ISSN 1973-2201 *STATISTICA, anno LXXXIV, n. 4, 2024*
DOI: https://doi.org/10.60923/issn.1973-2201/14913

A QUANTILE-BASED INCOME ANALYSIS OF POWER-PARETO DISTRIBUTION

Varkey Ashlin[1]
Department of Statistics, Farook College (Autonomous), Kozhikode, University of Calicut, Kerala, India

N. Haridas Haritha
Department of Statistics, Farook College (Autonomous), Kozhikode, University of Calicut, Kerala, India

SUMMARY

This paper conducts a quantile-based income study of the Power-Pareto distribution. The major income inequality measures of the Power-Pareto distribution are derived, and the Lorenz ordering is studied. A simulation study is conducted to assess the performance of four competing estimation methods. The model is applied to a real income dataset, and both empirical and estimated income inequality measures are computed.

Keywords: Power-Pareto distribution; Income inequalities; Lorenz ordering; Simulation.

1. INTRODUCTION

The distribution and quantile function approaches are two alternative ways to define a probability distribution. Most statistical models used in reliability and income studies are built in terms of the distribution function $F(x)$ and some on quantile function $Q(u)$, even though they both offer the same information about the distribution with a different interpretation. The quantile function is defined by

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

The Belgian sociologist Quetelet (1846) is considered a pioneer in utilizing quantiles for statistical analysis, particularly for introducing the concept of inter-quantile range. Although the term quantile was first used by Kendall (1940), the idea appears to have

[1] Corresponding Author. E-mail: ashlinvarkey@gmail.com

originated in an article by Galton (1875), titled "Statistics by Intercomparison, with Remarks on the Law of Frequency of Error", published in Philosophy Magazine. The papers of Hastings *et al.* (1947), Tukey (1977), and Parzen (1979) sparked the development of the quantile function as a vital tool in statistical analysis, instead of the distribution function. Gilchrist (2000) and Nair *et al.* (2013) provided a comprehensive overview of statistical modeling with quantile functions. Ramberg and Schmeiser (1974), Ramberg (1975), Ramberg *et al.* (1979), Freimer *et al.* (1988), Hankin and Lee (2006), Haritha *et al.* (2008), Van Staden and Loots (2009), Midhu *et al.* (2013), Sankaran *et al.* (2016), Sankaran and Dileep Kumar (2018b), Sankaran and Dileep Kumar (2018a), Ashlin and Haritha (2024) etc worked on quantile-based models.

Pareto (1897) was the first to model income data over a hundred years ago. Other distributions emerged in the literature due to the Pareto distribution's severe limitation in explaining only the upper tail of the distribution. The lognormal distribution captures a substantial portion of the income range but diverges significantly at the extremes. Over time, standard continuous distributions such as exponential, gamma, beta, Weibull, lognormal, and their generalizations were proposed as income distributions. The well-known works by Arnold (1983) and Kleiber and Kotz (2003) provide a comprehensive overview of income data modeling and analysis.

Income inequality indicates how income is distributed unequally across a population. Through Lorenz's work (Lorenz, 1905), income inequality curves have piqued people's interest. Later, Gini (1914) introduced an index defined as two times the area between the Lorenz curve and the egalitarian line. Then several income inequality measures such as the coefficient of variation, the variance of logarithms, Pietra index (Pietra, 1932), Bonferroni curve and index (Bonferroni, 1930), Frigyes measures (Frigyes, 1965), Atkinson measures (Atkinson, 1970), generalized entropy measures (Rohde, 2008), Zenga curve (Zenga, 1984), etc., came into existence.

The primary goal of this work is to explore the potential of the Power-Pareto distribution in income modeling within a quantile framework. We discuss various income inequality measures for this distribution, including the Lorenz curve, Gini index, Leimkuhler curve, Bonferroni curve, and index. Additionally, we derive other income inequality measures such as Pietra, Atkinson, generalized entropy, Frigyes measures, and Zenga curve, along with income gap ratios and truncated Gini indices for both the poor and rich. Lorenz ordering of Power-Pareto distribution is also studied. Finally, we apply the Power-Pareto distribution to real income data and compare empirical and theoretical inequality measures.

This paper is organized as follows: In Section 2, we discuss the genesis and properties of the Power-Pareto distribution. In Section 3, we derive the major income inequality measures of Power-Pareto in terms of quantiles and study Lorenz ordering. We perform a simulation study with the above distribution to analyze the capability of four estimation methods in Section 4. In Section 5, applications to real income data are presented. Finally, Section 6 provides the conclusion of the study.

2. The Power-Pareto Distribution

The quantile function of Power-Pareto distribution is

$$Q(u;C,\lambda_1,\lambda_2)=\frac{C\,u^{\lambda_1}}{(1-u)^{\lambda_2}},\quad 0\leq u\leq 1, C>0, \lambda_1>0, \lambda_2>0, \tag{1}$$

where C is the scale parameter, λ_1 and λ_2 are shape parameters that govern the left and right tails, respectively. It is also possible that one of the λ's can be zero but not both. This distribution is also known as the Davies distribution and the Hankin and Lee distribution. The quantile density function of the distribution is given as

$$q(u;C,\lambda_1,\lambda_2)=\frac{C\,u^{\lambda_1}}{(1-u)^{\lambda_2}}\left(\frac{\lambda_1}{u}+\frac{\lambda_2}{(1-u)}\right).$$

The density and distribution functions of the Power-Pareto distribution are not available in closed form but a numerical inversion of the quantile function can be used to determine it. The above quantile function is obtained as the product of quantile functions of the Power and Pareto distributions. Hence the quantile function in Eq. (1) reduces to Pareto and Power distributions respectively, as λ_1 and λ_2 tend to zero. Also, the scaled logistic distribution becomes a special case of Power-Pareto when $\lambda_1=\lambda_2=\lambda$.

