Quantile-Based Study of Income Distributions and Income Inequality Measures

Thesis submitted to the University of Calicut

for the award of the degree of

DOCTOR OF PHILOSOPHY

IN

STATISTICS

under the Faculty of Science

by

Ashlin Varkey

under the guidance of

Dr. Haritha N Haridas

Postgraduate and Research Department of Statistics

Farook College (Autonomous)

Kozhikode, Kerala - 673632

Affiliated to the University of Calicut

August 2025

CERTIFICATE

I hereby certify that the thesis titled ***"Quantile-Based Study of Income Distributions and Income Inequality Measures"*** submitted by **Ms. Ashlin Varkey** to the **University of Calicut** in partial fulfillment of the requirements for the award of the Degree of ***Doctor of Philosophy in Statistics*** is a record of original research work carried out by her under my supervision. The content of this work has not been included in any other thesis submitted previously for the award of any degree or diploma. Also certify that the contents of the thesis have been checked using an anti-plagiarism database and no unacceptable similarity was found through the software check.

Kozhikode
18 August, 2025

Dr. Haritha N Haridas
Research Supervisor

CERTIFICATE

Certified that the corrections/suggestions, recommended by the adjudicators, of the Ph.D. Thesis titled ***"Quantile-Based Study of Income Distributions and Income Inequality Measures"*** submitted by **Ms. Ashlin Varkey**, Research scholar, Department of Statistics, Farook College (Autonomous), under my supervision and guidance, have been incorporated in this copy of the thesis, and the contents in the thesis and the soft copy are one and the same.

Kozhikode

18 August, 2025

Dr. Haritha N Haridas

Research Supervisor

DECLARATION

I, **Ashlin Varkey**, hereby declare that the work presented in the thesis titled ***"Quantile-Based Study of Income Distributions and Income Inequality Measures"*** is based on the original work done by me under the guidance of Dr. Haritha N Haridas, Associate Professor, Department of Statistics, Farook College (Autonomous), Kozhikode, Kerala and has not been included in any other thesis submitted previously for the award of any degree. The contents of the thesis have undergone a plagiarism check using iThenticate software at C.H.M.K. Library, University of Calicut, and the similarity index found within the permissible limit. I also declare that the thesis is free from AI-generated content.

Kozhikode | **Ashlin Varkey**

18 August, 2025

ACKNOWLEDGEMENTS

I would like to take this opportunity to extend my sincere gratitude to all those who have supported and guided me throughout my PhD journey.

Above all, I offer my deepest reverence to God, whose unwavering presence has been my source of strength and grace through every challenge and success, enabling me to bring this work to completion.

I am truly fortunate to have had Dr. Haritha N Haridas, Associate Professor in the Department of Statistics at Farook College (Autonomous), Kozhikode, as my guide and supervisor. Her steadfast support, gentle encouragement, and ability to foster a positive and productive research environment have been instrumental in driving my work forward. I am deeply grateful for her invaluable insights and mentorship throughout this journey.

I also extend my sincere thanks to Dr. Abdul Rasheed K. V., Head of the Department of Statistics at Farook College, along with all the faculty members of the department, for their support and guidance.

My heartfelt gratitude goes to Dr. Aysha Swapna K. A., our esteemed Principal, for her constant encouragement. I would like to express my sincere gratitude to Farook College for generously providing the necessary infrastructural facilities. I also thank Dr. Mansoor Babu V., our college librarian, for his invaluable assistance.

I am grateful to Dr. N. Unnikrishnan Nair (CUSAT), Dr. P. Anilkumar (St. Joseph's College, Devagiri), Dr. K. M. Naseer (Farook College), Dr. R. M. Juvairiyya (Farook College), Dr. Krishnarani S. D. (University of Calicut), and Dr. Dileepkumar M. (University of Calicut) for sharing their knowledge and providing valuable guidance throughout my studies.

My heartfelt thanks go to all my former teachers and friends who have supported me throughout this journey. I am especially grateful to my dear friends, Anakha K. K. and Fathima Jafna, for their support, valuable suggestions, and assistance in clarifying doubts during my research. I extend my love and gratitude to all the scholars from our research room who made these years memorable and inspiring, especially Nitha K. U., Vidya V. P., Jouhara P. N., Fathima Thensi N., and Reham Ali, for their invaluable encouragement and unconditional support.

I am grateful to the Kerala State Council for Science, Technology, and Environment (KSCSTE) for providing me with financial support during the first three years of my research.

I could not have completed my thesis without the unwavering love and support of my family. I am truly blessed to have a family that has stood by me with patience and understanding throughout this journey. I am profoundly grateful to my parents for being my pillars of strength in the pursuit of my dream. I also extend my heartfelt thanks to my sister for her constant encouragement, keen interest in my research, and inspiring support.

Once again, I extend my heartfelt thanks to everyone who has supported me in various ways throughout the completion of this research work.

