Statistics and Applications {ISSN 2454-7395 (online)}
Volume 22, No. 1, 2024 (New Series), pp 321–341
http://www.ssca.org.in/journal

Singh Maddala Dagum Distribution with Application to Income Data

Ashlin Varkey and Haritha N. Haridas
Department of Statistics, Farook College (Autonomous), Kozhikode,
University of Calicut, Kerala, India 673632
Received: 15 September 2023; Revised: 08 October 2023; Accepted: 12 October 2023

Abstract

This article introduces the Singh Maddala Dagum distribution as the sum of the quantile functions of the Singh-Maddala and Dagum distributions. The distributional properties, income inequality measures, and poverty measures of this distribution are derived. Poverty measures such as the poverty gap ratio and the Foster-Greer-Thorbecke measure were converted to quantile forms. The least squares method is used to estimate the parameters of the proposed distribution, and the model is applied to two real datasets.

Key words: Singh-Maddala distribution; Dagum distribution; Quantile function; Income inequality measures; Poverty measures.

AMS Subject Classifications: 33B15, 33B20, 60E05, 62G07, 62P20

1. Introduction

The two equivalent techniques for modeling and analyzing statistical data are by using the distribution functions and quantile functions. The quantile function for a real-valued and continuous random variable X with distribution function $F(x)$ is given as

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}, \; 0 \leq u \leq 1.$$

Even though Galton (1875) first proposed the formal concept of quantiles, the work of Hastings *et al.* (1947) provided a notable advancement in depicting quantile functions to represent distributions. Parzen's (1979) paper and Tukey's (1977) research on exploratory data analysis stimulated the development of the quantile functions as a vital tool in statistical analysis instead of the distribution functions.

The quantile function holds a number of characteristics that the distribution function does not have. In particular, two quantile functions added together and two positive quantile functions multiplied together are again quantile functions. Also, $\frac{1}{Q(1-u)}$ is the quantile function of $\frac{1}{X}$, if $Q(u)$ is the quantile function of X. For a comprehensive review of this concept, one can refer to Nair *et al.* (2013), Gilchrist (2000), Sankaran and Dileep Kumar (2018), and the references therein.

Corresponding Author: Ashlin Varkey
Email: ashlinvarkey@gmail.com

Tarsitano (2004) used a general form of the Tukey lambda family of distributions proposed by Ramberg and Schmeiser (1972), to provide a good start for quantile-based income modeling. However, the model put forth by Tarsitano (2004) is not valid throughout the parametric space. To solve this issue, Haritha *et al.* (2007) utilized the four-parameter generalized lambda distribution proposed by Freimer *et al.* (1988) for income modeling. Later, using the quantile function method, the Zenga measure and other measures of income inequality were examined by Sreelakshmi and Nair (2014).

The objective of this paper is to introduce a new quantile function that is useful for the analysis of income data. Since Singh-Maddala (SM) and Dagum distributions are adaptable and frequently used in income modeling, we propose the Singh Maddala Dagum (SMD) distribution derived from the sum of the quantile functions of the two models.

Singh and Maddala introduced the SM distribution in 1975 and refined it in 1976, has received special attention among income distributions. The SM distribution is a special case of the generalized beta 2 (GB2) distribution and is known as Burr XII or simply Burr distribution. For a detailed study on the SM distribution, one could refer to Kleiber and Kotz (2003),Shahzad and Asghar (2013b), and Kumar (2017). The distribution and quantile functions of the SM distribution are given by

$$G(x) = 1 - \left[1 + \left(\frac{x}{b}\right)^{a}\right]^{-q}, \ \ x > 0, \tag{1}$$

and

$$Q_1(u) = b\left[(1-u)^{-\frac{1}{q}} - 1\right]^{\frac{1}{a}}, \ \ 0 < u < 1, \tag{2}$$

where all three parameters a, b, q are positive.

Dagum distribution proposed by Dagum (1977) is also a special case of GB2 distribution and is known as Burr III distribution. Dagum distribution has numerous applications in the fields of reliability, meteorology, quality control, insurance, business failure data, and income modeling. A detailed discussion of the Dagum distribution can be found in Kleiber and Kotz (2003) and Shahzad and Asghar (2013a). Using the SM and Dagum distributions Saulo *et al.* (2023) proposed parametric quantile regressions. The distribution and quantile functions of the Dagum distribution are given by

$$H(x) = \left[1 + \left(\frac{x}{b}\right)^{-a}\right]^{-p}, \ \ x > 0, \tag{3}$$

and

$$Q_2(u) = b\left[u^{-\frac{1}{p}} - 1\right]^{-\frac{1}{a}}, \ \ 0 < u < 1, \tag{4}$$

where all three parameters a, b, p are positive.

The remaining portion of the article is structured as follows. We define SMD distribution and its basic aspects in Section 2. Section 3 deals with some popular distributions that belong to the proposed class or that result from pertinent transformations on the proposed quantile function. Section 4 covers the distributional properties, such as skewness, kurtosis, L-moments, order statistics, etc. Section 5 discusses the major income inequalities and poverty measures of the proposed class. The inference method and its application to real data are carried out in Section 6. Overall findings from the study are given in the final Section 7.

2. Singh Maddala Dagum (SMD) quantile function

If X and Y are two non-negative random variables with quantile functions $Q_1(u)$ and $Q_2(u)$ respectively. Then

$$Q(u) = Q_1(u) + Q_2(u),$$

is again a quantile function. Likewise, the sum of two quantile density functions results in a quantile density function. Now we define a new quantile function

$$Q(u) = b\left[\left((1-u)^{-\frac{1}{q}} - 1\right)^{\frac{1}{a}} + \left(u^{-\frac{1}{p}} - 1\right)^{-\frac{1}{a}}\right],\ 0 < u < 1,\ a,\ b,\ p,\ q > 0, \tag{5}$$

which is the sum of quantile functions in (2) and (4). The proposed class of distribution is known as SMD distribution and its support is $(0, \infty)$. The quantile density function of the SMD distribution is

$$\begin{aligned} q(u) &= \frac{dQ(u)}{du} \\ &= b\left[\frac{(1-u)^{-\frac{1}{q}-1}\left((1-u)^{-\frac{1}{q}} - 1\right)^{\frac{1}{a}-1}}{aq} + \frac{u^{-\frac{1}{p}-1}\left(u^{-\frac{1}{p}} - 1\right)^{-\frac{1}{a}-1}}{ap}\right]. \end{aligned}$$

The density and distribution functions are not available in closed form for the family of distributions given in (5). However, these can be computed by numerical inversion of the quantile function. In terms of the distribution function, the density function $f(x)$ of the proposed class can be written as

$$f(x) = \frac{1}{b}\left[\frac{apq\ \ F(x)^{\frac{1}{p}+1}(1-F(x))^{\frac{1}{q}+1}}{pF(x)^{\frac{1}{p}+1}[(1-F(x))^{-\frac{1}{q}} - 1]^{\frac{1}{a}-1} + q(1-F(x))^{\frac{1}{q}+1}(F(x)^{-\frac{1}{p}} - 1)^{-\frac{1}{a}-1}}\right]. \tag{6}$$

The density function is plotted for various parameter combinations and is given in Figure 1. For various parameter values, it can be seen that the family includes decreasing, unimodal, positive, and negatively skewed models.