Gilchrist (2000) examined the combined version of Power and Pareto distributions, referring to it as the Power-Pareto distribution. Hankin and Lee (2006) studied the properties of this distribution and compared it to other distributions. They used the method of maximum likelihood and a simple technique using least squares for parameter estimation. Nair *et al.* (2013) examined the basic properties of the Power-Pareto distribution, such as L-moments, percentile-based measurements, and some characterizations based on reliability measures. Using the Power-Pareto distribution Sankaran and Preetha Kumari (2010) developed a parameter regression model and evaluated the model attributes in a reliability framework. Nair and Vineshkumar (2010) demonstrated the utility of the Power-Pareto distribution in reliability analysis and verified it with real data. Sreelakshmi and Nair (2014) briefly studied and derived economic measures like the Lorenz curve and income share elasticity of the above distribution. Recently, Caeiro and Norouzirad (2024) evaluated the performance of different estimation methods for the Power-Pareto distribution. There has been limited research on income studies using the Power-Pareto distribution, so this paper focuses on a quantile-based income analysis of Power-Pareto.

3. Income Inequality Measures

Income inequality refers to the extent to which income is distributed unequally throughout a population. Even now it is a fertile field for discussions due to the huge income disparity existing among the population.

Although there exist several income inequality measures in literature, the Lorenz curve (Lorenz, 1905) still holds a significant place among other inequalities. For finite populations, the Lorenz curve is defined as a function $L(u)$ on $[0,1]$ such that, $L(u)$ indicates the fraction of total income in the population accounted for by the $100u\%$ poorest individuals in the population for fixed u. When X is a non-negative random variable with a positive mean, Gastwirth (1971) proposed a more precise definition for the Lorenz curve using the quantile function and is given by

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

where $\mu=\int_0^1 Q(p)dp$. The Lorenz curve for the Power-Pareto distribution is given as

$$\begin{aligned} L(u) &= \frac{\beta_u(\lambda_1+1,1-\lambda_2)}{\beta(\lambda_1+1,1-\lambda_2)} \\ &= I_u(\lambda_1+1,1-\lambda_2), \quad \lambda_2<1, \end{aligned} \tag{2}$$

where $\beta(\lambda_1+1,1-\lambda_2)$ denotes the complete beta function and $\beta_u(\lambda_1+1,1-\lambda_2)=\int_0^u u^{\lambda_1}(1-u)^{-\lambda_2}$, denotes an incomplete beta function. The regularized incomplete beta function is represented as $I_u(\lambda_1+1,1-\lambda_2)$.

The Gini index (Gini, 1914), is defined as twice the area between the Lorenz curve and the egalitarian line. It is the most frequently used inequality measure and is given by

$$\begin{aligned} G &= 2\int_0^1 (u-L(u))du \\ &= 1-2\int_0^1 L(u)du. \end{aligned}$$

Now Gini index corresponding to the Power-Pareto distribution is

$$G=\frac{\lambda_1+\lambda_2}{\lambda_1-\lambda_2+2}, \quad \lambda_2<1. \tag{3}$$

The generalized Gini index that is responsive to both high and low incomes was introduced by Kakwani (1980a), Donaldson and Weymark (1980, 1983), and Yitzhaki (1983) as an extension of the Gini index and is given by

$$G_v=1-v(v-1)\int_0^1 L(u)(1-u)^{v-2}du, \quad v>1.$$

For Power-Pareto the index is

$$G_v=1-\frac{v\beta(\lambda_1+1,v-\lambda_2)}{\beta(\lambda_1+1,1-\lambda_2)}, \quad \lambda_2<1. \tag{4}$$

Based on the first-moment distribution, Bonferroni (1930) introduced an income inequality curve called the Bonferroni curve, and the corresponding index is called the Bonferroni index. The Bonferroni curve and index are represented in quantile terms as

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

$$B = 1 - \int_0^1 B_F(u)du,$$

where $\mu = \int_0^1 Q(p)dp$. For the Power-Pareto distribution, these measures are

$$B_F(u) = \frac{1}{u} I_u(\lambda_1 + 1, 1 - \lambda_2), \tag{5}$$

$$B = 1 - \frac{1}{(1+\lambda_1)^2 \beta(1+\lambda_1, 1-\lambda_2)} {}_3F_2(1+\lambda_1, 1+\lambda_1, \lambda_2; 2+\lambda_1, 2+\lambda_1; 1), \tag{6}$$

where $\lambda_2 < 1$, ${}_3F_2(1+\lambda_1, 1+\lambda_1, \lambda_2; 2+\lambda_1, 2+\lambda_1; 1) = \sum_{k=0}^{\infty} \frac{(1+\lambda_1)_k (1+\lambda_1)_k (\lambda_2)_k}{(2+\lambda_1)_k (2+\lambda_1)_k} \left(\frac{1}{k!}\right)$ is a generalized hypergeometric function and $(1+\lambda_1)_k$ denotes the ascending factorial. The Lorenz curve, Gini index, Bonferroni curve, and Bonferroni index of the Power-Pareto distribution are given in Giorgi and Nadarajah (2010).

The Pietra index (Pietra, 1932) also known as the Schutz index (Duclos and Araar, 2006) is a classical income inequality measure. It has a geometric attribute of being the maximum deviation between the Lorenz curve and the line of equality. In quantile terms, the formula for the Pietra index and relative mean deviation is given as

$$P = \frac{\tau_1}{2\mu},$$

$$\tau_2 = \frac{\tau_1}{\mu},$$

where $\tau_1 = \int_0^1 |Q(u) - Q(u_0)| du$ and $\mu = Q(u_0)$ for some $0 < u_0 < 1$. Also, we can find u_0 by solving for u in the equation $\mu = Q(u)$. Now, the Pietra index corresponding to Power-Pareto is,

$$P = \frac{u_0^{\lambda_1+1}(1-u_0)^{-\lambda_2} - \beta_{u_0}(\lambda_1+1, 1-\lambda_2)}{\beta(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1. \tag{7}$$