Ashlin Varkey

ABSTRACT

Income distributions and income inequality measures are key topics in economics and statistics, focusing on the allocation of wealth and resources across different segments of a population. Understanding these concepts is crucial for evaluating economic policies, social equity, and economic development. There are two primary approaches to modeling data: the distribution function approach and the quantile function approach. This thesis emphasizes a quantile-based analysis of income distributions and income inequality measures. We explore various quantile functions from the literature and assess their potential for modeling income data. We conduct a comprehensive quantile-based income analysis of the Power-Pareto (PP) distribution by deriving key income inequality measures and examining the Lorenz ordering associated with it. Additionally, simulation, estimation, and application aspects are investigated. This study introduces the Singh-Maddala-Dagum (SMD) distribution, defined as the sum of the quantile functions of the Singh-Maddala (SM) and Dagum distributions. Its distributional properties, along with measures of income inequality and poverty, are derived. The poverty gap ratio and Foster-Greer-Thorbecke (FGT) measures are formulated in quantile terms. Estimation of parameters and practical applications are conducted. Furthermore, this thesis includes a quantile-based comparative analysis of income inequality across Indian states. Six parametric models with closed-form quantile functions, including Weibull, PP, SM, Dagum, SMD, and Modified Lambda Family (MLF), are employed to model per capita household income across Indian states using data from the India Human Development Survey-II (IHDS-II). Parameter estimation and validation are performed for each model. For every state, empirical income inequality measures, including the Gini, Pietra, Atkinson, generalized entropy, Bonferroni, and Frigyes measures, are derived and compared with theoretical values from the best-fitting model. This thesis concludes by highlighting the significance of quantile functions in income modeling and outlining potential directions for future research.

Keywords: Income distributions; Income inequality measures; Power-Pareto distribution; Singh-Maddala Dagum distribution; Indian states.

പഠനസംഗ്രഹം

വരുമാന വിതരണങ്ങളും വരുമാന അസമത്വ അളവുകളും സാമ്പത്തിക ശാസ്ത്രത്തിലും സ്ഥിതിവിവരക്കണക്കുകളിലും പ്രധാന വിഷയങ്ങളാണ്, ഇവ ഒരു ജനസംഖ്യയുടെ വിവിധ വിഭാഗങ്ങളിലുടനീളം സമ്പത്തിന്റെയും വിഭവങ്ങളുടെയും വിഹിതത്തിൽ ശ്രദ്ധ കേന്ദ്രീകരിക്കുന്നു. സാമ്പത്തിക നയങ്ങൾ, സാമൂഹിക തുല്യത, സാമ്പത്തിക വികസനം എന്നിവ വിലയിരുത്തുന്നതിന് ഈ ആശയങ്ങൾ മനസ്സിലാക്കുന്നത് നിർണായകമാണ്. ഡാറ്റ മോഡലിംഗ് ചെയ്യുന്നതിന് രണ്ട് പ്രാഥമിക സമീപനങ്ങളുണ്ട്: വിതരണ ഏകദ സമീപനവും, ക്വാണ്ടൈൽ ഏകദ സമീപനവും. വരുമാന വിതരണങ്ങളുടെയും വരുമാന അസമത്വ അളവുകളുടെയും ക്വാണ്ടൈൽ അടിസ്ഥാനമാക്കിയുള്ള വിശകലനത്തിന് ഈ പ്രബന്ധം ഊന്നൽ നൽകുന്നു. ഗ്രന്ഥസൂചിക അവലോകനത്തിൽ നിന്ന് വിവിധ ക്വാണ്ടൈൽ ഏകദങ്ങൾ ഞങ്ങൾ പര്യവേക്ഷണം ചെയ്യുകയും വരുമാന ഡാറ്റ മോഡലിംഗ് ചെയ്യുന്നതിനുള്ള അവയുടെ സാധ്യതകൾ വിലയിരുത്തുകയും ചെയ്യുന്നു. പ്രധാന വരുമാന അസമത്വ അളവുകൾ ഉരുത്തിരിഞ്ഞുകൊണ്ടും അതുമായി ബന്ധപ്പെട്ട ലോറൻസ് ക്രമം പരിശോധിച്ചുകൊണ്ടും പവർ-പാരെറ്റോ (പി. പി.) വിതരണത്തിന്റെ സമഗ്രമായ ക്വാണ്ടൈൽ അടിസ്ഥാനമാക്കിയുള്ള വരുമാന വിശകലനം ഞങ്ങൾ നടത്തുന്നു. കൂടാതെ, സിമുലേഷൻ, എസ്റ്റിമേഷൻ, ആപ്ലിക്കേഷൻ വശങ്ങൾ എന്നിവ പഠിക്കുന്നു. സിംഗ്-മദ്ദാല (എസ്. എം.), ഡാഗം വിതരണങ്ങളുടെ ക്വാണ്ടൈൽ ഏകദങ്ങളുടെ ആകെത്തുകയായി നിർവചിച്ചിരിക്കുന്ന സിംഗ് മദ്ദാല ഡാഗം (എസ്. എം. ഡി.) വിതരണത്തെ ഈ പഠനം പരിചയപ്പെടുത്തുന്നു. വരുമാന അസമത്വത്തിന്റെയും ദാരിദ്ര്യത്തിന്റെയും അളവുകൾക്കൊപ്പം അതിന്റെ വിതരണ സവിശേഷതകളും കണ്ടെത്തിയിട്ടുണ്ട്. ദാരിദ്ര്യ വിടവ് അനുപാതവും ഫോസ്റ്റർ-ഗ്രീർ-തോർബെക്ക് (എഫ്. ജി. ടി.) അളവുകളും ക്വാണ്ടൈൽ പദങ്ങളിലാണ് രൂപപ്പെടുത്തിയിരിക്കുന്നത്. എസ്. എം. ഡി. യുടെ എസ്റ്റിമേഷനും പ്രായോഗിക പ്രയോഗങ്ങളും നടത്തുന്നു. കൂടാതെ, ഇന്ത്യൻ സംസ്ഥാനങ്ങളിലുടനീളമുള്ള വരുമാന അസമത്വത്തിന്റെ ക്വാണ്ടൈൽ അടിസ്ഥാനമാക്കിയുള്ള താരതമ്യ വിശകലനം ഈ പ്രബന്ധത്തിൽ ഉൾപ്പെടുന്നു. ഇന്ത്യ ഹ്യൂമൻ ഡെവലപ്മെന്റ് സർവേ- II (ഐ. എച്ച്. ഡി. എസ്.-II) യിൽ നിന്നുള്ള ഡാറ്റ ഉപയോഗിച്ച് ഇന്ത്യൻ സംസ്ഥാനങ്ങളിലുടനീളമുള്ള പ്രതിശീർഷ ഗാർഹിക വരുമാന മോഡലിങ്ങിനു വെയ്ബുൾ, പി. പി., എസ്. എം., ഡാഗം, എസ്. എം. ഡി., മോഡിഫൈഡ് ലാംഡ ഫാമിലി (എം. എൽ. എഫ്.) എന്നിവയുൾപ്പെടെ ക്ലോസ്ഡ്-ഫോം ക്വാണ്ടൈൽ ഏകദങ്ങളുള്ള ആറ് പാരാമെട്രിക് മോഡലുകൾ ഉപയോഗിക്കുന്നു. ഓരോ മോഡലിനും പാരാമീറ്റർ എസ്റ്റിമേഷനും സാധൂകരണവും നടത്തുന്നു. ഓരോ സംസ്ഥാനത്തിനും, ഗിനി, പിയട്ര, അറ്റ്കിൻസൺ, സാമാന്യവൽക്കരിച്ച എൻട്രോപ്പി, ബോൺഫെറോണി, ഫ്രിഗീസ് അളവുകൾ ഉൾപ്പെടെയുള്ള എംപിരിക്കൽ വരുമാന അസമത്വ അളവുകൾ ഏറ്റവും അനുയോജ്യമായ മോഡലിൽ നിന്ന് ഉരുത്തിരിഞ്ഞ സൈദ്ധാന്തിക മൂല്യങ്ങളുമായി താരതമ്യം ചെയ്യുന്നു. വരുമാന മോഡലിംഗിൽ ക്വാണ്ടൈൽ ഏകദങ്ങളുടെ പ്രാധാന്യം എടുത്തുകാണിച്ചുകൊണ്ടും ഭാവി ഗവേഷണത്തിനു സാധ്യതയുള്ള ദിശകൾ രൂപപ്പെടുത്തിക്കൊണ്ടുമാണ് ഈ പ്രബന്ധം അവസാനിക്കുന്നത്.

