\alpha}{\alpha-1} \right\}^{\beta+i+1} \left\{ \frac{B\left(\frac{\alpha-1}{\alpha} \,\middle|\, \beta+i+1, \beta+1 \right)}{B^{-1}(\beta+i+1, \beta+1)} \right\} \tag{31}$$

where $B(a,b) = \dfrac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$ is the Beta function. Numerical routines for evaluating the incomplete Beta function (30) are well known for a long time and are provided in standard PC software such as e.g. Microsoft Excel. However, for $\alpha \in (0,1)$ expression (29) cannot be further simplified and one has to resort to numerical integration. For the moments of the original Topp and Leone [1] distribution (cf. (5)) the cumulative moments were derived by Nadarajah and Kotz [2]. For $\alpha \in (1,2]$, we have for the cumulative moments M_0, M_1, M_2 and M_3

$$
\begin{aligned}
\mathrm{M}_0 &= \alpha^{\beta}\left\{\frac{\alpha}{\alpha-1}\right\}^{\beta+1}\left\{\frac{\mathrm{B}\left(\frac{\alpha-1}{\alpha}\middle|\beta+1,\beta+1\right)}{\mathrm{B}^{-1}(\beta+1,\beta+1)}\right\}\\
\mathrm{M}_1 &= \mathrm{M}_0-\alpha^{\beta}\left\{\frac{\alpha}{\alpha-1}\right\}^{\beta+2}\left\{\frac{\mathrm{B}\left(\frac{\alpha-1}{\alpha}\middle|\beta+2,\beta+1\right)}{\mathrm{B}^{-1}(\beta+2,\beta+1)}\right\}\\
\mathrm{M}_2 &= -\mathrm{M}_0+2\mathrm{M}_1+\alpha^{\beta}\left\{\frac{\alpha}{\alpha-1}\right\}^{\beta+3}\left\{\frac{\mathrm{B}\left(\frac{\alpha-1}{\alpha}\middle|\beta+3,\beta+1\right)}{\mathrm{B}^{-1}(\beta+3,\beta+1)}\right\}\\
\mathrm{M}_3 &= \mathrm{M}_0-3\mathrm{M}_1+3\mathrm{M}_2-\alpha^{\beta}\left\{\frac{\alpha}{\alpha-1}\right\}^{\beta+4}\left\{\frac{\mathrm{B}\left(\frac{\alpha-1}{\alpha}\middle|\beta+4,\beta+1\right)}{\mathrm{B}^{-1}(\beta+4,\beta+1)}\right\}
\end{aligned}
\tag{32}
$$

Substituting $\alpha = 2$, in (32) yields the mean $\mu_1' = \mathrm{M}_0 = 4^{\beta}\,\mathrm{B}(\beta+1,\beta+1)$ of a Standard Reflected Topp and Leone (SRTL) distribution and hence $1-4^{\beta}\,\mathrm{B}(\beta+1,\beta+1)$ is the mean of a Standard Topp and Leone (STL) distribution on $(0,1)$ (see, Nadarajah and Kotz [2]).

Inverse Cumulative Distribution Function

Utilizing the inverse cdf technique random samples from RGTL distributions may straightforwardly be generated. From (9) we derive that $\left\{1-\mathrm{F}^{-1}(\mathrm{z}\,|\,\alpha,\beta)\right\}$, $\mathrm{z}\in[0,1]$ is one of the roots of the quadratic equation in y

$$(\alpha-1)\,\mathrm{y}^2-\alpha\,\mathrm{y}+\sqrt[\beta]{1-\mathrm{z}}=0 \tag{33}$$

Noting that (similarly to equation (20)) the symmetry axis associated with the l.h.s. of the quadratic (33) has a value (21) which is strictly larger than 1 for $1<\alpha\le 2$, it follows that out of the two solutions of (33) only the solution

$$\frac{\alpha-\sqrt{\alpha^2-4(\alpha-1)\sqrt[\beta]{1-\mathrm{z}}}}{2(\alpha-1)}$$

can yield $\left\{1-\mathrm{F}^{-1}(\mathrm{z}\,|\,\alpha,\beta)\right\}\in[0,1]$. Analogously, it follows that for $0\le\alpha<1$ only the solution