3. Members of the family

We can obtain several popular distributions from the suggested model (5) for various parameter values and by utilizing some transformations given in Gilchrist (2000).
Case 1. $b > 0, q > 0, a = 1$ and $p \to 0$
The quantile function of the suggested class tends to the Lomax distribution and is given as

$$Q(u) = b\left[(1-u)^{-\frac{1}{q}} - 1\right]. \tag{7}$$

Case 2. $b > 0, a > 0, q = 1$ and $p \to 0$
The quantile function of the suggested class tends to the Fisk distribution and is given as

$$Q(u) = b\left[(1-u)^{-1} - 1\right]^{\frac{1}{a}}. \tag{8}$$

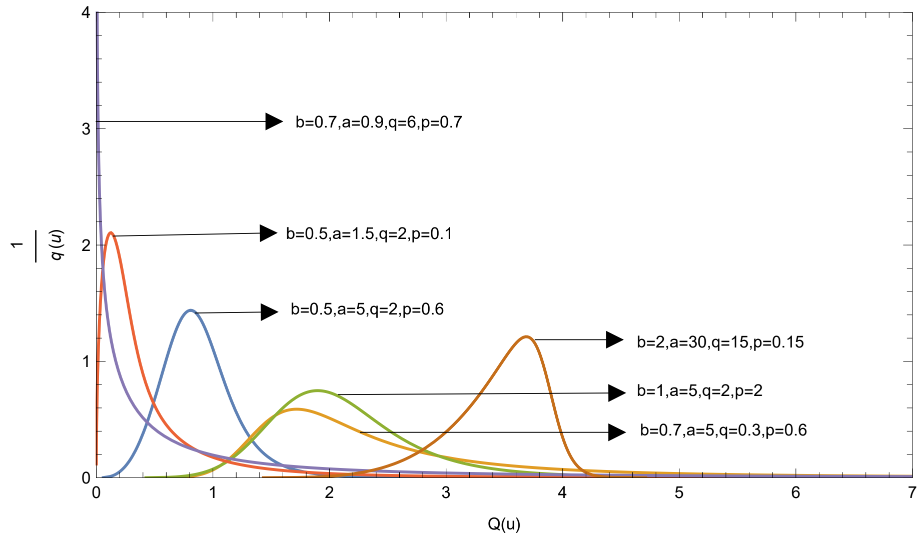

Figure 1: Plots of density function for different values of parameters

Case 3. $b > 0, a = q$ and $p \to 0$
The quantile function of the suggested class tends to the Paralogistic distribution and is given as

$$Q(u) = b\left[(1-u)^{-\frac{1}{a}} - 1\right]^{\frac{1}{a}}. \tag{9}$$

On applying reciprocal transformation on (9), we get the inverse Paralogistic distribution with quantile function

$$Q(u) = \frac{1}{Q(1-u)} = k\left(u^{-\frac{1}{a}} - 1\right)^{-\frac{1}{a}},$$

where $k = \frac{1}{b}$ and a are the parameters. Further details on paralogistic and inverse paralogistic distributions can be found in Klugman *et al.* (2019).

The following theorems give the relationships between the random variables representing the SM, SMD, and Dagum distributions.

Theorem 1: If $V \sim SM(a,b,q)$ then the random variable,
$U = V + b\left\{\left[1 - \left(1 + \left(\frac{V}{b}\right)^a\right)^{-q}\right]^{-\frac{1}{p}} - 1\right\}^{-\frac{1}{a}}$ has $SMD(a,b,p,q)$ distribution.

Proof:

Let S and R represent two random variables with distribution functions $F_S(x)$ and $F_R(x)$ and quantile functions $Q_S(u)$ and $Q_R(u)$ respectively. Assume $Q^*(u) = Q_S(u) + Q_R(u)$, then the random variable that corresponds to the quantile function $Q^*(u)$ is $S + Q_R(F_S(S))$ or $R + Q_S(F_R(R))$ (Sankaran *et al.*, 2016).

Let $V \sim SM(a,b,q)$ and $W \sim Dagum(a,b,p)$; then $V + Q_W(F_V(V))$ has $SMD(a,b,p,q)$ distribution by above result.
We have, $Q_W(u) = b\left(u^{-\frac{1}{p}} - 1\right)^{-\frac{1}{a}}$ and $F_V(V) = 1 - \left[1 + \left(\frac{V}{b}\right)^a\right]^{-q}$

Therefore, $V + Q_W\left(F_V(V)\right) = V + b\left\{\left[1-\left(1+\left(\frac{V}{b}\right)^{a}\right)^{-q}\right]^{-\frac{1}{p}}-1\right\}^{-\frac{1}{a}}$ has $SMD(a,b,p,q)$ distribution. □

Theorem 2: If $W \sim Dagum(a,b,p)$, then the random variable,
$U = W + b\left\{\left[1-\left(1+\left(\frac{W}{b}\right)^{-a}\right)^{-p}\right]^{-\frac{1}{q}}-1\right\}^{\frac{1}{a}}$ has $SMD(a,b,p,q)$ distribution.

Proof: The proof is omitted since it is similar to that of Theorem 1. □

4. Distributional characteristics

The use of quantile functions reduces the effort needed to describe a distribution through its moments. Hence it is common in statistical analysis to use quantile-based measurements of distributional features like location, dispersion, skewness, and kurtosis. These measurements can be used to estimate the model's parameters by matching population characteristics with corresponding sample characteristics.

4.1. Measures of location, spread and shape

The r^{th} order traditional moment is given as

$$E(X^r) = \int_0^1 (Q(u))^r du.$$

In particular, the mean of the SMD distribution is

$$\mu = b\left[\frac{\Gamma\left(1+\frac{1}{a}\right)\Gamma\left(q-\frac{1}{a}\right)}{\Gamma(q)} + \frac{\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(1-\frac{1}{a}\right)}{\Gamma(p)}\right].$$

For the model given in (5), the median (M) is

$$\begin{aligned} M &= Q(0.5) \\ &= b\left[\left(2^{\frac{1}{q}}-1\right)^{\frac{1}{a}} + \left(2^{\frac{1}{p}}-1\right)^{-\frac{1}{a}}\right]. \end{aligned} \tag{10}$$

The interquartile range (IQR) is

$$\begin{aligned} IQR &= Q(0.75) - Q(0.25) \\ &= b\Bigg\{\left[(0.25)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}} - \left[(0.75)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}} \\ &\quad + \left[(0.75)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}} - \left[(0.25)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}\Bigg\}. \end{aligned} \tag{11}$$

Galton's skewness (S) and Moors kurtosis (T) measures are given in (12) and (13) respectively.

$$\begin{aligned} S &= \frac{Q(0.25)+Q(0.75)-2M}{IQR} \\ &= \frac{S_1+S_2}{\left[(0.25)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}-\left[(0.75)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}+\left[(0.75)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}-\left[(0.25)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}}, \end{aligned} \tag{12}$$

where $S_1=\left[(0.25)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}+\left[(0.75)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}-2\left[2^{\frac{1}{q}}-1\right]^{\frac{1}{a}}$,
and $S_2=\left[(0.25)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}+\left[(0.75)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}-2\left[2^{\frac{1}{p}}-1\right]^{-\frac{1}{a}}$.

$$\begin{aligned} T &= \frac{Q(0.875)-Q(0.625)+Q(0.375)-Q(0.125)}{IQR} \\ &= \frac{T_1+T_2}{\left[(0.25)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}-\left[(0.75)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}+\left[(0.75)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}-\left[(0.25)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}}, \end{aligned} \tag{13}$$

where $T_1=\left[(0.125)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}-\left[(0.375)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}+\left[(0.625)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}-\left[(0.875)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}$,
and $T_2=\left[(0.875)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}-\left[(0.625)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}+\left[(0.375)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}-\left[(0.125)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}$.