A measure of income inequality proposed by Atkinson (1970) is called the Atkinson index and its quantile version is given by

$$A_\epsilon = 1 - \frac{1}{\int_0^1 Q(u)du}\left\{\int_0^1 Q(u)^{1-\epsilon}du\right\}^{\frac{1}{1-\epsilon}},$$

where $\epsilon > 0$ is a sensitivity parameter that gives small incomes more weight as it increases. The generalized entropy measure (Cowell and Kuga, 1981) in quantile terms is given as

$$G.E_\theta = \frac{1}{\theta(\theta-1)}\int_0^1 \left\{\left[\frac{Q(u)}{\int_0^1 Q(u)du}\right]^\theta - 1\right\}du,$$

where $\theta \in \mathbb{R} \setminus \{0,1\}$, is a sensitivity parameter. The Atkinson index and generalized entropy measure of the Power-Pareto distribution are given in Eq. (8) and Eq. (9) respectively.

$$A_\epsilon = 1 - \frac{[\beta(\lambda_1(1-\epsilon)+1, 1-\lambda_2(1-\epsilon))]^{\frac{1}{1-\epsilon}}}{\beta(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1, \tag{8}$$

$$GE_\theta = \frac{1}{\theta(\theta-1)}\left[\frac{\beta(\lambda_1\theta+1, 1-\lambda_2\theta)}{(\beta(\lambda_1+1, 1-\lambda_2))^\theta} - 1\right], \quad \lambda_2 < 1. \tag{9}$$

Frigyes (1965) introduced three measures with clear economic implications and are given as

$$\rho = \frac{m}{m_1}, \quad \nu = \frac{m_2}{m_1}, \quad \eta = \frac{m_2}{m},$$

where $m = E(X)$, $m_1 = E(X|X < m)$, $m_2 = E(X|X \geqslant m)$. The measure ν can be thought of as a measure of inequality over the entire income distribution, whereas the measures ρ and η represent inequalities in the two sections of the distribution below and above the mean, respectively. Eltetö and Frigyes (1968) have discussed the properties and applications of these three measures. In quantile terms, these measures are expressed as

$$\rho = \frac{u_0 Q(u_0)}{\int_0^{u_0} Q(u)du},$$

$$\nu = \frac{u_0}{1-u_0}\frac{\int_{u_0}^1 Q(u)du}{\int_0^{u_0} Q(u)du},$$

$$\eta = \frac{\int_{u_0}^1 Q(u)du}{(1-u_0)Q(u_0)}.$$

The Frigyes measures ρ, ν and η for the Power-Pareto are

$$\begin{aligned}
\rho &= \frac{u_0^{\lambda_1+1}}{(1-u_0)^{\lambda_2}} \frac{1}{\beta_{u_0}(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1, \\
\nu &= \frac{u_0}{(1-u_0)} \left[\frac{1}{I_{u_0}(\lambda_1+1, 1-\lambda_2)} - 1 \right], \quad \lambda_2 < 1, \\
\eta &= \frac{(1-u_0)^{\lambda_2-1}}{u_0^{\lambda_1}} \left[\beta(\lambda_1+1, 1-\lambda_2) - \beta_{u_0}(\lambda_1+1, 1-\lambda_2) \right], \quad \lambda_2 < 1.
\end{aligned} \tag{10}$$

Zenga (1984) utilized the first-moment distribution and the quantiles of the size distribution to define a new measure, known as the Zenga concentration curve. A new inequality measure based on the conditional expectations of the corresponding distribution was introduced by Zenga (2007) and is given as

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

The functional relationship between the Lorenz curve and the Zenga curve is given by

$$I(u) = \frac{u - L(u)}{u(1-L(u))}.$$

The Zenga curve corresponding to the Power-Pareto is

$$I(u) = \frac{u\beta(\lambda_1+1, 1-\lambda_2) - \beta_u(\lambda_1+1, 1-\lambda_2)}{u\left[\beta(\lambda_1+1, 1-\lambda_2) - \beta_u(\lambda_1+1, 1-\lambda_2)\right]}, \quad \lambda_2 < 1. \tag{11}$$

The Leimkuhler curve in information theory is related to the Lorenz curve through the equation

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

The difference between the construction of the Lorenz curve and the Leimkuhler curve is that the first one arranges the sources in increasing order of productivity whereas the latter arranges in decreasing order. Now the Leimkuhler curve corresponding to the Power-Pareto, as given in Eq. (12), is also provided in Nair and Vineshkumar (2022).

$$\begin{aligned}
K(u) &= 1 - \frac{\beta_{1-u}(\lambda_1+1, 1-\lambda_2)}{\beta(\lambda_1+1, 1-\lambda_2)} \\
&= 1 - I_{1-u}(\lambda_1+1, 1-\lambda_2), \quad \lambda_2 < 1.
\end{aligned} \tag{12}$$

The relationships between the Lorenz, Leimkhuler, Zenga, and Bonferroni curves are given in Sreelakshmi and Nair (2014).

The income gap ratio and the truncated Gini index are used to calculate the poverty or affluence indices linked with income data. Sen (1976), Takayama (1979) and Sen (1986) dealt with these indices. Let u and $(1-u^*)$ denote the proportion of poor and rich people in the population, respectively. The income gap ratio of the poor and the rich in quantile terms is given in Eq. (13) and Eq. (14).

$$\alpha(u) = 1 - \frac{\int_0^u Q(p)dp}{uQ(u)}, \tag{13}$$

$$\alpha^*(u^*) = 1 - \frac{(1-u^*)Q(u^*)}{\int_{u^*}^1 Q(p)dp}. \tag{14}$$

Now the truncated Gini index for the poor and the rich in quantile terms are given in Eq. (15) and Eq. (16).