$$\frac{\alpha+\sqrt{\alpha^2-4(\alpha-1)\sqrt[\beta]{1-z}}}{2(\alpha-1)}$$

can result in $\{1-F^{-1}(z\,|\,\alpha,\beta)\}\in[0,1]$. Hence, we have

$$F^{-1}(z\,|\,\alpha,\beta)=\begin{cases}1-\dfrac{\alpha-\sqrt{\alpha^2-4(\alpha-1)\sqrt[\beta]{1-z}}}{2(\alpha-1)} & 1<\alpha\le 2\\ 1-\sqrt[\beta]{1-z} & \alpha=1\\ 1-\dfrac{\alpha+\sqrt{\alpha^2-4(\alpha-1)\sqrt[\beta]{1-z}}}{2(\alpha-1)} & 0\le\alpha<1\end{cases} \tag{34}$$

where the case $\alpha=1$ follows from the cdf of a standard reflected power (SRP) distribution ($\alpha=1$ in (9)).

4. Maximum likelihood estimation

Below we shall discuss an approximate MLE procedure for a total of N observations grouped in m intervals $[x_{i-1},x_i]$ with n_i observations each and interval mean values $\overline{x}_i$, where $x_0\equiv 0$, $x_m\equiv 1$ and

$$N=\sum_{i=1}^{m} n_i$$

The data described above may be summarized in an m-vector x whose elements are the interval mean values and an m-vector n containing the number of observations in each interval. The approximate MLE procedure below may easily be modified to a non-approximate MLE procedure utilizing order statistics, but here our approach is tailored to the format of the income distribution data to be presented in Table 1. The approximate MLE procedure will assume that the probability mass is concentrated at the interval mean $\overline{x}_i$ of the intervals $[x_{i-1},x_i]$. Utilizing (10) we have the likelihood $L(\alpha,\beta\,|\,x,n)$ to be proportional to

$$\beta^N\prod_{i=1}^{m}\left[\left\{\alpha y_i-(\alpha-1)y_i{}^2\right\}^{\beta-1}\left\{\alpha-2(\alpha-1)y_i\right\}\right]^{n_i} \tag{35}$$

where

$$y_i=1-\overline{x}_i \tag{36}$$

Instead of maximizing $L(\alpha,\beta\,|\,x,n)$ we may equivalently maximize the log-likelihood. Taking the logarithm of (35) and calculating the derivative with respect to β we obtain

$$\frac{N}{\beta}+\sum_{i=1}^{m}n_i\,\mathrm{Ln}\left\{\alpha y_i-(\alpha-1)y_i{}^2\right\} \tag{37}$$

It follows from (37) that

$$\hat{\beta} = N\left[\sum_{i=1}^{m} n_i \,\mathrm{Ln}\left\{\frac{1}{\alpha y_i - (\alpha-1)y_i^{\,2}}\right\}\right]^{-1} \tag{38}$$

is the unique MLE of β given a particular value of α. Taking the logarithm of (35) and calculating the derivative with respect to α, one obtains

$$(\beta-1)\sum_{i=1}^{m} \frac{n_i(1-y_i)}{\alpha-(\alpha-1)y_i} + \sum_{i=1}^{m} \frac{n_i(1-2y_i)}{\alpha-2(\alpha-1)y_i} \tag{39}$$

Substituting (38) into (39) (utilizing $\hat{\beta}$ instead of β and expressing $\hat{\beta}$ in terms of α) the following function $\Psi(\alpha)$ is derived:

$$\Psi(\alpha) = \left[\frac{N}{\sum_{i=1}^{m} n_i \,\mathrm{Ln}\left\{\frac{1}{\alpha y_i - (\alpha-1)y_i^{\,2}}\right\}} - 1\right]\sum_{i=1}^{m} \frac{n_i(1-y_i)}{\alpha-(\alpha-1)y_i} + \sum_{i=1}^{m} \frac{n_i(1-2y_i)}{\alpha-2(\alpha-1)y_i} \tag{40}$$

where y_i is given by (36) and the function is defined on a bounded range of $0 \le \alpha \le 2$ .The MLE $\hat{\alpha}$ follows as one of the roots of the equation $\Psi(\alpha) = 0$ or as one of the boundary values $\alpha = 0$ or $\alpha = 2$. The bounded domain of $\Psi(\alpha)$ allows for straightforward plotting of the function in standard spreadsheet software such as Microsoft Excel and subsequent determination of an approximate solution of the MLE $\hat{\alpha}$. Using the root finding algorithm Goalseek, available in Microsoft Excel, and the approximate solution of $\hat{\alpha}$ allows us to calculate $\hat{\alpha}$ up to a desired level of accuracy. Finally, substitution of $\hat{\alpha}$ in to (38) yields the MLE $\hat{\beta}$. The MLE procedure above will be demonstrated in the next section using U.S. 2001 income data.