പ്രധാന പദങ്ങൾ: വരുമാന വിതരണങ്ങൾ; വരുമാന അസമത്വ അളവുകൾ; പവർ-പാരെറ്റോ വിതരണം; സിംഗ് മദ്ദാല ഡാഗം വിതരണം; ഇന്ത്യൻ സംസ്ഥാനങ്ങൾ.

To My Family

CONTENTS

LIST OF FIGURES

List of Figures

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Statistical modeling of income distributions and analysis of income inequality measures have been extensively studied in the literature. Beginning from Pareto's model by Pareto (1897), a vast collection of statistical distributions has been developed for modeling income data. There are two main approaches to modeling data: the distribution function approach and the quantile function approach. While a substantial amount of research exists on the former, studies on the latter remain relatively limited.

The quantile function approach offers several advantages over the distribution function approach. Similar to the distribution function, but with minimal effort, the distributional characteristics such as location, dispersion, skewness, kurtosis, etc., can be directly derived from the quantile function. Motivated by this, we have tried to use the potential of quantile functions for modeling income data. The quantile-based approach to income modeling was initially introduced by Tarsitano (2004). Later, Haritha et al. (2007) and Sreelakshmi and Nair (2014) have utilized the quantile function approach for income modeling.

The study of income inequality and poverty measures has been widely explored in the literature. There is a scope in evaluating these measures in quantile form when the distribution function does not exist in an explicit form. So the present work focuses on modeling income data and examining income inequality measures in quantile form.

1.2 Quantile function - The basic concepts

The distribution function and quantile function approaches are two distinct methods of defining a probability distribution. The quantile function is a pivotal concept in probability and statistics, serving as the inverse of the cumulative distribution function (CDF). It is widely used in various fields, such as finance, econometrics, and reliability analysis.

The quantile function $Q(u)$ is defined for a real-valued continuous random variable X with a right continuous distribution function $F(x)$ as

$$Q(u) = F^{-1}(u) = \inf \{x : F(x) \geq u\}, \; 0 \leq u \leq 1. \tag{1.2.1}$$

We have $F(x) \geq u$ if and only if $Q(u) \leq x$ for all $-\infty < x < \infty$ and $0 < u < 1$. If we can find an x such that $F(x) = u$, then $F(Q(u)) = u$ and the smallest value of x that satisfies $F(x) = u$ is $Q(u)$. When $F(x)$ is continuous and strictly increasing, $Q(u)$ is the only value of x that satisfies $F(x) = u$. Hence the quantile function of X can be obtained by solving the equation $F(x) = u$.

Quetelet (1846) pioneered the use of quantiles in statistical analysis, particularly by introducing a concept similar to the inter-quantile range. Although the term 'quantile' was first introduced by Kendall (1940), the concept itself appears to have originated in Galton's (1875) article titled 'Statistics by intercomparison, with remarks on the Law of Frequency of Error' published in Philosophical Magazine. The works of Hastings et al. (1947), Tukey (1977), and Parzen (1979) played a significant role in establishing the quantile function as an essential tool in statistical analysis, in contrast to the distribution function. Later, Gilchrist (2000) and Nair et al. (2013a) provided a comprehensive overview of statistical modeling with quantile functions.