$$G(u) = \frac{2[\mu(u)]^{-1}}{u^2} \int_0^u pQ(p)dp - 1, \tag{15}$$

$$G^*(u^*) = 1 - \frac{2[\mu^*(u^*)]^{-1}}{(1-u^*)^2} \int_{u^*}^1 (1-p)Q(p)dp, \tag{16}$$

where $\mu(u) = \frac{1}{u}\int_0^u Q(p)dp$ and $\mu^*(u^*) = \frac{1}{1-u^*}\int_{u^*}^1 Q(p)dp$. For the Power-Pareto distribution, the income gap ratio and the truncated Gini index for poor and rich are given as

$$\alpha(u) = 1 - \frac{(1-u)^{\lambda_2}}{u^{\lambda_1+1}} \beta_u(\lambda_1+1, 1-\lambda_2), \tag{17}$$

$$\alpha^*(u^*) = 1 - \frac{u^{*\lambda_1}(1-u^*)^{1-\lambda_2}}{\beta(\lambda_1+1, 1-\lambda_2) - \beta_{u*}(\lambda_1+1, 1-\lambda_2)}, \tag{18}$$

$$G(u) = \left(\frac{2}{u}\right) \frac{\beta_u(\lambda_1+2, 1-\lambda_2)}{\beta_u(\lambda_1+1, 1-\lambda_2)} - 1, \tag{19}$$

$$G^*(u^*) = 1 - \left(\frac{2}{1-u^*}\right) \frac{\beta(\lambda_1+1, 2-\lambda_2) - \beta_{u*}(\lambda_1+1, 2-\lambda_2)}{\beta(\lambda_1+1, 1-\lambda_2) - \beta_{u*}(\lambda_1+1, 1-\lambda_2)}. \tag{20}$$

Here in Eq. (17) to Eq. (20) we have $\lambda_2 < 1$. For the Power-Pareto distribution, we derive the generalized Gini index, Pietra index, Atkinson index, generalized entropy, Frigyes measures, Zenga curve, income gap ratios, and truncated Gini indices.

REMARK 1. *For a Power-Pareto distribution with* $(C, \lambda_1, 0)$*, the income inequality curves and measures given in Eq.(2) to Eq. (12), and Eq. (17) to Eq. (20) reduces to corresponding income inequality curves and measures of the Power distribution.*

REMARK 2. *For a Power-Pareto distribution with* $(C, 0, \lambda_2)$*, the income inequality curves and measures given in Eq.(2) to Eq. (12), and Eq. (17) to Eq. (20) reduces to corresponding income inequality curves and measures of the Pareto distribution.*

3.1. Lorenz Order

Let $\mathscr{L}$ represent the class of non-negative random variables with finite and positive mean. Now, Lorenz ordering for $X, Y \in \mathscr{L}$ is defined as

$$X \geq_L Y \Longleftrightarrow F_X \geq_L F_Y \Longleftrightarrow L_X(u) \leq L_Y(u),$$

for all $u \in [0, 1]$. Here, in Lorenz's notion, X is more unequal than Y. When two Lorenz curves do not intersect, the lower curve is said to be "more unequal", and certainly, a stochastic ordering based on this concept, the Lorenz partial ordering, has proven to be useful in a variety of situations. Whenever the Lorenz curves intersect, the Lorenz dominance has no meaning, and the use of the Gini index becomes unavoidable. Also, the Lorenz order is a partial order that is scale invariant.

When the quantile function is available in closed form, we can utilize star-shaped ordering to obtain the Lorenz ordering. If X and Y denote random variables in $\mathscr{L}$, each having quantile functions Q_X and Q_Y, then Arnold (1987) proved that the star-shaped ordering implies the Lorenz ordering. The star-shaped ordering is defined as follows.

DEFINITION 3. *We state that* X *is star-shaped with respect to* Y*, if* $Q_X(u)/Q_Y(u)$ *is a non-increasing function of* u *and denote* $X \leqslant_* Y$.

THEOREM 4. *The Lorenz ordering for a Power-Pareto distribution is,*

Case 1: $X \sim Power-Pareto(1, \lambda_1, \lambda_2)$ *and* $Y \sim Power-Pareto(1, \lambda_1, \lambda_2')$ *then* $Q_X(u)/Q_Y(u)$ *is non-increasing in* u *iff* $\lambda_2' > \lambda_2$.

Case 2: $X \sim Power-Pareto(1, \lambda_1, \lambda_2)$ *and* $Y \sim Power-Pareto(1, \lambda_1', \lambda_2)$ *then* $Q_X(u)/Q_Y(u)$ *is non-increasing in* u *iff* $\lambda_1' > \lambda_1$.

Case 3: $X \sim Power-Pareto(1, \lambda_1, \lambda_2)$ *and* $Y \sim Power-Pareto(1, \lambda_1', \lambda_2')$ *then* $Q_X(u)/Q_Y(u)$ *is non-increasing in* u *iff* $\lambda_1' > \lambda_1$ *and* $\lambda_2' > \lambda_2$.

Note that in the above three cases, the scale parameter C is assumed to be equal to 1 because Lorenz ordering is scale invariant.

4. SIMULATION STUDY FOR COMPARING FOUR ESTIMATION METHODS

In this Section, we discuss four estimation methods, including the method of L-moments and three percentile methods.

L-moments estimation is the process of equating sample L-moments to population L-moments and obtaining the estimators. Hosking (1990) introduced a unified theory of L-moments and systematically examined their properties and significance in statistical analysis. These moments generally exhibit lower sample variances and demonstrate greater robustness against outliers. Let $X_1, X_2, \cdots, X_n$ be a random sample of size n, then the r^{th} population and sample L-moments are given as

$$L_r = \int_0^1 \sum_{k=0}^{r-1} (-1)^{r-1-k} \binom{r-1}{k} \binom{r-1+k}{k} u^k Q(u) du \tag{21}$$

and

$$l_r = \frac{1}{r} \sum_{i=1}^{n} \frac{\sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \binom{i-1}{r-k-1} \binom{n-i}{k}}{\binom{n}{r}} x_{i:n}, \tag{22}$$

where $x_{i:n}$ denotes the i^{th} order statistic. By equating, $L_r = l_r$, $r = 1, 2, \ldots$ we get estimates of the parameters. Here the number of equations to be considered equals the number of parameters in the model.