5. Fitting 2001 U.S. income distribution data

In 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. [8] an illuminating and comprehensive analysis of the African-American and Caucasian (Non-Hispanic) wealth gap was presented based on a longitudinal survey of approximately over 6000 households over the period 1968–1992. The authors argue that a parametric estimation of the wealth-earning relationship by race is not an appropriate approach. Their main objection is that the wealth-earning relationship is non-linear with an unknown functional form which is difficult to parameterize and parametric estimation may thus likely yield inaccurate estimates. The authors also provide an extensive and up-to-date bibliography up to and including 2001. Barsky et al. [8] note that the racial wealth gap far exceeds the racial income gap at the higher wealth ranges, suggesting that the racial wealth gap is too large to be explained by income gap alone. On the other hand, they 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 earnings account for all of the racial wealth difference in the first quartile of the wealth distribution. Interested readers are also referred to Couch and Daly [9] and O'Neill et al. [11] who study the related topic of the racial wage gap in the U.S.

Our approach to this problem is somewhat different. We attempt to use the distribution developed in the previous sections to fit the more recent household income data in the U.S. for the year 2001 (Source: U.S. Census Bureau, Current Population Survey, March 2002) classified according to the Caucasian (Non-Hispanic), African-American and Hispanic populations and draw some tentative conclusions about the racial income gap based on this data. Parametric estimation of income data is quite common for almost 100 years and a wide variety of distributions have been proposed (see, Kleiber and Kotz [12] for an extensive bibliography). RGTL distributions (which are not discussed in Kleiber and Kotz [12]) allow for a strictly positive density value at its lower bound, which is observed in a non parametric kernel density estimate of the 1989 income data (see Figure 2 of Barsky et al. [8] p. 668). The new distribution we are proposing turns out to be appropriate for the U.S. 2001 household income data, especially for that of the African-American sub population. We emphasize that the main purpose of the numerical analysis below is to illustrate the fitting attributes of the RGTL distribution and properties of its parameters. The numerical analysis herein in no way yields a conclusive answer to the problem of racial income gaps (nor that of racial wealth and racial wage gaps) while providing indications of the current state of affairs and further study is in order.

Table 1 below contains income distribution data for households in the year 2001 for the different ethnic groups: Caucasian (Non-Hispanic), African-American and Hispanic throughout the U.S.A. The MLE procedure above will be used to fit RGTL distributions for incomes of these three groups. Only the data up to $250,000 will be used in Table 1 since the U.S. Census Bureau data does not provide the maximum observed income in their statistics. Of the total number of U.S. households surveyed 98.58%, 99.44%, 99.65% have in 2001 an income less than $250,000 for the Caucasian (Non-Hispanic), Hispanic and African-American ethnic groups, respectively.

Figure 4 displays a graph of the function $\varsigma(\alpha)$ (cf. (40)) for the income data of Caucasian (Non-Hispanic) Americans presented in Table 1. From Figure 4 we observe an approximate root of the equation $\varsigma(\alpha) = 0$ to be the value $\alpha^* \approx 1.70$. Since, $\varsigma(\alpha) > 0$ (< 0) for $0 \le \alpha < \alpha^*$ ($\alpha^* \le \alpha \le 2$) it follows that $\hat{\alpha} = \alpha^*$ is the unique MLE of (35) for α . Using Goalseek (a standard root finding algorithm in Microsoft Excel) with a accuracy of 1 . 10^{-6}, utilizing the approximate solution 1.70 we obtain $\hat{\alpha} = \alpha^* = 1.679$. The unique MLE $\hat{\beta} = 6.767$ follows from substituting $\hat{\alpha} = 1.679$ in (38). Figure 5 below plots both the empirical and their fitted RGTL counterparts (cf. (1) and (2)) with a = \$0, b = \$250,000 , $\alpha = 1.679$ and $\beta = 6.767$. Differences between the empirical cdf and fitted cdf can be observed in Figure 5A. The Kolmogorov-Smirnov Statistic D, which is the maximum observed difference between the empirical and fitted cdf's (see, e.g., DeGroot [13]), in Figure 5A equals 8.60%.