The key properties of the quantile functions that will be useful in the following discussion are

(a) $Q(u)$ is non-decreasing on the interval $(0, 1)$ and $Q(F(x)) \leq x$ for all $x \in \mathbb{R}$ for which $0 < F(x) < 1$.

(b) For any $0 < u < 1$, $F(Q(u)) \geq u$.

(c) $Q(u)$ is left continuous i.e., $Q(u^-) = Q(u)$.

(d) $Q(u^+) = \inf\{x : F(x) > u\}$, indicates that $Q(u)$ has limits from above.

(e) Any jumps in $F(x)$ corresponds to flat point in $Q(u)$, and vice versa.

(f) If U is a random variable having uniform distribution on $(0, 1)$, and $F(x)$ is the distribution function of the random variable $X = Q(U)$ then we have $P(Q(U) \leq x) = P(U \leq F(x)) = F(x)$. Thus $Q(U)$ and X are identically distributed and this property enables us to generate data from any distribution having closed form quantile function.

(g) $H(Q(u))$ is a quantile function whenever $H(x)$ is a non-decreasing function of x. Similarly, $H(Q(1-u))$ is a quantile function whenever $H(x)$ is a non-increasing function of x.

(h) Let X be a continuous random variable with quantile function $Q(u)$ and $H(u)$ is a non-decreasing function meeting the boundary constraints $H(0) = 0$ and $H(1) = 1$, then $Q(H(u))$ is a quantile function of a random variable having the same support as X.

(i) The sum of two quantile functions and the product of two positive quantile functions are again quantile functions.

(j) If the random variable X has quantile function $Q(u)$, then $\frac{1}{X}$ has quantile function $\frac{1}{Q(1-u)}$.

These properties make quantile functions useful for modeling and analyzing statistical data.

1.3 Quantile density function

If X is a random variable with quantile function $Q(u)$ then the quantile density function is simply the derivative of $Q(u)$, i.e.,

$$q(u) = Q'(u). \tag{1.3.1}$$

If X has probability density function $f(x)$, then the density quantile function is given as $f(Q(u))$. We observe that the quantile density function and density quantile function are reciprocal to each other, as differentiating the identity $F(Q(u)) = u$ leads to the relationship

$$q(u)f(Q(u)) = 1. \tag{1.3.2}$$

1.4 Quantile-based definitions

This section includes several general concepts that will be useful in the following sections, such as percentiles, order statistics, moments, and L-moments in quantile terms.

1.4.1 Percentiles

Percentiles are used in statistics to indicate the value below which a specific percentage of observations in a dataset occurs. The dataset is divided into 100 equal portions, with p^{th} percentile indicating the value below which $100.p\%$ of the observations lie. Several distributional properties can be studied using percentile-based measures, and frequently used measures are given below.

The median, a measure of location is determined using the equation

$$M = Q(0.5). \tag{1.4.1}$$

The interquartile range is used to quantify dispersion and is given as

$$IQR = Q\left(0.75\right) - Q\left(0.25\right). \tag{1.4.2}$$

The quartile deviation is

$$QD = \frac{IQR}{2}. \tag{1.4.3}$$

Galton's coefficient of skewness, also referred to as Bowley's coefficient of skewness (Bowley (1920)), is

$$S = \frac{Q\left(0.75\right) + Q\left(0.25\right) - 2M}{IQR}. \tag{1.4.4}$$

Moor's coefficient of kurtosis, introduced by Moors (1988), is calculated using the equation

$$T = \frac{Q\,(0.875) - Q\,(0.625) + Q\,(0.375) - Q\,(0.125)}{IQR}. \tag{1.4.5}$$

In the cases of extreme positive skewness, $Q(0.25) \longrightarrow M$ while in cases of extreme negative skewness $Q(0.75) \longrightarrow M$, ensuring that S takes values between -1 and 1. Also, when the distribution is symmetric, i.e., $M = \frac{Q(0.25)+Q(0.75)}{2}$, then $S = 0$. The coefficient of kurtosis T is validated because the differences $Q(0.875) - Q(0.625)$ and $Q(0.375) - Q(0.125)$ increase (decrease) when a relatively small (large) amount of probability mass is dense around $Q(0.75)$ and $Q(0.25)$. This corresponds to a greater (lesser) spread near the $\mu \pm \sigma$.

1.4.2 Order statistics

Let $X_1, X_2 \ldots X_n$ be a random sample of size n having common distribution function $F(x)$, i.e., $X_1, X_2 \ldots X_n$ are independently and identically distributed random variables with distribution function $F(x)$. The order statistics of the random sample are obtained by arranging the sample values in ascending order and are denoted as $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$. Here, $X_{(1)} = \min_{1 \le i \le n} X_i$, represents the minimum value and $X_{(n)} = \max_{1 \le i \le n} X_i$, represents the maximum value in the sample.