The first four L-moments of the Power-Pareto distribution are given in Nair *et al.* (2013) and are,

$$\begin{aligned}
L_1 &= C\beta(\lambda_1 + 1, 1 - \lambda_2), \\
L_2 &= \frac{C(\lambda_1 + \lambda_2)}{\lambda_1 - \lambda_2 + 2} \beta(\lambda_1 + 1, 1 - \lambda_2), \\
L_3 &= \frac{C(\lambda_1^2 + \lambda_2^2 + 4\lambda_1\lambda_2 + \lambda_2 - \lambda_1)}{(\lambda_1 - \lambda_2 + 2)_{(2)}} \beta(\lambda_1 + 1, 1 - \lambda_2), \\
L_4 &= \frac{C(\lambda_1 + \lambda_2)(\lambda_1^2 + \lambda_2^2 + 8\lambda_1\lambda_2 - 3\lambda_1 + 3\lambda_2 + 2)}{(\lambda_1 - \lambda_2 + 2)_{(3)}} \beta(\lambda_1 + 1, 1 - \lambda_2).
\end{aligned} \tag{23}$$

Percentile method I refer to the percentile-based estimation approach discussed in Karian and Dudewicz (1999). Let $\tilde{\pi}_p$ be the $(100\,p)^{\text{th}}$ percentile of the dataset $X_1, X_2, \ldots, X_n$. We calcualte $\tilde{\pi}_p$ by primarily expressing $(n+1)p$ as $r + \frac{a}{b}$, where r is a positive integer, and the fraction $(\frac{a}{b})$ lies in the interval $[0, 1]$. Now, $\tilde{\pi}_p$ for the ordered data $X_{(1)}, X_{(2)} \ldots X_{(n)}$ is defined as

$$\tilde{\pi}_p = X_{(r)} + \frac{a}{b}(X_{(r+1)} - X_{(r)}). \tag{24}$$

Now the three sample percentiles $\tilde{\tau}_1, \tilde{\tau}_2$ and $\tilde{\tau}_3$, are given by

$$\begin{aligned}
\tilde{\tau}_1 &= \tilde{\pi}_{0.5}, \\
\tilde{\tau}_2 &= \tilde{\pi}_{0.9} - \tilde{\pi}_{0.1}, \\
\tilde{\tau}_3 &= \frac{\tilde{\pi}_{0.5} - \tilde{\pi}_{0.1}}{\tilde{\pi}_{0.9} - \tilde{\pi}_{0.5}},
\end{aligned} \tag{25}$$

where they represent the sample median, inter-decile range, and the ratio of left-right tail weight, respectively.
For the Power-Pareto distribution, the population percentiles are given as

$$\begin{aligned}
\tau_1 &= C2^{\lambda_2-\lambda_1}, \\
\tau_2 &= \frac{C(9^{\lambda_1} - 9^{-\lambda_2})}{10^{\lambda_1-\lambda_2}}, \\
\tau_3 &= \frac{5^{\lambda_1-\lambda_2} - 9^{-\lambda_2}}{9^{\lambda_1} - 5^{\lambda_1-\lambda_2}}.
\end{aligned} \tag{26}$$

By solving $\tilde{\tau}_1 = \tau_1$, $\tilde{\tau}_2 = \tau_2$, and $\tilde{\tau}_3 = \tau_3$, we get estimates of the parameters C, λ_1, and λ_2.

The percentile method II proposed by Haritha *et al.* (2008) uses quantile-based measurements of location, dispersion, skewness, and kurtosis. In this method population median, quartile deviation, Galton's coefficient of skewness, and Moor's kurtosis are equated to corresponding sample measures to obtain the estimators. These measures of the Power-Pareto distribution are discussed in Nair *et al.* (2013). Since the Power-Pareto distribution has only three parameters, we need only the median, quartile deviation, and Galton's skewness coefficient for estimation, which are given as follows

$$\begin{aligned}
M &= C\,2^{\lambda_2-\lambda_1}, \\
QD &= \frac{C\,4^{\lambda_2-\lambda_1}}{2}(3^{\lambda_1} - 3^{-\lambda_2}), \\
S &= \frac{3^{\lambda_1} + 3^{-\lambda_2} - 2^{\lambda_2-\lambda_1+1}}{3^{\lambda_1} - 3^{-\lambda_2}}.
\end{aligned} \tag{27}$$

The percentile method III, discussed in Caeiro and Norouzirad (2024), is popular due to its simplicity. Estimators are derived by utilizing the relationship between probabilities and percentile values through either the cumulative distribution function or the quantile function. In this estimation method, the number of percentiles considered must also equal the number of parameters. Thus, for three different cumulative probability levels p_1, p_2, and p_3 satisfying $0 < p_1 < p_2 < p_3 < 1$, the corresponding $100p_i\%$ percentiles, $i = 1,2,3$, are denoted as q_1, q_2, and q_3. Thus

$$F(q_i) = p_i \iff q_i = Q(p_i), \quad i = 1,2,3,$$

where Q(.) is the quantile function of the Power-Pareto distribution. Now, by performing a log transformation on the ratio of two consecutive percentiles, we get the following

$$\log\frac{q_2}{q_1} = \lambda_1 \log\frac{p_2}{p_1} + \lambda_2 \log\frac{1-p_1}{1-p_2},$$

and

$$\log\frac{q_3}{q_2} = \lambda_1 \log\frac{p_3}{p_2} + \lambda_2 \log\frac{1-p_2}{1-p_3}.$$

By solving the above two equations, we obtain λ_1 and λ_2 as

$$\lambda_1 = \frac{\log\frac{1-p_2}{1-p_3}\log\frac{q_2}{q_1} - \log\frac{1-p_1}{1-p_2}\log\frac{q_3}{q_2}}{\log\frac{p_2}{p_1}\log\frac{1-p_2}{1-p_3} - \log\frac{p_3}{p_2}\log\frac{1-p_1}{1-p_2}}, \tag{28}$$

and

$$\lambda_2 = \frac{-\log\frac{p_3}{p_2}\log\frac{q_2}{q_1} + \log\frac{p_2}{p_1}\log\frac{q_3}{q_2}}{\log\frac{p_2}{p_1}\log\frac{1-p_2}{1-p_3} - \log\frac{p_3}{p_2}\log\frac{1-p_1}{1-p_2}}. \tag{29}$$

Now from the equation for the second percentile, we get

$$q_2 = C p_2^{\lambda_1}(1-p_2)^{-\lambda_2},$$

$$\text{i.e.}\quad C = q_2 p_2^{-\lambda_1}(1-p_2)^{\lambda_2}. \tag{30}$$

The estimators can derived by replacing q_i in Eq. (28) to Eq. (30) by the corresponding sample percentiles. One can choose the probabilities (p_1, p_2, p_3) as $(0.1, 0.5, 0.9)$. Similarly, let I be a set containing three distinct values selected from the first n positive integers, $\{1, 2, \ldots, n\}$, where n represents the sample size. An alternative selection of percentiles is by setting $q_i = x_{(i)}$ for $i \in I$, with the corresponding cumulative probabilities given by

$$p_i = \frac{i-a}{n+b},$$

where a and b are real constants. A commonly used choice for these constants is $a = 0$ and $b = 1$. The percentile method III is computationally simple and resistant to outliers, however, it is generally less efficient compared to alternative approaches.