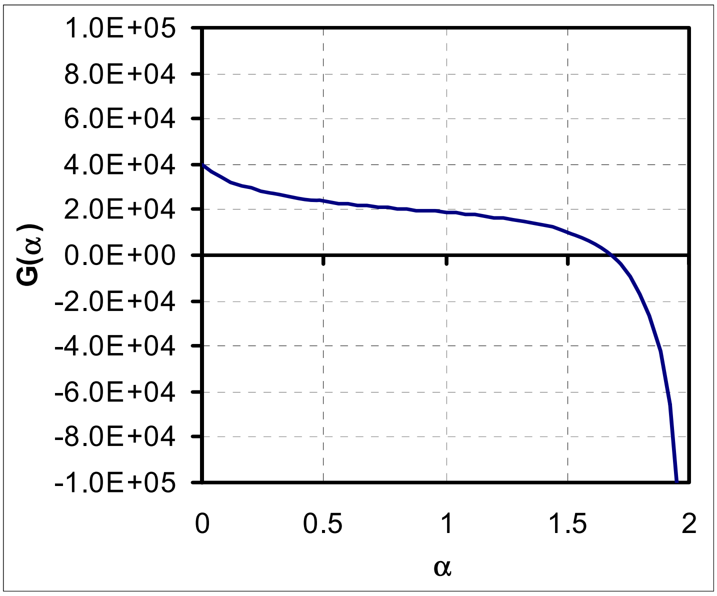

Figure 4. A graph of the function $\varsigma(\alpha)$ (cf. (40)) for the income data of Caucasian (Non-Hispanic) Americans presented in Table 1

Table 1. U.S. income distribution for households in year 2001 (Source: U.S. Census Bureau, Current Population Survey, March 2002. Numbers in thousands, households as of March of the following year)

	Caucasian (Non-Hispanic)		African American		Hispanic	
Income of Household	Number	Mean Income	Number	Mean Income	Number	Mean Income
Under $2,500..............	1,443	$168	520	$439	273	$426
$2,500 to $4,999...........	773	$3,808	388	$3,842	137	$3,767
$5,000 to $7,499...........	2,141	$6,450	698	$6,359	304	$6,387
$7,500 to $9,999...........	2,561	$8,749	756	$8,658	404	$8,663
$10,000 to $12,499.........	3,142	$11,220	621	$11,173	458	$11,214
$12,500 to $14,999.........	2,946	$13,615	543	$13,672	411	$13,659
$15,000 to $17,499.........	3,167	$16,091	660	$16,089	553	$15,993
$17,500 to $19,999.........	2,803	$18,660	479	$18,655	418	$18,579
$20,000 to $22,499.........	3,099	$21,082	610	$21,094	490	$21,005
$22,500 to $24,999.........	2,697	$23,706	447	$23,682	373	$23,691
$25,000 to $27,499.........	3,055	$26,064	570	$26,061	477	$26,011
$27,500 to $29,999.........	2,446	$28,673	464	$28,544	330	$28,617
$30,000 to $32,499.........	3,277	$31,059	492	$31,040	479	$30,998
$32,500 to $34,999.........	2,330	$33,679	375	$33,655	335	$33,601
$35,000 to $37,499.........	2,950	$36,045	437	$35,944	412	$36,082
$37,500 to $39,999.........	2,114	$38,713	310	$38,626	249	$38,641
$40,000 to $42,499.........	2,846	$41,052	434	$41,004	424	$40,938
$42,500 to $44,999.........	1,924	$43,679	260	$43,693	231	$43,668
$45,000 to $47,499.........	2,236	$46,058	289	$45,908	291	$46,044
$47,500 to $49,999.........	1,966	$48,709	256	$48,655	205	$48,607
$50,000 to $52,499.........	2,403	$51,042	350	$50,924	247	$51,021
$52,500 to $54,999.........	1,736	$53,679	210	$53,553	153	$53,725
$55,000 to $57,499.........	2,014	$56,127	249	$55,972	224	$55,992
$57,500 to $59,999.........	1,528	$58,650	177	$58,680	177	$58,764
$60,000 to $62,499.........	2,047	$61,053	248	$60,979	219	$61,106
$62,500 to $64,999.........	1,417	$63,719	162	$63,761	141	$63,801
$65,000 to $67,499.........	1,710	$66,048	175	$65,990	157	$66,018
$67,500 to $69,999.........	1,325	$68,677	150	$68,705	124	$68,734
$70,000 to $72,499.........	1,622	$71,067	190	$71,090	159	$71,112
$72,500 to $74,999.........	1,248	$73,707	142	$73,589	128	$73,711
$75,000 to $77,499.........	1,608	$75,981	133	$75,974	132	$75,860
$77,500 to $79,999.........	1,073	$78,662	100	$78,693	72	$78,726
$80,000 to $82,499.........	1,380	$81,051	100	$80,950	125	$80,976
$82,500 to $84,999.........	993	$83,688	90	$83,584	90	$83,708
$85,000 to $87,499.........	1,144	$86,057	103	$85,984	76	$85,830
$87,500 to $89,999.........	803	$88,696	86	$88,754	55	$88,636
$90,000 to $92,499.........	985	$91,051	78	$91,103	83	$90,997
$92,500 to $94,999.........	701	$93,658	83	$93,666	41	$93,579
$95,000 to $97,499.........	915	$96,071	71	$95,901	65	$95,999
$97,500 to $99,999.........	712	$98,682	65	$98,639	48	$98,811
$100,000 to $149,999.......	8,374	$119,083	554	$117,549	515	$119,016
$150,000 to $199,999.......	2,689	$169,312	115	$172,222	113	$164,692
$200,000 to $249,999.......	993	$219,285	29	$218,672	43	$221,737
$250,000 and above.........	1,345	$462,675	46	$433,097	59	$474,843
Total	**90,682**	**$60,512**	**13,315**	**$39,248**	**10,499**	**$44,383**