The distribution function and density function of the r^{th} order statistic are given in (1.4.6) and (1.4.7), respectively.

$$F_r(x) = P(X_{(r)} \leq x) = \sum_{k=r}^{n} \binom{n}{r} F(x)^k \, (1 - F(x))^{n-k} \tag{1.4.6}$$

$$f_r(x) = \frac{1}{\beta(r, n-r+1)} \, f(x) \, F(x)^{r-1} \, (1 - F(x))^{n-r}. \tag{1.4.7}$$

Specifically, the distribution function of $X_{(1)}$ and $X_{(n)}$ are

$$F_1(x) = 1 - (1 - F(x))^n \quad \text{and} \quad F_n(x) = (F(x))^n.$$

The quantile function corresponding to the r^{th} order statistic is defined as

$$Q_{(r)}(u) = Q\left(I_u^{-1}(r, n-r+1)\right) \tag{1.4.8}$$

where, $I_u^{-1}(.,.)$ denotes the inverse of regularized incomplete beta function. The derivation of the r^{th} order statistic in the quantile terms is given in Nair et al. (2013a). The quantile functions of the order statistics $X_{(1)}$ and $X_{(n)}$, are given by

$$Q_{(1)}(u) = Q\left(1 - (1-u)^{\frac{1}{n}}\right) \text{ and } Q_{(n)}(u) = Q\left(u^{\frac{1}{n}}\right). \tag{1.4.9}$$

1.4.3 Moments

This section delves into moments based on quantile functions. The r^{th} traditional moment is defined as

$$\mu_r^{'} = E(X^r) = \int_0^{\infty} x^r f(x) \, dx. \tag{1.4.10}$$

By substituting $x = Q(u)$ in (1.4.10), we get moments in terms of the quantile function and is

$$\mu_r^{'} = \int_0^1 (Q(u))^r \, du. \tag{1.4.11}$$

The mean is obtained by putting $r = 1$ in (1.4.11) and it takes the form

$$\mu = \int_0^1 Q(u) \, du = \int_0^1 (1-u) \, q(u) \, du. \tag{1.4.12}$$

The r^{th} central moment in quantile form is

$$\mu_r = E(X-\mu)^r = \int_0^1 (Q(u)-\mu)^r \, du. \tag{1.4.13}$$

Using (1.4.13) additional moment-based measures such as spread, skewness, and kurtosis can be calculated in quantile terms.

1.4.4 L-moments

The expected values of linear functions of order statistics are known as the L-moments. The L-moments are an alternative to the classical moments. Although Hosking (1990) developed a thorough theory on L-moments, the groundwork for L-moments was built by the work on a linear combination of order statistics by Sillitto (1969) and Greenwood et al. (1979). These moments usually have lower sample variances and are resistant to outliers. L-moments are useful for summarizing probability distributions, identifying distributions, and fitting models to data.

The r^{th} L-moment is given as

$$L_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}), \quad r = 1, 2, \ldots. \tag{1.4.14}$$

We know that

$$\begin{aligned} E(X_{r:n}) &= \int x\, f_r(x)\, dx \\ &= \frac{n!}{(r-1)!(n-r)!} \int_0^1 u^{r-1}(1-u)^{n-r}\, Q(u)\, du. \end{aligned} \tag{1.4.15}$$

By substituting the value of $E(X_{r:n})$ in (1.4.14), we get

$$L_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \frac{r!}{k!(r-k-1)!} \int_0^1 u^{r-k-1}(1-u)^k\, Q(u)\, du.$$

Using the binomial theorem to expand $(1-u)^k$ and combining the powers of u will give the expression of L-moment as

$$L_r = \int_0^1 \sum_{k=0}^{r-1} (-1)^{r-1-k} \binom{r-1}{k} \binom{r-1+k}{k} u^k\, Q(u)\, du. \tag{1.4.16}$$

To establish the above relationship, Jones (2004) has provided a different approach. Specifically, the first four L-moments are

$$L_1 = \int_0^1 Q(u)\, du, \tag{1.4.17}$$

$$L_2 = \int_0^1 (2u-1)\, Q(u)\, du, \tag{1.4.18}$$

$$L_3 = \int_0^1 (6u^2 - 6u + 1)\, Q(u)\, du, \tag{1.4.19}$$

$$L_4 = \int_0^1 (20u^3 - 30u^2 + 12u - 1)\, Q(u)\, du. \tag{1.4.20}$$

Alternatively, we can use the quantile density function to compute the L-moments given in (1.4.17) - (1.4.20) and are given as

$$L_1 = \int_0^1 (1 - u)\, q(u)\, du, \tag{1.4.21}$$

$$L_2 = \int_0^1 (u - u^2)\, q(u)\, du, \tag{1.4.22}$$

$$L_3 = \int_0^1 (3u^2 - 2u^3 - u)\, q(u)\, du, \tag{1.4.23}$$

$$L_4 = \int_0^1 (u - 6u^2 + 10u^3 - 5u^4)\, q(u)\, du. \tag{1.4.24}$$

L-moments exist as long as the $E(X)$ is finite, whereas classical moments may need more specific conditions to be finite across various distributions. The L-moments $(L_r; r = 1, 2, \ldots)$ can be used to characterize a distribution whose mean exists. However, any collection containing every L-moment except one is insufficient to characterize a distribution.

Let X be a non-degenerate random variable with finite mean, then the ratios $\tau_r = \frac{L_r}{L_2}, r = 3, 4, \ldots$ are scale-free, dimensionless, and bounded. Also, the coefficients τ_r lie between -1 and 1. L-moment ratios can be used to determine a distribution's skewness and kurtosis. The L-coefficient of skewness and L-coefficient of kurtosis are defined in (1.4.25) and (1.4.26), respectively.

$$\tau_3 = \frac{L_3}{L_2}, \tag{1.4.25}$$

$$\tau_4 = \frac{L_4}{L_2}. \tag{1.4.26}$$

The coefficient τ_3 lies in $(-1, 1)$ and that of τ_4 is $\frac{1}{4}(5\,\tau_3^2 - 1) \leq \tau_4 < 1$. These results can be found in Hosking (1996) and Jones (2004). Similarly, the L-coefficient of variation is defined as

$$\tau_2 = \frac{L_2}{L_1}. \tag{1.4.27}$$

When the random variable X is non-negative, $L_1 \,\&\, L_2 > 0$ and $L_2 < L_1$, we get $0 < \tau_2 < 1$.