4.2. L-moments

The L-moments are alternatives to the classical moments and are the expected values of linear functions of order statistics. The work on order statistics by Sillitto (1969) and Greenwood *et al.* (1979) laid the foundation for L-moments, but Hosking (1990) developed a comprehensive theory on L-moments. These moments are resistant to outliers and typically have reduced sample variances. Like classical moments, L-moments can be used to identify distributions, summarise measures of probability distributions, and fit models to data. The r^{th} L-moment is represented as

$$L_r=\int_0^1\sum_{k=0}^{r-1}(-1)^{r-1-k}\binom{r-1}{k}\binom{r-1+k}{k}u^kQ(u)du.$$

The first four L-moments of SMD distributions are

$$L_1=b\left[A_1O_1+A_2R_1\right],$$

$$L_2=b\left[A_1\left(O_1-O_2\right)-A_2\left(R_1-R_2\right)\right],$$

$$L_3=b\left[A_1\left(O_1-3O_2+2O_3\right)+A_2\left(R_1-3R_2+2R_3\right)\right],$$

$$L_4=b\left[A_1\left(O_1-6O_2+10O_3-5O_4\right)-A_2\left(R_1-6R_2+10R_3-5R_4\right)\right],$$

where $A_1 = \Gamma\left(1+\frac{1}{a}\right)$, $A_2 = \Gamma\left(1-\frac{1}{a}\right)$, $O_i = \frac{\Gamma\left(iq-\frac{1}{a}\right)}{\Gamma(iq)}$, $R_i = \frac{\Gamma\left(ip+\frac{1}{a}\right)}{\Gamma(ip)}$ and $i = 1, 2, 3, 4$.
The L-coefficient of variation (τ_2), which is an alternative to the coefficient of variation based on traditional moments is

$$\begin{aligned}\tau_2 &= \frac{L2}{L1}\\ &= \frac{A_1\left(O_1-O_2\right)-A_2\left(R_1-R_2\right)}{A_1O_1+A_2R_1}.\end{aligned} \tag{14}$$

The L-coefficient of skewness (τ_3) and L-coefficient kurtosis (τ_4) of the SMD distribution, is given in (15) and (16).

$$\begin{aligned}\tau_3 &= \frac{L3}{L2}\\ &= \frac{A_1\left(O_1-3O_2+2O_3\right)+A_2\left(R_1-3R_2+2R_3\right)}{A_1\left(O_1-O_2\right)-A_2\left(R_1-R_2\right)}.\end{aligned} \tag{15}$$

$$\begin{aligned}\tau_4 &= \frac{L4}{L2}\\ &= \frac{A_1\left(O_1-6O_2+10O_3-5O_4\right)-A_2\left(R_1-6R_2+10R_3-5R_4\right)}{A_1\left(O_1-O_2\right)-A_2\left(R_1-R_2\right)}.\end{aligned} \tag{16}$$

The plots of L-coefficients of skewness (τ_3) and kurtosis (τ_4) for different parameter values are given in Figures 2, 3 and 4. In Figure 2, the curve of τ_3 decreases with a for fixed value of q and p but the curve of τ_4 decreases with a for fixed value of q and p, when $p > 1$. In Figure 3, the curves of τ_3 and τ_4 increase with p for fixed values of a and q when $q \geq 1$. The curves of τ_3 and τ_4 for fixed values of a and p and for varying q are given in Figure 4.

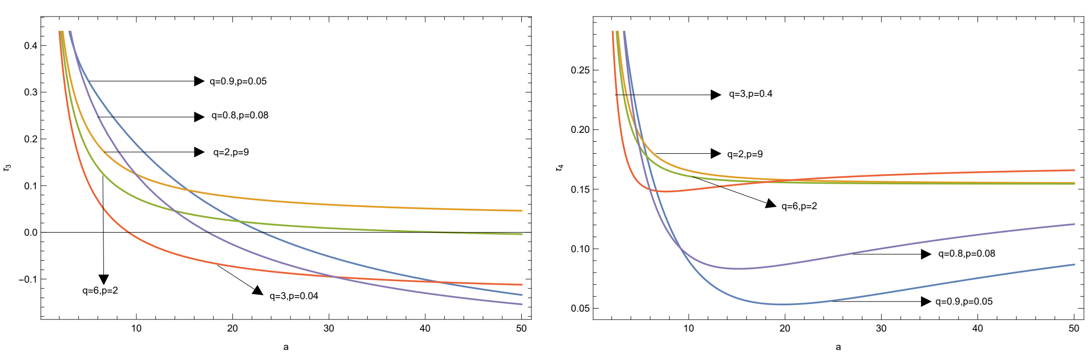

Figure 2: Plot of L-coefficients of skewness and kurtosis for particular values of q and p as a function of the parameter a

4.3. Order statistics

In a random sample of size n, let $X_{r:n}$ represent the r^{th} order statistic. Then, $X_{r:n}$ has density function $f_r(x)$ and is given as

$$f_r(x) = \frac{1}{\beta(r, n-r+1)} f(x)F(x)^{r-1}(1-F(x))^{n-r}.$$

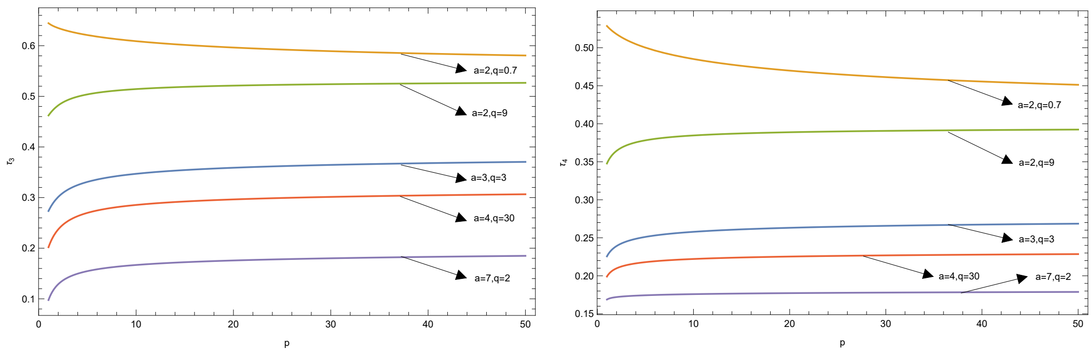

Figure 3: Plot of L-coefficients of skewness and kurtosis for particular values of a and q as a function of the parameter p