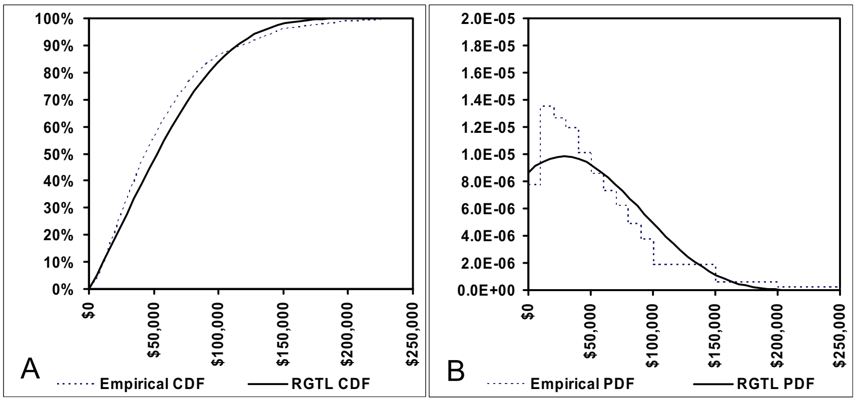

Figure 5. Empirical and an MLE fitted RGTL distribution ($\hat{\alpha} = 1.643$ and $\hat{\beta} = 6.179$) of the Caucasian (Non-Hispanic) income data in Table 1; A: CDF; B: PDF

Hence, with 43 degrees of freedom (Table 1 has 43 rows up to \$250,000) the Kolmogorov-Smirnov test accepts the fitted RGTL distribution at the 10% ($D_{0.10} \approx 0.182$), 5% ($D_{0.05} \approx 0.203$) as well as 1% ($D_{0.01} \approx 0.243$)levels, respectively. Table 2 provides the unique MLE estimators for $\hat{\alpha}$ and $\hat{\beta}$ (obtained using the procedure described in Section 4) for the Caucasian (Non-Hispanic), African-American and Hispanic income data presented in Table 1. Figure 6A (Figure 6B) plots the empirical and fitted RGTL pdf with MLE $\hat{\alpha} = 1.613$, $\hat{\beta} = 10.629$ ($\hat{\alpha} = 1.685$ and $\hat{\beta} = 10.306$) for the African-American (Hispanic) income data as presented in Table 1. The Kolmogorov-Smirnov Statistic D for the African-American (Hispanic) income data equals 6.01% (8.09%) which is smaller than that of the Caucasian (Non-Hispanic) income data (indicating a better fit). Hence the Kolmogorov-Smirnov test accepts both MLE fitted RGTL distributions in Figure 6A and 6B at the 10%, 5% and 1% levels, respectively.