Due to their aforementioned characteristics, L-moments are widely used in many fields, including hydrology, civil engineering, and meteorology. They can also be utilized for parameter estimation by employing the technique of equating population characteristics to sample characteristics, similar to the moment method of estimation.

Consider a random sample $X_1, X_2, ..., X_n$ of size n from the population, and let $X_{(1)}, X_{(2)}, ..., X_{(n)}$ be the ordered sample observations. We have sample counterparts for population L-moments and the r^{th} sample L-moment is represented as

$$l_r = \sum_{k=0}^{r-1} p_{r-1,k}\; h_k\,, \tag{1.4.28}$$

where

$$p_{r-1,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}$$

and

$$h_k = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \frac{(i-1)(i-2)\dots(i-k)}{(n-1)(n-2)\dots(n-k)} X_{(i)}, \quad \text{for } k = 0, 1, 2, \dots, n-1.$$

Asymptotic properties of sample L-moments are proved in Hosking (1990).

1.5 Income inequality measures

The unequal distribution of income among individuals or groups within a society, region, or nation is referred to as income inequality. It highlights the gap between the rich and the poor and shows how unequally income is distributed among a population. This means that a small segment of the population holds a substantial portion of the total income or wealth, while a larger segment has significantly less. Income inequality is caused by several reasons such as variations in education, inheritance, wealth accumulation, resource accessibility, and governmental policy. Income disparity can negatively affect the economy, such as slower economic growth, higher rates of poverty, and more social unrest. Also, Dalton (1920) provided the requirements that should be satisfied by an efficient income inequality measure. In this section, we are discussing different methods to measure income inequality.

1.5.1 Lorenz curve

The work of Lorenz in 1905 revolutionized economic and statistical income distribution studies. Even today, the Lorenz curve is a fertile field of investigation into the bordering area between statistics and economics. Despite various inequality measures, the Lorenz curve remains a crucial tool for assessing income inequality. For a n-point distribution, by plotting the share $L(i/n)$ of total income obtained by the $(i/n) * 100\%$ of households with lower incomes, $i = 0, 1, 2, \ldots, n$, and interpolating linearly yields the Lorenz curve. The empirical (or discrete) Lorenz curve is thus defined in terms of $n + 1$ points as

$$L\left(\frac{i}{n}\right) = \frac{\sum_{j=1}^{i} x_{j:n}}{\sum_{j=1}^{n} x_{j:n}}, \quad i = 0, 1, 2, \ldots, n, \tag{1.5.1}$$

where $x_{j:n}$ represents the j^{th} smallest income.

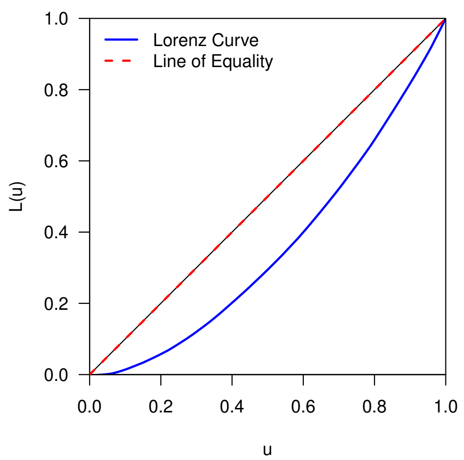


Figure 1.1: Empirical Lorenz curve

In Figure 1.1, a straight line is drawn from the origin $(0,0)$ to the point $(1,1)$, which is referred to as the 'line of equality' or 'egalitarian line'. The Lorenz curve typically bows below the line of equality. If the Lorenz curve coincides with this line it represents a fictitious situation in which income is equally distributed throughout the population. On the other hand, the amount of income inequality in the distribution is indicated by how much the Lorenz curve deviates from the line of equality.

The 'moment distribution form' of the Lorenz curve is given in (1.5.2) and is defined using first-moment distribution. It has advantages, particularly for parametric families where the quantile function cannot be represented using basic functions. Thus the Lorenz curve is given as

$$\{(u, L(u))\} = \{(u, w) \mid u = F(x),\, w = F_{(1)}(x)\,;\, x \geq 0\}. \tag{1.5.2}$$

Here $F_{(1)}(x)$ is the first-moment distribution and is

$$\begin{aligned} F_{(1)}(x) &= \frac{\int_0^x t f(t)\, dt}{E(X)} \\ &= \frac{\int_0^x t f(t)\, dt}{\int_0^\infty t f(t)\, dt}. \end{aligned} \tag{1.5.3}$$

given that $E(X)$ is finite. $F_{(1)}(x)$ indicates the proportional share of total income of individuals whose income is less than or equal to x.

The Lorenz curve $L(u)$ is a function defined on the interval $[0,1]$ for finite populations. For a fixed value of u, $L(u)$ indicates the proportion of the total income in the population that is accounted for by the bottom $100u\%$ of individuals in the population. A more accurate definition of the Lorenz curve using the quantile function was proposed by Gastwirth (1971). If X is a non-negative random variable with a finite positive mean, then the Lorenz curve is defined as

$$L(u) = \frac{1}{\mu}\int_0^u Q(p)\,dp\,, \tag{1.5.4}$$

where $\mu = \int_0^1 Q(p)\,dp\,.$ On differentiating (1.5.4), we get

$$Q(u) = \mu\, L^{'}(u).$$

The above equation makes it obvious that the Lorenz curve uniquely determines the distribution as the quantile function can characterize any distribution.