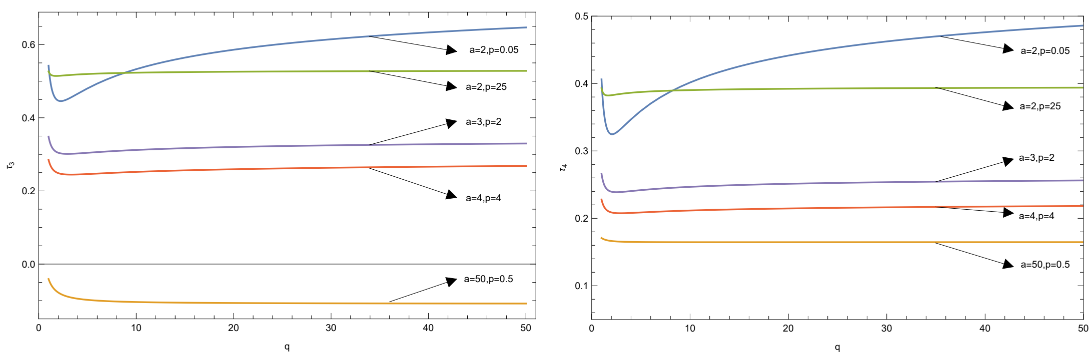

Figure 4: Plot of L-coefficients of skewness and kurtosis for particular values of a and p as a function of the parameter q

From (6) we get

$$f_r(x) = \frac{apq}{b\beta(r,n-r+1)} \frac{F(x)^{r+\frac{1}{p}}(1-F(x))^{n+\frac{1}{q}+1-r}}{pF(x)^{\frac{1}{p}+1}\left[(1-F(x))^{-\frac{1}{q}}-1\right]^{\frac{1}{a}-1}+q(1-F(x))^{\frac{1}{q}+1}\left(F(x)^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}-1}}.$$

Thus

$$E(X_{r:n}) = \frac{apq}{b\beta(r,n-r+1)} \times \int_0^\infty \frac{xF(x)^{r+\frac{1}{p}}(1-F(x))^{n+\frac{1}{q}+1-r}}{pF(x)^{\frac{1}{p}+1}\left[(1-F(x))^{-\frac{1}{q}}-1\right]^{\frac{1}{a}-1}+q(1-F(x))^{\frac{1}{q}+1}\left(F(x)^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}-1}}dx.$$

In quantile terms, the above expression can be written as

$$E(X_{r:n}) = \frac{apq}{b\beta(r,n-r+1)} \int_0^1 \frac{Q(u)\,u^{r+\frac{1}{p}}(1-u)^{n+\frac{1}{q}+1-r}}{p\,u^{\frac{1}{p}+1}\left((1-u)^{-\frac{1}{q}}-1\right)^{\frac{1}{a}-1}+q(1-u)^{\frac{1}{q}+1}\left(u^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}-1}}du.$$

For SMD distribution the first-order statistic $X_{1:n}$ has a quantile function

$$Q_{(1)}(u) = Q\left[1-(1-u)^{\frac{1}{n}}\right]$$
$$= b\left\{\left[(1-u)^{-\frac{1}{nq}}-1\right]^{\frac{1}{a}} + \left[\left(1-(1-u)^{\frac{1}{n}}\right)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}\right\}, \quad (17)$$

and the n^{th} order statistic $X_{n:n}$ has the quantile function

$$Q_{(n)}(u) = Q\left(u^{\frac{1}{n}}\right)$$
$$= b\left\{\left[\left(1-u^{\frac{1}{n}}\right)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}} + \left[u^{-\frac{1}{np}}-1\right]^{-\frac{1}{a}}\right\}. \quad (18)$$

5. Income inequality and poverty measures

In statistical and economics literature, the study of income inequality and poverty measures are always popular and favorite subjects. A measure of income inequality is intended to give an index, that can reduce the differences in income that exist among the members of a group, whereas a poverty measure evaluates the severity of poverty experienced by those whose income is below a pre-determined poverty level.

5.1. Income inequality measures

The Lorenz curve proposed by Lorenz (1905) is a flexible tool for reporting and graphically depicting income inequality. When the income is arranged in increasing order of magnitude, the points $(u, L(u))$ define a Lorenz curve, where u denotes the cumulative frequency of income receiving units and $L(u)$ denotes the cumulative frequency of income. Gastwirth (1971) gave a general definition of Lorenz curve as

$$L(u) = \frac{1}{\mu}\int_0^u Q(p)dp,$$

where $\mu = \int_0^1 Q(p)dp$. For SMD distribution the Lorenz curve is

$$L(u) = \frac{q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}{q\beta\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}, \quad (19)$$

where $\beta_*(.,.)$, is an incomplete beta function.

The Gini index is a well known income inequality proposed by Gini (1914) and is defined as two times the area between the Lorenz curve and the egalitarian line. The Gini index for the class of distributions in (5) is

$$G = 1-2\int_0^1 L(u)du$$
$$= 1-2\left[\frac{q\beta\left(1+\frac{1}{a}, 2q-\frac{1}{a}\right) + p\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right) - p\beta\left(2p+\frac{1}{a}, 1-\frac{1}{a}\right)}{q\beta\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}\right]. \quad (20)$$

Pietra (1932) developed the Pietra index which measures the maximal vertical distance between the Lorenz curve and the line of equality. The Pietra index and relative mean deviation in quantile terms are

$$P = \frac{\vartheta_1}{2\mu},$$
$$\tau_2 = \frac{\vartheta_1}{\mu},$$

where $\vartheta_1 = \int_0^1 |Q(u) - Q(u_0)|du$ and $\mu = Q(u_0)$ for some $0 < u_0 < 1$. Further, by solving for u in the equation $\mu = Q(u)$, u_0 can be obtained, and μ represents the mean of the distribution.
Now, the Pietra index of the SMD distribution is given as

$$P = \frac{u_0 Q\left(u_0\right) - b\left[q\beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)\right]}{\mu}. \tag{21}$$

The Bonferroni curve proposed by Bonferroni (1930) is used to quantify the variability in income distribution. For an absolutely continuous and non-negative random variable, the Bonferroni curve in quantile terms is given as

$$\begin{aligned} B_F(u) &= \frac{L(u)}{u} \\ &= \frac{1}{u\mu}\int_0^u Q(p)dp. \end{aligned}$$

For SMD distribution the Bonferroni curve is

$$B_F(u) = \frac{q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}{u\left[q\beta\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)\right]}. \tag{22}$$

A more realistic curve was introduced by Zenga (2007) based on the conditional expectation of the concerned distribution. The Zenga curve in quantile terms is

$$Z(u) = 1 - \frac{(1-u)}{u}\frac{\int_0^u Q(p)dp}{\int_u^1 Q(p)dp}.$$

For SMD distribution the Zenga curve is given as

$$Z\left(u\right) = \frac{q\,\mathfrak{z}_1 + p\,\mathfrak{z}_2}{q\,\mathfrak{z}_3 + p\,\mathfrak{z}_4}, \tag{23}$$

where

$$\begin{aligned} \mathfrak{z}_1 &= \left[\beta\left(1+\frac{1}{a}, q-\frac{1}{a}\right) - u^{-1}\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right)\right], \\ \mathfrak{z}_2 &= \left[\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right) - u^{-1}\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)\right], \\ \mathfrak{z}_3 &= \left[\beta\left(1+\frac{1}{a}, q-\frac{1}{a}\right) - \beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right)\right], \\ \mathfrak{z}_4 &= \left[\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right) - \beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)\right]. \end{aligned}$$

The Lorenz, Bonferroni, and Zenga curves of SMD distribution are given in Figure 5, 6, and 7 respectively.

The Frigyes measures developed by Éltetö and Frigyes (1968) have clear economic interpretations and are given as

$$\varphi = \frac{m}{m_1},\ \psi = \frac{m_2}{m_1},\ \omega = \frac{m_2}{m},$$

where $m = E(X)$, $m_1 = E(X|X < m)$, and $m_2 = E(X|X \geqslant m)$. The measure ψ can be considered as an inequality measure for the complete income distribution, whereas φ and ω denote the inequalities of the two respective portions of the distribution below and above the mean.