A simulation study was conducted to evaluate the performance of the L-moment method and the percentile methods I, II, and III. The performance of the aforementioned four competing estimation techniques is justified using data generated with the inversion method. Here, samples of sizes 100, 500, 1000, and 3000 are generated from the Power-Pareto distribution with parameters $C = 8.326$, $\lambda_1 = 0.319$, and $\lambda_2 = 0.181$. For each sample size, the above four estimation methods are employed to calculate the bias and mean square error (MSE), and the process is repeated 1000 times. This simulation process is carried out entirely with the statistical software R.

TABLE 1
Simulation table.

Estimation methods	Parameters	Sample size (n)	Absolute bias	MSE
Method of L-moments	C	100	0.04413	0.54136
		500	0.01914	0.09035
		1000	0.00994	0.05612
		3000	0.00463	0.01435
	λ_1	100	0.00222	0.00359
		500	0.00129	0.00059
		1000	0.00028	0.00037
		3000	0.00009	0.00009
	λ_2	100	0.00255	0.00178
		500	0.00091	0.00032
		1000	0.00041	0.00019
		3000	0.00026	0.00005
Method of percentiles I	C	100	0.08471	0.88213
		500	0.03844	0.18391
		1000	0.01860	0.08614
		3000	0.01192	0.02857
	λ_1	100	0.00893	0.00524
		500	0.00342	0.00116
		1000	0.00134	0.00054
		3000	0.00080	0.00018
	λ_2	100	0.00163	0.00309
		500	0.00094	0.00066
		1000	0.00044	0.00030
		3000	0.00043	0.00010
Method of percentiles II	C	100	0.15141	2.95712
		500	0.03979	0.60647
		1000	0.00831	0.28698
		3000	0.00710	0.09558
	λ_1	100	0.00337	0.01799
		500	0.00052	0.00398
		1000	0.00167	0.00188
		3000	0.00060	0.00062
	λ_2	100	0.00195	0.01499
		500	0.00078	0.00325
		1000	0.00041	0.00152
		3000	0.00030	0.00052
Method of percentiles III	C	100	0.02704	0.83185
		500	0.02527	0.18228
		1000	0.01339	0.08571
		3000	0.01026	0.02854
	λ_1	100	0.00600	0.00481
		500	0.00030	0.00114
		1000	0.00012	0.00053
		3000	0.00033	0.00018
	λ_2	100	0.00369	0.00294
		500	0.00179	0.00065
		1000	0.00100	0.00030
		3000	0.00062	0.00010

In this simulation study, as the sample size increases, the absolute bias and mean square error (MSE) in the L-moments and percentile method I of estimation diminish. From Table 1, it is evident that as the sample size increases, the MSE decreases; however, the absolute bias shows variation at a point for the parameter λ_1 in the percentile method of estimation II and III. When comparing the numerical values of bias and MSE, none of the estimation methods consistently exhibit lower bias. However, the L-moments method consistently exhibits the lowest MSE at all points compared to the other three methods. Hence, the method of L-moments provides the best estimates for the Power-Pareto distribution.

5. Data Analysis

To demonstrate the application of the Power-Pareto distribution, we consider the data from Census Bureau (2007). Table 2 represents the per capita personal income (in dollars) by industries in Indiana state in 2005. From Section 4, we found that the method of L-moments gives the best estimate for the Power-Pareto distribution. Hence we use the L-moments estimation method to estimate the parameters. The sample L-moments are

$$l_1 = 28353.7 \quad l_2 = 2081.596 \quad \text{and} \quad l_3 = 506.7143. \tag{31}$$

The estimates are then obtained by equating these sample L-moments to the corresponding population L-moments in Eq. (23). The Newton-Raphson method is used to solve the resulting system of three nonlinear equations and the estimates obtained are

$$\hat{C} = 26104.50 \quad \hat{\lambda}_1 = 0.03399 \quad \text{and} \quad \hat{\lambda}_2 = 0.10744. \tag{32}$$

TABLE 2
Per capita personal income (in dollars).

25696	23481	26678	27250	36286	23934	30617	27867
31722	27212	29649	24338	26905	26885	28778	25802
33955	31122	25756	26896	25680	25287	21667	33586
27723	27758	25704	25231	25046	35605	27352	25408
23577	27950	44354	31845	28781	32045	22699	26422
42946	27431	36466	28238	27484	27137	23922	27738
31456	36752	27744	32054	31049	26809	28639	25651
27767	31725	32246	22795	24940	25611	31784	28926
27168	26666	26787	29136	25974	26753	27767	
30067	35413	30713	27222	25419	29040	34194	
24498	26399	27469	26500	23583	32354	27402	
25635	27984	27777	28688	24077	24571	27425	

Two goodness of fit methodologies are used to assess the model's adequacy. The first is the Q-Q plot, given in Figure 1. It shows that the above data shows agreement with the Power-Pareto distribution. The chi-square test is also used to determine the goodness of fit. The test statistic value is 9.58381, with a p-value of 0.65242, indicating that the data follows the Power-Pareto distribution.