The following properties of the Lorenz curve are directly observable from the equation (1.5.4) and are

1. $L(u)$ is continuous in the interval $0 \leq u \leq 1$, with $L(0)$ equals 0 and $L(1)$ equals 1 .

2. $L(u)$ is increasing.

3. $L(u)$ is convex.

On the other hand, any function that satisfies these properties can be identified as the Lorenz curve of a statistical distribution (Thompson (1976)). Aaberge (2000) recommended and justified the use of some inequality measures for encapsulating the fundamental insights derived from the Lorenz curve.

There are two methods for generating Lorenz curve models. The first method starts with an income distribution function, from which the corresponding Lorenz curve is derived using the previously mentioned methods. The second one is to start from simple curves that meet the desired conditions for the Lorenz curve. Although this generally leads to cumbersome distribution functions, it may provide enough flexibility for empirical Lorenz curves. One can refer to Kleiber and Kotz (2003) for additional details on parametric families of Lorenz curves.

Lorenz ordering provides a method for assessing the inequality between two distributions. When the Lorenz curve of a distribution consistently remains above the other, the former is considered to Lorenz dominate the latter, indicating a lower level of inequality. Let $\mathcal{L}$ be the class of non-negative random variables with positive and finite mean. Hence, Lorenz ordering for two random variables $X, Y \in \mathcal{L}$ is defined as,

$$X \geq_L Y \iff F_X \geq_L F_Y \iff L_X(u) \leq L_Y(u) \tag{1.5.5}$$

for all $u \in [0, 1]$. In this case, X is said to be larger than Y in Lorenz's notion. When two Lorenz curves do not intersect the distribution with the lower curve may be referred to as 'more unequal'. Furthermore, the Lorenz ordering is scale-invariant. Lorenz dominance loses its significance in cases where Lorenz curves cross each other, thereby making it essential to employ the Gini index for evaluation.

We can use star-shaped ordering to get the Lorenz ordering when the quantile function of the distribution is available in closed form. If X and Y are random variables in $\mathcal{L}$ with quantile functions Q_X and Q_Y respectively, then Arnold (1987) determined that star-shaped implies Lorenz ordering and is defined as follows.

Definition 1.5.1. *If* $Q_X(u)/Q_Y(u)$ *is a non-increasing function of* u*, then* X *is star-shaped with respect to* Y*, and denoted as* $X \leqslant_* Y$*.*

1.5.2 Gini index

The Gini index is the most widely used measure of inequality. Gini (1914) developed the concept of the ratio of concentration, which evolved into what is now referred to as the Gini index. The initial way to express it is by calculating the ratio of the mean of absolute differences (a notion first put forward by Gini in 1912) to two times the mean. If X and Y are two random variables having the same distribution function F and finite mean μ then Gini as the mean of absolute differences is

$$G = \frac{1}{2\mu} \int_0^{\infty} \int_0^{\infty} |x - y| \, \mathrm{d}F(x) \, \mathrm{d}F(y) \tag{1.5.6}$$

Alternatively, the Gini index is defined as twice the area between the Lorenz curve and the egalitarian line. i.e.,

$$\begin{aligned} G &= 2 \int_0^1 (u - L(u)) du \\ &= 1 - 2 \int_0^1 L(u) du. \end{aligned} \tag{1.5.7}$$

It can also be interpreted as the ratio of the area between the line of equality and the Lorenz curve to the area under the line of equality. The Gini index ranges from 0 to 1, where $G = 0$ denotes perfect equality and $G = 1$ denotes perfect inequality. The Gini index is also known as the Gini coefficient, Lorenz concentration ratio, and coefficient of concentration.

In literature, there are several representations for the Gini index however an equation in terms of expected values of order statistics is

$$\begin{aligned} G &= 1 - \frac{E(X_{1:2})}{E(X)} \\ &= 1 - \frac{\int_0^\infty [1 - F(x)]^2 \, dx}{E(X)}. \end{aligned} \tag{1.5.8}$$

A direct generalization of the above equation is given in Muliere and Scarsini (1989) and is

$$G_\delta = 1 - \frac{E(X_{1:\delta})}{E(X)}. \tag{1.5.9}$$

The generalized Gini index, which is sensitive to both high and low incomes, was established by Kakwani (1980b), Donaldson and Weymark (1980), Donaldson and Weymark (1983), and Yitzhaki (1983). It can be calculated using (1.5.10)

$$G_\delta = 1 - \delta(\delta - 1) \int_0^1 L(u)(1 - u)^{\delta - 2} du, \tag{1.5.10}$$

where $\delta > 1$ is the inequality aversion parameter. More specifically, the upper-income percentiles are given more weight when $\delta \in (0, 1)$ and when $\delta > 2$ the index focuses on the lower-income percentiles. When $\delta = 2$, the index is the same as the traditional Gini index.