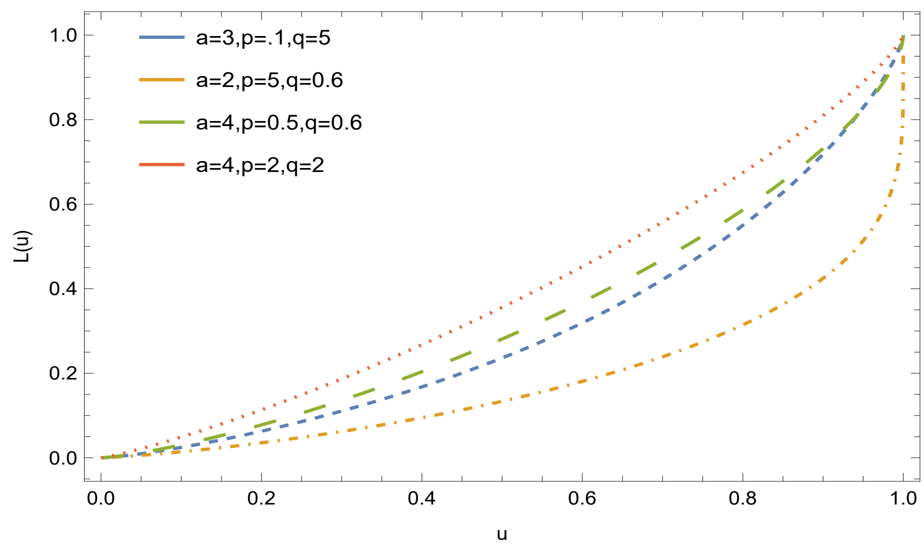

Figure 5: Graph of SMD Lorenz curve

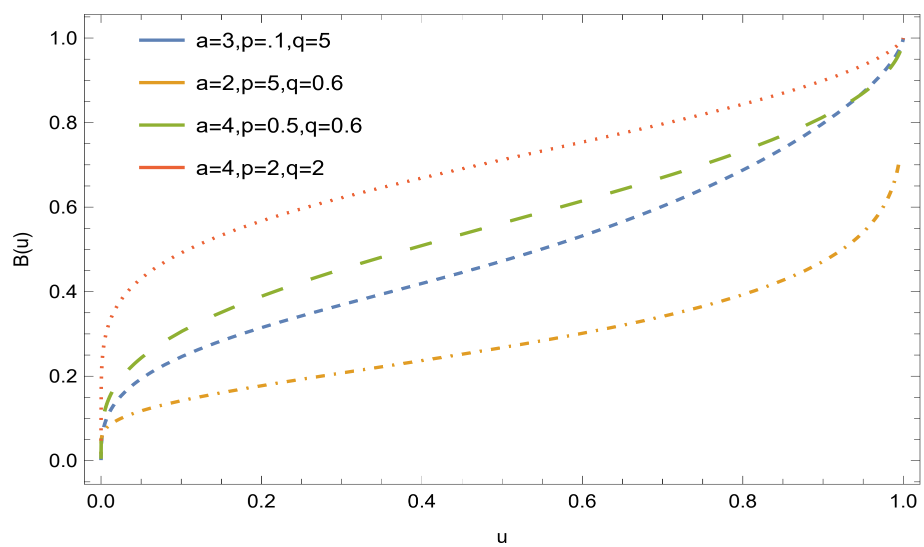

Figure 6: Graph of SMD Bonferroni curve

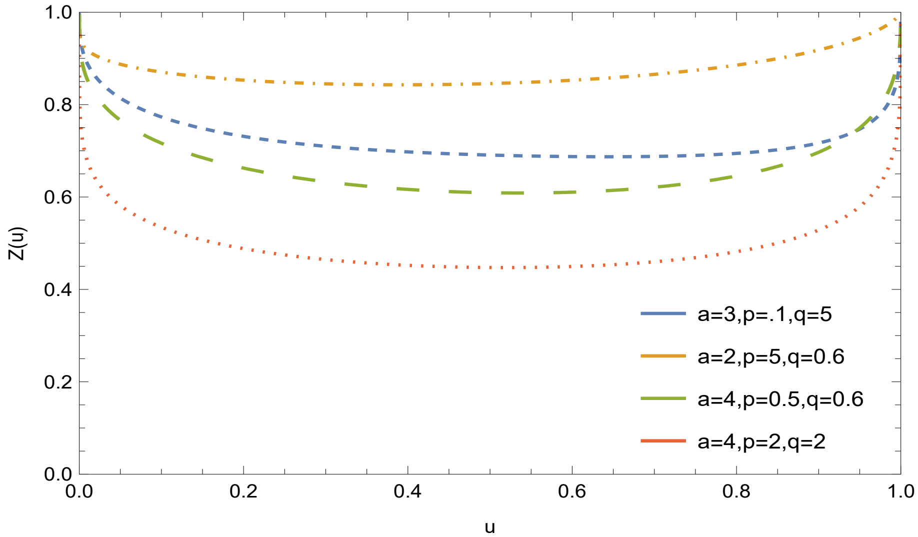

Figure 7: Graph of SMD Zenga curve

In quantile terms, these measures are given as

$$\varphi = \frac{u_0 Q(u_0)}{\int_0^{u_0} Q(u)du},$$
$$\psi = \frac{u_0}{1-u_0}\frac{\int_{u_0}^{1} Q(u)du}{\int_0^{u_0} Q(u)du},$$
$$\omega = \frac{\int_{u_0}^{1} Q(u)du}{(1-u_0)Q(u_0)}.$$

For SMD distribution these measures are

$$\varphi = \frac{u_0\left[\left((1-u_0)^{-\frac{1}{q}}-1\right)^{\frac{1}{a}} + \left(u_0^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}\right]}{q\beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}, \tag{24}$$

$$\psi = \frac{u_0}{(1-u_0)}\left[\frac{q\beta\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}{q\beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)} - 1\right], \tag{25}$$

$$\omega = \frac{q\,\mathfrak{w}_1 + p\,\mathfrak{w}_2}{(1-u_0)\left[\left((1-u_0)^{-\frac{1}{q}}-1\right)^{\frac{1}{a}} + \left(u_0^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}\right]}. \tag{26}$$

where

$$\mathfrak{w}_1 = \left[\beta\left(1+\frac{1}{a}, q-\frac{1}{a}\right) - \beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right)\right],$$
$$\mathfrak{w}_2 = \left[\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right) - \beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)\right].$$

5.2. Poverty measures

Measures of poverty are primarily used to track socioeconomic development and set goals for success or failure. Most of the poverty measurements can be stated as the average deprivation faced by the poor. If the function $D(z, y)$ describes the level of deprivation experienced by an individual whose income y is less than the poverty line z.
Hence

$$\begin{aligned} P &= E_y \left[D\left(z, y\right) I\left(y < z\right) \right] \\ &= \int_0^z D\left(z, y\right) f(y) dy, \end{aligned} \tag{27}$$

where $I\left(y < z\right)$ represents an indicator function which takes value 1 when $y < z$ and 0 otherwise, and $f(y)$ represents probability density function. For a detailed reading on poverty measures, one can refer to Kakwani (1980) and Morduch (2008). Chotikapanich *et al.* (2013) derived poverty measures from generalized beta distribution and examined how poverty has changed in south and southeast Asian nations.