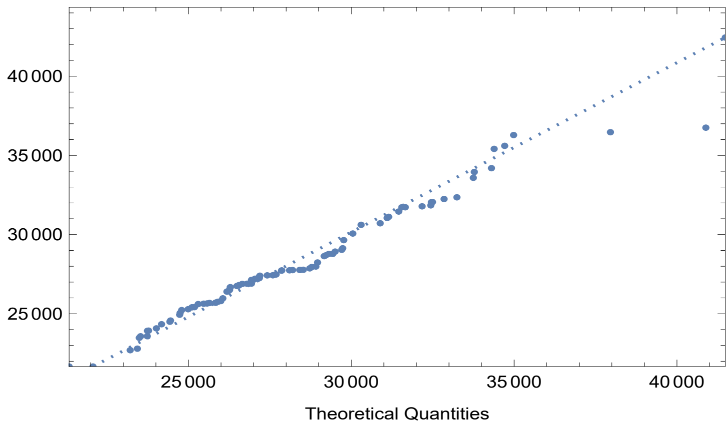

Figure 1 – Q-Q plot for the per capita personal income of Indiana in 2005.

The theoretical estimates and empirical values of the Gini, Bonferroni, Pietra, Atkinson, generalized entropy, and Frigyes measures of Indiana state are given in Table 3. The empirical values of these inequality measures except Bonferroni and Frigyes measures are computed using the 'ineq' package in R, while the theoretical estimates of these income inequality measures are computed for the Power-Pareto distribution. Let x_i denote the income of the i^{th} unit and n be the total number of observations. Then the empirical Bonferroni index is determined after arranging the observations in ascending order and is given by

$$\widehat{B} = \frac{\sum_{i=1}^{n}\left(x_i - \sum_{j=1}^{i}\frac{x_j}{i}\right)}{\sum_{i=1}^{n} x_i}. \tag{33}$$

The empirical Frigyes measures can be calculated using the equations

$$\begin{aligned} \widehat{\rho} &= \left(\frac{m-1}{n}\right)\frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{m-1} x_i}, \\ \widehat{\nu} &= \left(\frac{m-1}{n-m}\right)\frac{\sum_{i=m}^{n} x_i}{\sum_{i=1}^{m-1} x_i}, \\ \widehat{\eta} &= \left(\frac{n}{n-m}\right)\frac{\sum_{i=m}^{n} x_i}{\sum_{i=1}^{n} x_i}, \end{aligned} \tag{34}$$

where $m = E(X)$. To calculate the empirical and theoretical values of Atkinson index and generalized entropy, we use $\epsilon = 0.5$ and $\theta = 0.5$, respectively.

The Power-Pareto distribution provides the closest estimate to the empirical value, with a maximum absolute difference (Abs .diff) of 0.01709. The term 'Abs. diff' in Table 3 denotes the absolute difference between the estimated and the empirical value.

TABLE 3
Income inequality indices of Indiana.

Income inequality		**Empirical**	**Estimated**	**Abs. diff**
Gini index		0.07262	0.07342	0.00080
Bonferroni index		0.10123	0.10306	0.00183
Pietra index		0.05163	0.05191	0.00028
Atkinson index		0.00454	0.00472	0.00018
Generalized entropy		0.00910	0.00945	0.00035
Frigyes measures	ρ	1.08597	1.09389	0.00792
	ν	1.24716	1.23756	0.00960
	η	1.14843	1.13134	0.01709

6. Conclusion

In this paper, we carried out a quantile-based income study of the Power-Pareto distribution. Major income inequality measures of the Power-Pareto distribution, such as the Pietra index, Atkinson index, generalized entropy, Frigyes measures, and the Zenga curve, are derived, along with income gap ratios and truncated Gini indices for both the poor and the rich. Additionally, the Lorenz ordering of the Power-Pareto distribution is examined. The performance of four estimation methods is evaluated through a simulation study, with the L-moment method outperforming the other three. The model is applied to real income data, and both empirical and estimated income inequality indices are calculated. For future research, we suggest exploring whether other quantile-based models used in reliability analysis have potential applications in income modeling.

Acknowledgements

We thank the referee and the editor for their constructive comments. The authors also express their gratitude to the Department of Science & Technology (DST) and the Department of Biotechnology (DBT) for their technical support. The first author is thankful to the Kerala State Council for Science, Technology and Environment (KSC-STE) for the financial support.

REFERENCES

B. C. ARNOLD (1983). *Pareto Distribution*. International Cooperative Publishing House, Silver Spring Maryland.

B. C. ARNOLD (1987). *Some related orderings*. In *Majorization and the Lorenz Order: A Brief Introduction*, Springer, New York, pp. 77–89.

V. ASHLIN, N. H. HARITHA (2024). *Singh Maddala Dagum distribution with application to income data*. Statistics and Applications, 22, no. 1, pp. 321–341.

A. B. ATKINSON (1970). *On the measurement of inequality*. Journal of Economic Theory, 2, no. 3, pp. 244–263.

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

F. CAEIRO, M. NOROUZIRAD (2024). *Comparing estimation methods for the Power–Pareto distribution*. Econometrics, 12, no. 3.

U. S. CENSUS BUREAU (2007). *County and City Data Book:2007.* Washington, DC.

F. COWELL, K. KUGA (1981). *Additivity and the entropy concept: An axiomatic approach to inequality measurement*. Journal of Economic Theory, 25, pp. 131–143.

D. DONALDSON, J. A. WEYMARK (1980). *A single-parameter generalization of the Gini indices of inequality*. Journal of Economic Theory, 22, no. 1, pp. 67–86.

D. DONALDSON, J. A. WEYMARK (1983). *Ethically flexible Gini indices for income distributions in the continuum*. Journal of Economic Theory, 29, no. 2, pp. 353–358.

J. Y. DUCLOS, A. ARAAR (2006). *Poverty and Equity: Measurement, Policy and Estimation with DAD.* Springer, New York.

Ö. ELTETÖ, E. FRIGYES (1968). *New income inequality measures as efficient tools for causal analysis and planning*. Econometrica, 36, no. 2, pp. 383–396.

M. L. FREIMER, G. KOLLIA, G. S. MUDHOLKAR, C. T. LIN (1988). *A study of the generalized Tukey lambda family*. Communications in Statistics - Theory and Methods, 17, no. 10, pp. 3547–3567.

E. FRIGYES (1965). *Analysis and planning of the income distribution of workers and employees (in Hungarian).* Candidate's dissertation, Budapest.