Table 2. Maximum Likelihood Estimators for the parameters $\hat{\alpha}$ and $\hat{\beta}$ of RGTL distributions for the income data in Table 1 up to \$250,000

	$\hat{\alpha}$	$\hat{\beta}$
Caucasian (Non-Hispanic)	1.679	6.767
Hispanic	1.685	10.306
African-American	1.613	10.629

Table 3 contains the (standardized) cumulative moments $M_0 = \mu_1'$, M_1, M_2, M_3 and the central moments μ_2, μ_3 and μ_4 calculated utilizing (32) and (28) . Note that there is a strict ordering column-wise for all the values in Table 3 in the order: Caucasian American (non-Hispanic), Hispanic, African-American. From Table 3 we can calculate values for the mean and standard deviation utilizing the transformation Y = \$250,000 X. In a similar manner, the median and mode of the MLE RGTL distributions can be evaluated utilizing the parameter values in Table 2, (34) and (23). In addition, we may utilize Table 3 to calculate the coefficient of skewness β_1 and coefficient of kurtosis β_2 given by

$$\beta_1 = \frac{\mu_3^{\,2}}{\mu_2^{\,3}}\ ,\ \beta_2 = \frac{\mu_4}{\mu_2^{\,2}}$$

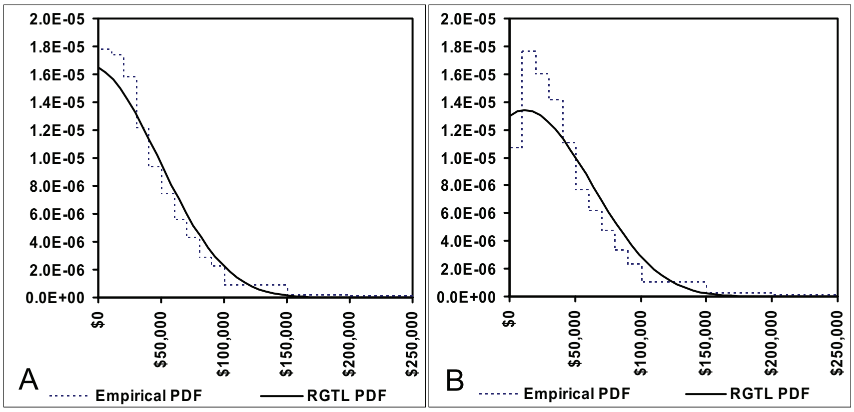

Figure 6. Empirical and MLE fitted RGTL pdf's for the income data in Table 1; A: African-American ($\hat{\alpha} = 1.613$ and $\hat{\beta} = 10.629$); B: Hispanic ($\hat{\alpha} = 1.685$ and $\hat{\beta} = 10.306$)

These estimated statistics are provided in Table 4 for the three subpopulations under consideration.

Table 3. Cumulative Moments M_k and Central Moments μ_{k+1} of the MLE fitted RGTL distributions for the income data in Table 1 up to \$250,000 calculated utilizing (32) and (28), k = 1,…,3

	$M_0 = \mu_1'$	M_1	M_2	M_3	μ_2	μ_3	μ_4
Caucasian (Non-Hispanic)	2.34e-1	3.97e-2	1.11e-2	3.80e-3	2.47e-2	2.54e-3	1.75e-3
Hispanic	1.77e-1	2.38e-2	5.26e-3	1.51e-3	1.60e-2	1.66e-3	8.40e-3
African-American	1.59e-1	1.98e-2	4.17e-3	1.14e-3	1.44e-2	1.60e-3	7.31e-3

Table 4. Statistics associated with the MLE fitted RGTL distributions for the income data in Table 1 up to $250,000

	Mean	Median	Mode	St. Dev	β_1	β_2
Caucasian (Non-Hispanic)	$58393	$52534	$28306	$39326	0.424	2.858
Hispanic	$44316	$38606	$11851	$31710	0.660	3.248
African-American	$39786	$33599	$0	$30002	0.858	3.522

A similar ordering as observed in Table 3 can be observed throughout Table 4. Note that the difference in the point estimates in Table 4 between the Caucasian (Non-Hispanic) population and the African-American Population is approximately $18607 or more and those associated with the Hispanic population $13928 or more. The latter observation is amplified somewhat in Table 4 by the fact that the fitted mean income for the Caucasian (Non-Hispanic) population overestimates the empirical mean (of income up to $250,000)by $3936 whereas the fitted mean income for the African-American (Hispanic) population is overestimated by only $1898 ($2357). Perhaps the most notable difference is the modal income value of $0 for the MLE fitted RGTL distribution for the African-American population while the modal income value for the Caucasian (Non-Hispanic) and Hispanic population have a value substantially larger than zero (and the mode for the Caucasian (Non-Hispanic) population is more than twice that of Hispanics). A similar observation can be made by comparing the RGTL distributions in Figure 5B, 6A and 6B. Finally, from Table 2 and (3) we may evaluate the density values at the lower bound, i.e. $f(0 \mid 0,\$250{,}000,\hat{\alpha},\hat{\beta})$, presented in Table 5. Hence, in comparison with Americans of Caucasian origin, African-Americans appear to be approximately 1.9 times as likely and Hispanics 1.5 times as likely, in the year 2001, to have negligible income. It is the fact that our MLE fitted RGTL pdf's may take any positive value at its lower bound, that allows us to reach such a conclusion.