The headcount ratio is the most basic and widely used measure of poverty, it represents the proportion of the population who are poor and is denoted by H. By definition

$$H = \frac{N_p}{N},$$

where N_p and N denotes the number of poor and total population respectively. That is, the head-count ratio ignores the severity of the deprivation experienced by the poor.

A number of alternatives to the head-count ratio have been proposed in order to establish a measure that takes into account both the proportion of poor as well as the intensity of poverty among those who are characterized as poor. The poverty gap ratio calculates the amount of money by which each person falls below the poverty line. It can be obtained from (27), by setting $D\left(z, y\right) = \left(\frac{z-y}{z}\right)$. Thus

$$\begin{aligned} PG &= \int_0^z D\left(z, y\right) f(y) dy \\ &= \int_0^z \left(\frac{z-y}{z}\right) f(y) dy. \end{aligned} \tag{28}$$

Using the transformation, $F(z) = u$ and $F(y) = p$, where $0 < u < 1$ and $0 < p < 1$ in (28), we get the poverty gap ratio in quantile form and is given as

$$PG = \int_0^u \left(\frac{Q(u) - Q(p)}{Q(u)}\right) dp. \tag{29}$$

The poverty gap ratio defined here is also known as the income gap ratio of the poor in Haritha *et al.* (2007). The poverty gap ratio can be written in terms of reversed mean residual quantile function as follows

$$PG = \frac{uR(u)}{Q(u)},$$

where $R(u) = u^{-1} \int_0^u (Q(u) - Q(p))\, dp$. For, SMD distribution the poverty gap ratio is

$$PG = u - \frac{q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}{\left((1-u)^{-\frac{1}{q}}-1\right)^{\frac{1}{a}} + \left(u^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}}. \tag{30}$$

The Foster-Greer-Thorbecke (FGT) measure proposed by Foster *et al.* (1984) generalizes the poverty gap ratio. Here, $D(z, y) = \left(\frac{z-y}{z}\right)^{\alpha}$ and the measure is

$$FGT(\alpha) = \int_0^z \left(\frac{z-y}{z}\right)^{\alpha} f(y)dy,$$

where $\alpha \geq 1$ is the inequality aversion parameter. The lower tail of the income distribution receives more emphasis as the value of α increases. When $\alpha = 1$, the FGT measure becomes equivalent to the poverty gap ratio. The quantile version of the FGT measure can be obtained by using the same transformation in the poverty gap ratio and is given in (31). Moreover, it does not have a closed form expression for the SMD distribution.

$$FGT(\alpha) = \int_0^u \left(\frac{Q(u) - Q(p)}{Q(u)}\right)^{\alpha} dp. \tag{31}$$

Watts (1968) introduced the first distribution-sensitive poverty index called Watt's index. This index satisfies the focus, monotonicity, and transfer axioms of poverty and in quantile terms, it is given as

$$W = \int_0^u \ln\left(\frac{Q(u)}{Q(p)}\right) dp. \tag{32}$$

Kakwani (1999) has proposed a measure that is closely related to the Watts index and is given by, $K^* = 1 - e^{-W}$. For SMD distribution these indices do not have simple algebraic expressions.

Sen (1976) put forward a measure that attempted to incorporate the effects of the number of poor, the severity of their poverty, and poverty distribution within the group. In quantile terms, it is

$$S = u\left(\frac{u\,\Delta\rho_1(u) + \rho_2(u)}{u\,\Delta\rho_1(u) + \rho_1(u)}\right),$$

where $\rho_1(u) = \frac{1}{u}\int_0^u Q(p)dp$, $\rho_2(u) = \frac{1}{u^2}\int_0^u (2p-u)\,Q(p)dp$, and $\Delta\,\rho_1$ denotes derivative of ρ_1 with respect to u. For SMD distribution $\rho_1(u)$ and $\rho_2(u)$ are given as

$$\rho_1(u) = \frac{b}{u}\left[q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)\right],$$

$$\rho_2(u) = \frac{b}{u^2}\Bigg[(2-u)q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) - 2q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, 2q-\frac{1}{a}\right)$$
$$+ 2p\beta_{u^{\frac{1}{p}}}\left(2p+\frac{1}{a}, 1-\frac{1}{a}\right) - up\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)\Bigg].$$

The Gini index for the poor has the quantile form

$$\eta(u) = 1 - \frac{2}{\rho_1(u)} \int_0^u Q(p)\left(\frac{u-p}{u^2}\right) dp$$
$$= \frac{\rho_2(u)}{\rho_1(u)}. \tag{33}$$

Now, for SMD distribution the above index can be written as

$$\eta(u) = \frac{2}{u} \times \frac{A}{B} - 1, \tag{34}$$

where

$$A = q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) - q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, 2q-\frac{1}{a}\right) + p\beta_{u^{\frac{1}{p}}}\left(2p+\frac{1}{a}, 1-\frac{1}{a}\right),$$
$$B = q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) + p\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right).$$

6. Inference and applications

Here, we estimate the parameters of the family of distributions (5) and use real data sets to assess the model's effectiveness and applications.

The parameters of the distribution in a quantile setup can be estimated using a variety of methods. The L-moments method, the percentile approach, the minimum absolute deviation method, the least squares method, and the maximum likelihood method are frequently used techniques. We employ the method of least squares to estimate the parameters of the model (5). In order to estimate the generalized Tukey lambda distribution, Öztürk and Dale (1985) utilized this estimation method. Hankin and Lee (2006) also used this method to estimate the parameters of the Davies distribution. The least square estimation is illustrated as follows.

Let $X_{(i)}$ denote the i^{th} order statistic from a random sample of size n from SMD distribution, and $u_{(i)}$ be the i^{th} order statistic of the associated uniformly distributed random variable, $u = F(X)$. In the ideal situation, both the random variables $X_{(i)}$ and $Q(u_{(i)}, \hat{\delta})$ have the same distribution, where $\hat{\delta}$ is the estimator of the model's parameter vector. In this estimation technique, we estimate $\delta = (a, b, p, q)$ that minimizes $\zeta(\delta)$

$$\zeta(\delta) = \sum_{i=1}^{n} \left(X_{(i)} - Q(u_{(i)}, \delta)\right)^2.$$

6.1. Real data analysis

The applicability of the model (5) can be demonstrated with the aid of two real income datasets. The first data is taken from `https://www.bea.gov`, which deals with the per capita personal income of 46 counties in South Carolina State, 2018. Using midyear population estimates from the Census Bureau, per capita personal income was calculated. We use the least squares method discussed above to estimate the parameters. The estimate is based on the parameter value that minimizes the residual sum of squares and is obtained as

$$\hat{a} = 8.2443,\ \hat{b} = 10618.9,\ \hat{q} = 2.83288,\ \text{and}\ \hat{p} = 1856.39.$$

The Q-Q plot and the chi-square test are the two goodness-of-fit criteria used here to evaluate how well the model fits the data. The Q-Q plot given in Figure 8, shows that the fit is satisfactory. We conducted the chi-square goodness-of-fit test and obtained the test statistic value as 6.89418 with p-value 0.648136. Hence, the proposed model (5) fits the given dataset reasonably well. Since the quantile functions of the SM and Dagum distributions are added to obtain our model, we fitted the above data to these distributions, and the results are given in Table 1. Figure 9 illustrates the histogram of the data along with the density functions for the SM, SMD, and Dagum distributions. It is clear from the figure that the SMD distribution fits the dataset more accurately than the other two models.