F. GALTON (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, no. 322, pp. 33–46.

J. L. GASTWIRTH (1971). *A general definition of the Lorenz curve*. Econometrica, 39, no. 6, pp. 1037–1039.

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

C. GINI (1914). *Sulla misura della concentrazione e della variabilità dei caratteri*. Atti del Reale Istituto veneto di scienze, lettere ed arti, 73, pp. 1203–1248.

G. M. GIORGI, S. NADARAJAH (2010). *Bonferroni and Gini indices for various parametric families of distributions*. Metron - International Journal of Statistics, LXVIII, pp. 23–46.

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

N. H. HARITHA, N. U. NAIR, K. R. M. NAIR (2008). *Modelling income using generalized lambda distributions*. Journal of income distribution, 17, no. 2, pp. 37–51.

J. C. HASTINGS, F. MOSTELLER, J. W. TUKEY, C. P. WINSOR (1947). *Low moments for small samples: A comparative study of order statistics*. The Annals of Mathematical Statistics, 18, no. 3, pp. 413–426.

J. R. M. HOSKING (1990). *L-moments: Analysis and estimation of distributions using linear combinations of order statistics*. Journal of the Royal Statistical Society: Series B (Methodological), 52, no. 1, pp. 105–124.

N. KAKWANI (1980a). *On a class of poverty measures*. Econometrica, 48, no. 2, pp. 437–446.

Z. A. KARIAN, E. J. DUDEWICZ (1999). *Fitting the generalized lambda distribution to data: A method based on percentiles*. Communications in Statistics - Simulation and Computation, 28, no. 3, pp. 793–819.

M. G. KENDALL (1940). *Note on the distribution of quantiles for large samples*. Supplement to the Journal of the Royal Statistical Society, 7, pp. 83–85.

C. KLEIBER, S. KOTZ (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, New Jersey.

M. O. LORENZ (1905). *Methods of measuring the concentration of wealth*. Publications of the American Statistical Association, 9, no. 70, pp. 209–219.

N. N. MIDHU, P. G. SANKARAN, N. U. NAIR (2013). *A class of distributions with the linear mean residual quantile function and it's generalizations*. Statistical Methodology, 15, pp. 1–24.

N. U. NAIR, P. G. SANKARAN, N. BALAKRISHNAN (2013). *Quantile-Based Reliability Analysis*. Springer, New York.

N. U. NAIR, B. VINESHKUMAR (2010). *L-moments of residual life*. Journal of Statistical Planning and Inference, 140, no. 9, pp. 2618–2631.

N. U. NAIR, B. VINESHKUMAR (2022). *Modelling informetric data using quantile functions*. Journal of Informetrics, 16, no. 2, p. 101266.

V. PARETO (1897). *Cours d'Économie Politique*, vol. 2. F. Rouge, Lausanne.

E. PARZEN (1979). *Nonparametric statistical data modeling*. Journal of the American Statistical Association, 74, no. 365, pp. 105–121.

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

L. A. J. QUETELET (1846). *Letters Addressed to HRH the Grand Duke of Saxe Coburg and Gotha in the Theory of Probability*. Charles and Edwin Laton, London.

J. S. RAMBERG (1975). *A probability distribution with applications to Monte Carlo simulation studies*. In G. P. PATIL, S. KOTZ, J. K. ORD (eds.), *A Modern Course on Statistical Distributions in Scientific Work*. Springer Netherlands, Dordrecht, pp. 51–64.

J. S. RAMBERG, E. J. DUDEWICZ, P. R. TADIKAMALLA, E. F. MYKYTKA (1979). *A probability distribution and its uses in fitting data*. Technometrics, 21, no. 2, pp. 201–214.

J. S. RAMBERG, B. W. SCHMEISER (1974). *An approximate method for generating asymmetric random variables*. Communications of the ACM, 17, pp. 78–82.

N. ROHDE (2008). *Lorenz curves and generalised entropy inequality measures*. In D. CHOTIKAPANICH (ed.), *Modeling Income Distributions and Lorenz Curves*, Springer, Economic Studies in Inequality, Social Exclusion, and Well-Being, chap. 15, pp. 271–283.

P. G. SANKARAN, M. DILEEP KUMAR (2018a). *A class of distributions with the quadratic mean residual quantile function*. Communications in Statistics - Theory and Methods, 48, no. 19, pp. 4936–4957.

P. G. SANKARAN, M. DILEEP KUMAR (2018b). *A new class of quantile functions useful in reliability analysis*. Journal of Statistical Theory and Practice, 12, no. 3, pp. 615–634.

P. G. SANKARAN, N. U. NAIR, N. N. MIDHU (2016). *A new quantile function with applications to reliability analysis*. Communications in Statistics - Simulation and Computation, 45, no. 2, pp. 566–582.

P. G. SANKARAN, P. V. PREETHA KUMARI (2010). *A parametric quantile regression model useful in reliability analysis*. Calcutta Statistical Association Bulletin, 62, no. 3-4, pp. 183–206.

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

P. K. SEN (1986). *The Gini coefficient and poverty indexes: Some reconciliations*. Journal of the American Statistical Association, 81, no. 396, pp. 1050–1057.

N. SREELAKSHMI, K. R. M. NAIR (2014). *A quantile based analysis of income data*. Ph.D. thesis, Cochin University of Science and Technology, Cochin.

N. TAKAYAMA (1979). *Poverty, income inequality, and their measures: Professor Sen's axiomatic approach reconsidered*. Econometrica, 47, no. 3, pp. 747–759.

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

P. J. VAN STADEN, M. T. LOOTS (2009). *Method of L-moment estimation for the generalized lambda distribution*. In Proceedings of the Third Annual ASEARC Conference, New Castle, Australia.

S. YITZHAKI (1983). *On an extension of the Gini inequality index*. International Economic Review, 24, no. 3, pp. 617–628.

M. ZENGA (1984). *Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili di reddito*. Giornale degli Economisti e Annali di Economia, 43, pp. 301–326.

M. ZENGA (2007). *Inequality curve and inequality index based on the ratios between lower and upper arithmetic means*. Statistica & Applicazioni, 5, no. 1, pp. 3–27.