Table 5. Density values at the lower bound of the MLE fitted RGTL distributions for the income data in Table 1 up to $250,000

	$f(0 \mid 0,\$250{,}000,\hat{\alpha},\hat{\beta})$
Caucasian (Non-Hispanic)	8.68e-6
Hispanic	1.30e-5
African-American	1.65e-5

The analysis of our investigations presented below seems to be, in our opinion, of some interest and value. In Figure 7A, we utilize the MLE fitted RGTL income distributions by plotting the percentiles of the African-American and Hispanic income distributions against those of the Caucasian (Non-Hispanic) one using (9), (34) and the corresponding MLE values for $\hat{\alpha}$ and $\hat{\beta}$ in Table 2. For example, from Figure 7A, we observe that approximately 70% (65%) of the African-American (Hispanic) population have less income than the median (50%) of the Caucasian (Non-Hispanic) income distribution. Similar comparisons can be made for other percentiles of the Caucasian (Non-Hispanic) income distribution utilizing Figure 7A. For example, 34%(29%) of the African-American (Hispanic) population earn less than what less than 20% of the Caucasian (Non-Hispanic) population earn (i.e. the 20% percentile of the Caucasian (Non-Hispanic) income distribution). Note that the solid curve in Figure 7A involving the African-American (Hispanic) income distribution is located completely above the unit diagonal which implies stochastic dominance of Caucasian (Non-Hispanic) income over that of the African-American (Hispanic) one. The latter can be directly concluded from the MLE values for $\hat{\alpha}$ and $\hat{\beta}$ in Table 2 and (16) for the African-American and Caucasian (Non-Hispanic) comparison but not the Hispanic and Caucasian (Non-Hispanic) comparison. This shows that the implication arrow in (16) cannot in general be reversed.

In a similar manner, Figure 7B utilizes the MLE fitted RGTL income distributions by plotting the percentiles of the African-American and Caucasian (Non-Hispanic) income distributions against those of the Hispanic one. For example, from Figure 7B, we observe that approximately 56% (37%) of the African-American (Caucasian Non-Hispanic) population have less income than the median (50%) of the Hispanic income distribution. We now conclude from Figure 7B that Hispanic income stochastically dominates the African-American one. The latter conclusion also follows immediately from the corresponding MLE values for $\hat{\alpha}$ and $\hat{\beta}$ in Table 2 and (16). However, we can conclude, once again, only by observation in Figure 7B that Hispanic income is stochastically dominated by Caucasian (Non-Hispanic) income (since the line associated with the Caucasian (Non-Hispanic) now happens to be completely below the unit diagonal). This conclusion, as before, cannot be directly obtained from the corresponding MLE values for $\hat{\alpha}$ and $\hat{\beta}$ in Table 2 and (16).