Table 1: Parameter estimates, chi-square statistic, and p-value of SM and Dagum distributions for dataset 1

Distribution	Parameter estimates	Chi-square statistic	p-value
SM	a = 24.0223		
	b = 33492.8	8.66856	0.468414
	q = 0.31186		
Dagum	a = 10.477		
	b = 36651.4	7.74593	0.559939
	p = 1.27739		

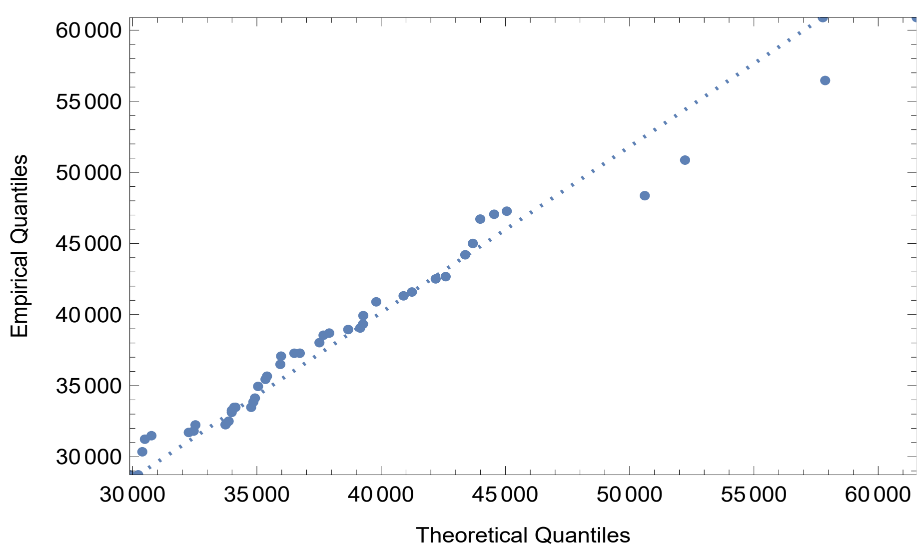

Figure 8: Q-Q plot for the per capita personal income of counties in South Carolina State in 2018

The second dataset is also taken from `https://www.bea.gov`, which deals with the per capita personal income of 120 counties in Kentucky State, 2020. The method of least squares is employed to estimate the parameters and is obtained as

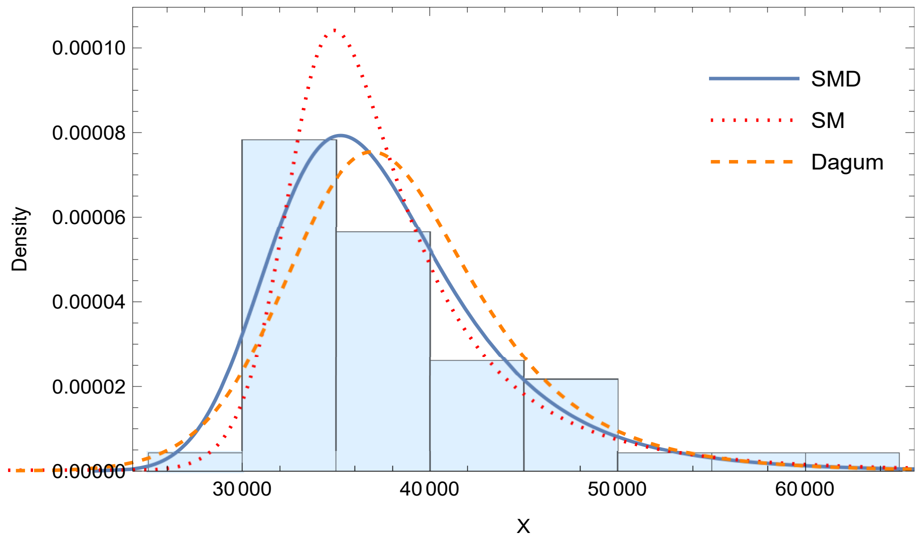

Figure 9: The densities of the SM, SMD, and Dagum distributions for the per capita personal income of counties in South Carolina State in 2018

$$\hat{a} = 9.46169, \ \hat{b} = 17513.9, \ \hat{q} = 3.20767, \text{ and } \hat{p} = 27.4554.$$

Two goodness-of-fit methods are used to evaluate how well the model fits the data. The first one is the Q-Q plot in Figure 10, which shows that the suggested model is appropriate for the given data set. In addition, we perform the chi-square goodness-of-fit test and get test statistic value 4.35791 with p-value 0.986754. This indicates the fit of SMD distribution for the given data. The SM and Dagum distributions are also fitted to income data of Kentucky State, and the results are given in Table 2. The histogram of the data and the density functions for the SM, SMD, and Dagum distributions are shown in Figure 11. From the figures and the chi-square values the SMD model appears to be better than SM and Dagum distributions.

Table 2: Parameter estimates, chi-square statistic, and p-value of SM and Dagum distributions for dataset 2

Distribution	Parameter estimates	Chi-square statistic	p-value
SM	a = 16.1799		
	b = 39576.5	7.20778	0.891129
	q = 0.640004		
Dagum	a = 9.86976		
	b = 36637.5	4.49484	0.984699
	p = 2.34804		

7. Conclusion

In this article, we propose the quantile function known as SMD distribution by adding the quantile functions of the SM and Dagum distributions. Several popular distributions are members of the proposed class of distributions. We studied the distributional properties, the major income inequality, and the poverty measures of the proposed class. We also derived the quantile version of poverty measures, such as the poverty gap ratio and the Foster-Greer-Thorbecke measure. The estimation of the parameters of the model was done using the method of least squares. The proposed class of distribution was used for the analysis of two real income data and it gives a better fit than SM and Dagum distributions.

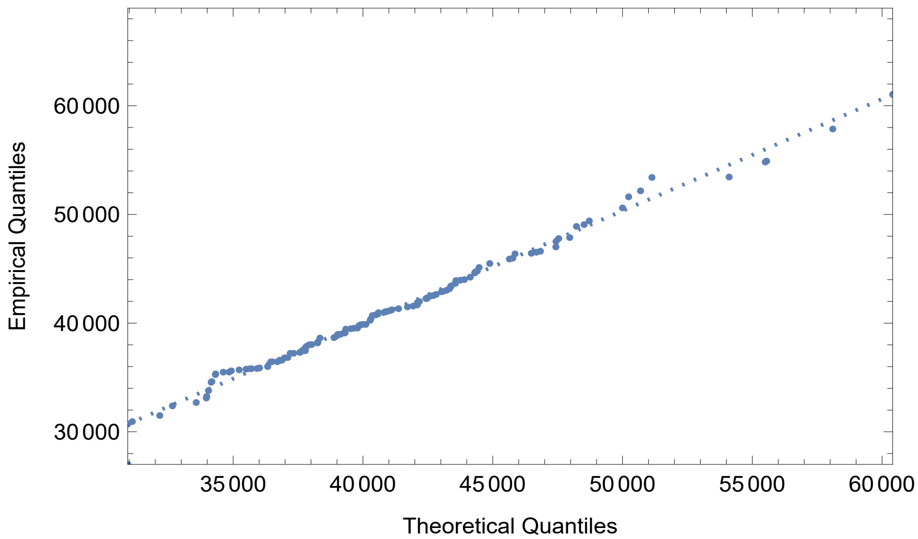

Figure 10: Q-Q plot for the per capita personal income of counties in Kentucky State in 2020