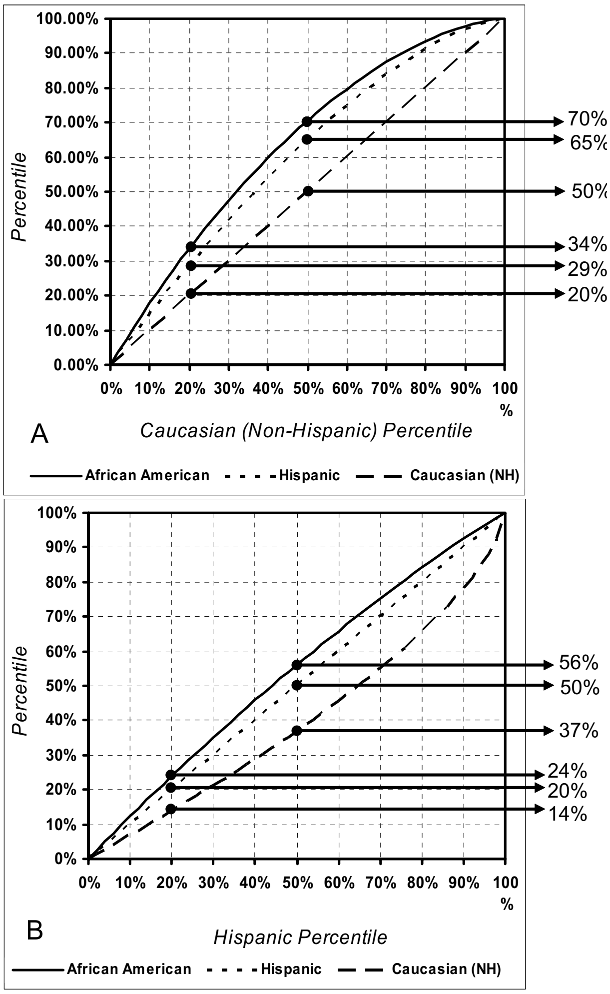

Figure 7. Stochastic Dominance Analysis by Ethnicity for the income data in Table 1 utilizing the MLE fitted RGTL cdf's

Summarizing, Table 2 and (16) alone imply that the chances of a Caucasian (Non-Hispanic) or Hispanic American earning more than a specified amount (anywhere within the range from \$0 to \$250,000) are higher than those for an African-American. In addition, the analysis in Figure 7 allows us to conclude that the chances of a Caucasian (Non-Hispanic) earning more than a specified amount (anywhere within the range from \$0 to \$250,000) are higher than those of a Hispanic. Moreover, Figure 7 and Table 4 demonstrate that although substantial advances have reportedly been made in reducing the income distribution gap amongst these three subpopulations in the U.S. during the last 20 years or so (see, e.g., Couch and Daly [9]), these differences still exist and are quite noticeable.

6. Concluding remarks

We have attempted to construct and investigate a new four-parameter continuous family of distributions on a bounded domain possessing arbitrary strictly positive density values at its lower bound. As an illustration, the new family is applied to fitting the distributions of income of Caucasians (Non-Hispanic), Hispanics and African-Americans in the U.S.A. in the year 2001 based on U.S. Census bureau data. The results seems to be quite satisfactory and allow us to compare the incomes of the above 3 groups in a novel manner which seems to be revealing by shedding additional light on features which are not obvious from a direct examination of the raw data.

Acknowledgments

We are indebted to T.A. Mazzuchi for his helpful comments in the course of developing this paper and to Dr. David Findley (U.S. Bureau of Census) for helping us to obtain recent data.

References

1. Topp, C.W. and Leone, F.C. (1955). A family of J-shaped frequency functions. Journal of the American Statistical Association, 50(269), 209–219.
2. Nadarajah, S. and Kotz, S. (2003). Moments of some J-shaped distributions. Journal of Applied Statistics, 30(3), 311–317.
3. Van Dorp, J.R. and Kotz, S. (2002). The standard two sided power distribution and its properties: With applications in financial engineering. The American Statistician, 56(2), 90–99.
4. Van Dorp, J.R. and Kotz, S. (2002). A novel extension of the triangular distribution and its parameter estimation. Journal of Royal Statistical Society, Series D, The Statistician, 51(1), 63–79.
5. Johnson, N.L. (1949). Systems of frequency curves generated by the methods of translation. Biometrika, 36, 149–176.
6. Weibull W. (1939). A statistical distribution of wide applicability. Journal of Applied Mechanics, 18, 293–297.
7. Van Dorp, J.R. and Kotz, S. (2003). Generalized trapezoidal distributions. Metrika, 58(1), 85–97.

8. Barsky, R., Bound, J., Kerwin, K.C. and Lupton, J.P. (2002). Accounting for the Black-White wealth gap: A nonparametric approach. Journal of the American Statistical Association, 97(459), 663–673.
9. Couch, K. and Daly, M.C. (2000). Black-White inequality in the 1990's: A decade of progress. Working Papers in Applied Economic Theory, No. 2000-07, Federal Reserve Bank of San Francisco.
10. Stuart, A. and Ord, J.K. (1994). Kendall's Advanced Theory of Statistics (Vol. 1, Distribution Theory). New York, Wiley.
11. O'Neill, D., Sweetman, O. and Van de Gaer, D. (2002). Estimating counterfactual densities: An application to Black-White wage differentials in the U.S., Economics Department Working Paper Series, Department of Economics, National University of Ireland – Maynooth.
12. Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences. New York, Wiley.
13. DeGroot, M.H. (1991). Probability and Statistics, 3rd ed. Reading, MA: Addison-Wesley.