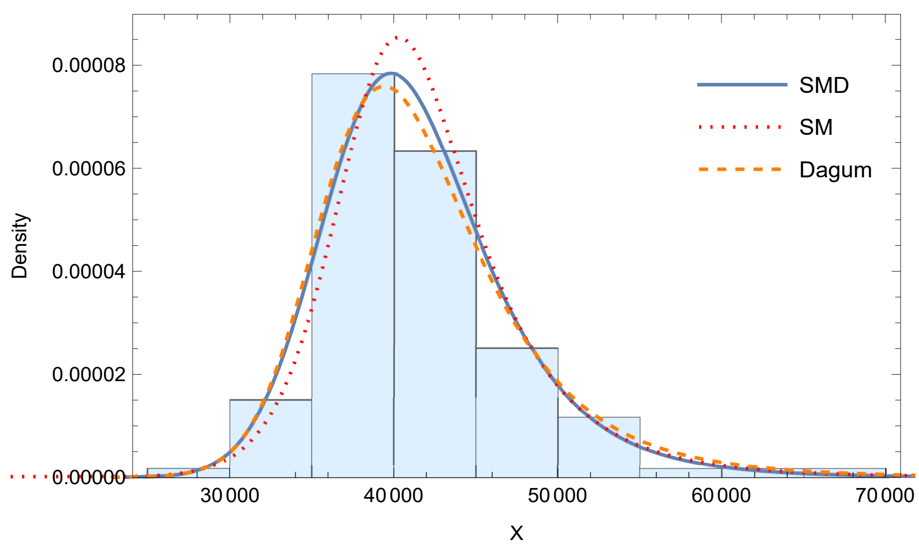

Figure 11: The densities of the SM, SMD, and Dagum distributions for the per capita personal income of counties in Kentucky State in 2020

Acknowledgements

We thank the referee and the editor for their valuable comments and suggestions. The first author is thankful to the Kerala State Council for Science, Technology and Environment

(KSCSTE) for the financial support. The authors are also thankful to the Department of Science & Technology (DST) for the technical support.

References

Bonferroni, C. E. (1930). *Elements of General Statistics*. Seeber, Florence.

Chotikapanich, D., Griffiths, W., Karunarathne, W., and Prasada Rao, D. (2013). Calculating poverty measures from the generalized beta income distribution. *Economic Record*, **89**, 48–66.

Dagum, C. (1977). A new model of personal income distribution: specification and estimation. *Économie Appliquée*, **30**, 413–437.

Éltetö, Ö. and Frigyes, E. (1968). New income inequality measures as efficient tools for causal analysis and planning. *Econometrica*, **36**, 383–396.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, **52**, 761–766.

Freimer, M. L., Kollia, G., Mudholkar, G. S., and Lin, C. T. (1988). A study of the generalized Tukey lambda family. *Communications in Statistics-Theory and Methods*, **17**, 3547–3567.

Galton, F. (1875). Statistics by intercomparison, with remarks on the law of frequency of error. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **49**, 33–46.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, **39**, 1037–1039.

Gilchrist, W. (2000). *Statistical Modeling with Quantile Functions (1st ed.).* Chapman and Hall/CRC, New York.

Gini, C. (1914). On the measurement of the concentration and variability of characters. *Proceedings of the Royal Veneto Institute of Sciences, Letters and Arts*, **73**, 1203–1248.

Greenwood, J. A., Landwehr, J. M., Matalas, N., and Wallis, J. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, **15**, 1049–1054.

Hankin, R. K. S. and Lee, A. (2006). A new family of non-negative distributions. *Australian & New Zealand Journal of Statistics*, **48**, 67–78.

Haritha, N. H., Nair, K. R. M., and Nair, N. U. (2007). *Income modeling using quantile functions*. Doctoral dissertation, Cochin University of Science and Technology, Cochin.

Hastings, J. C., Mosteller, F., Tukey, J. W., and Winsor, C. P. (1947). Low moments for small samples: A comparative study of order statistics. *The Annals of Mathematical Statistics*, **18**, 413–426.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, **52**, 105–124.

Kakwani, N. (1980). *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. Oxford University Press, Oxford.

Kakwani, N. (1999). *Inequality, Welfare And Poverty: Three Interrelated Phenomena*, pages 599–634. Springer Netherlands, Dordrecht.

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley Sons, Ltd, Hoboken, New Jersey.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2019). *Loss Models From Data to Decisions*. John Wiley, New York, 5^{th} edition.

Kumar, D. (2017). The Singh–Maddala distribution: Properties and estimation. *International Journal of System Assurance Engineering and Management*, **8 (Suppl. 2)**, 1297–1311.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**, 209–219.

Morduch, J. (2008). Poverty measures. In *United Nations Handbook of Poverty Statistics*, chapter 3, pages 52–83. United Nations, New York.

Nair, N. U., Sankaran, P. G., and Balakrishnan, N. (2013). Quantile Function Models. In *Quantile-Based Reliability Analysis*, chapter 3, pages 59–103. Springer, New York.

Öztürk, A. and Dale, R. F. (1985). Least squares estimation of the parameters of the generalized lambda distribution. *Technometrics*, **27**, 81–84.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, **74**, 105–121.

Pietra, G. (1932). New contributions to the methodology of variability and concentration indices. *Proceedings of the Royal Veneto Institute of Sciences, Letters and Arts*, **91**, 989–1008.

Ramberg, J. S. and Schmeiser, B. W. (1972). An approximate method for generating symmetric random variables. *Communications of the ACM*, **15**, 987–990.

Sankaran, P. G. and Dileep Kumar, M. (2018). Pareto-Weibull quantile function. *Journal of Applied Probability and Statistics*, **13**, 81–95.

Sankaran, P. G., Nair, N. U., and Midhu, N. N. (2016). A new quantile function with applications to reliability analysis. *Communications in Statistics - Simulation and Computation*, **45**, 566–582.

Saulo, H., Vila, R., Borges, G. V., Bourguignon, M., Leiva, V., and Marchant, C. (2023). Modeling income data via new parametric quantile regressions: Formulation, computational statistics, and application. *Mathematics*, **11**.

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, **44**, 219–231.

Shahzad, M. and Asghar, Z. (2013a). Comparing TL-moments, L-moments and conventional moments of Dagum distribution by simulated data. *Revista Colombiana de Estadística*, **36**, 79–93.

Shahzad, M. and Asghar, Z. (2013b). Parameter estimation of Singh-Maddala distribution by moments. *International Journal of Advanced Statistics and Probability*, **1**, 121–131.

Sillitto, G. P. (1969). Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika*, **56**, 641–650.

Singh, S. K. and Maddala, G. S. (1975). A stochastic process for income distribution and tests for income distribution functions. In *ASA Proceedings of the Business and Economic Statistics Section*, pages 551–553.

Sreelakshmi, N. and Nair, K. R. M. (2014). *A quantile based analysis of income data*. Doctoral dissertation, Cochin University of Science and Technology, Cochin.

Tarsitano, A. (2004). Fitting the generalized lambda distribution to income data. In *COMPSTAT 2004 Symposium*, pages 1861–1867, New York. Springer.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addsion Wesley, Massachusetts.

Watts, H. W. (1968). *An Economic Definition of Poverty*. Basic Books, New York.

Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, **V**.