The truncated Gini index is discussed in the studies of Takayama (1979) and Ord et al. (1983). However, Nair et al. (2008) provides a detailed study of the truncated Gini index for the poor and the rich separately. Let u and $(1 - u^*)$ denote the proportion of poor and rich in the population, respectively. Now in terms of the quantile function, the truncated Gini index for the poor and the rich is given as

$$G(u) = \frac{2\,(\mu(u))^{-1}}{u^2} \int_0^u p\,Q(p)\,dp - 1, \tag{1.5.11}$$

$$G^*(u^*) = 1 - \frac{2\,(\mu^*(u^*))^{-1}}{(1-u^*)^2} \int_{u^*}^1 (1-p)\,Q(p)\,dp, \tag{1.5.12}$$

where

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

and

$$\mu^*(u^*) = \frac{1}{1-u^*} \int_{u^*}^1 Q(p)\,dp.$$

1.5.3 Pietra index

Pietra index is another significant measure of income inequality obtained from the Lorenz curve, proposed by Pietra (1932). It is also called the Schutz index (Duclos and Araar (2006)). In geometrical terms, the Pietra index is defined as the maximal deviation between the Lorenz curve and the egalitarian line. i.e., $P =$

$\max_{0 \le u \le 1} \{u - L(u)\}$. Alternatively, it can also be described as twice the area of the biggest triangle that can be drawn within the space between the Lorenz curve and the egalitarian line.

An alternate expression in terms of the distribution function for relative mean deviation and the Pietra index is

$$\tau = \frac{E\,|X - E(X)|}{E(X)}, \tag{1.5.13}$$

$$P = \frac{E\,|X - E(X)|}{2\,E(X)}. \tag{1.5.14}$$

In quantile terms, the formula for relative mean deviation and the Pietra index is given as

$$\tau = \frac{\int_0^1 |Q(u) - Q(u_0)|du}{\mu}, \tag{1.5.15}$$

$$P = \frac{\int_0^1 |Q(u) - Q(u_0)|du}{2\mu}, \tag{1.5.16}$$

where μ is the mean of the corresponding distribution. Also, $\mu = Q(u_0)$ for some $0 < u_0 < 1$. We can determine u_0 by solving for u in the equation $\mu = Q(u)$.

The P ranges from 0 to 1. The minimum value of P is attained when there is complete income equality, which means the Lorenz curve aligns with the line of equality. The maximum value of P occurs when the Lorenz curve approaches the bottom right corner. These conditions are clear from Figure 1.1. The Pietra index represents the portion of the income that needs to be reallocated from wealthier individuals to poorer ones to achieve total income equality. The value of P indicates the proportion of total income that needs to be reallocated from households above the mean to those below the mean to achieve an equal distribution of incomes. Higher P values signal greater income inequality, necessitating more extensive redistribution to attain income equality. For this reason, the index is also referred to as the Robin Hood index.

1.5.4 Bonferroni curve and Bonferroni index

The Bonferroni curve is an income inequality measure based on the first moment distribution proposed by Bonferroni (1930). It is useful in situations where income inequality is largely due to the existence of individuals or groups with incomes significantly lower than the majority. Assume that X is a non-negative, absolutely continuous random variable with distribution function $F(x)$ and having a finite non-zero mean μ. Now the first-moment distribution and the partial mean μ_x of the probability distribution are given in (1.5.3) and (1.5.17) respectively.

$$\begin{aligned} \mu_x &= \frac{\int_0^x t f(t)\, dt}{\int_0^x f(t)\, dt} \\ &= \frac{\mu\, F_{(1)}(x)}{F(x)}. \end{aligned} \tag{1.5.17}$$

The Bonferroni measure is given as

$$B(F(x)) = \frac{\mu_x}{\mu}$$

$$= \frac{F_{(1)}(x)}{F(x)}. \tag{1.5.18}$$

The Bonferroni curve is represented in the orthogonal plane $(F(x), B(F(x)))$ within a unit square. The parametric expression for the Bonferroni curve is obtained by putting $u = F(x)$ in the above equation and is given as

$$\begin{aligned} B(u) &= \frac{F_{(1)}(F^{-1}(u))}{u} \\ &= \frac{1}{u\,\mu}\int_0^u F^{-1}(p)\,dp \\ &= \frac{1}{u\,\mu}\int_0^u Q(p)\,dp, \end{aligned} \tag{1.5.19}$$

where $u \in (0, 1]$. Bonferroni curve and the Lorenz curve are related through the equation

$$B(u) = \frac{L(u)}{u}. \tag{1.5.20}$$

The line that joins the points $(0, 1)$ and $(1, 1)$ in Figure 1.2 is referred to as the line of equality of the Bonferroni curve. The area bounded by the Bonferroni curve, the egalitarian line, and the y-axis in Figure 1.2 is known as the Bonferroni index, and it is determined by

$$B = 1 - \int_0^1 B(u)\,du. \tag{1.5.21}$$

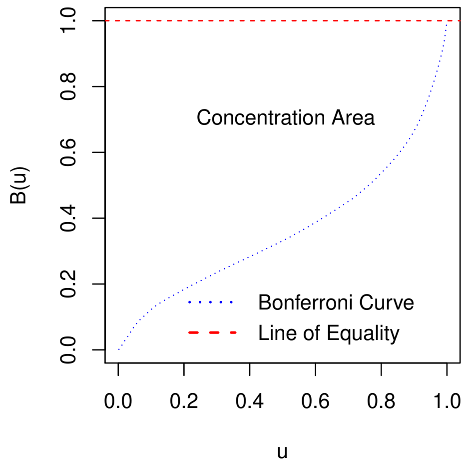


Figure 1.2: Empirical Bonferroni curve

The sensitivity of the Bonferroni curve to lower income levels is a feature that makes it useful for measuring poverty (Giorgi and Crescenzi (2001)). For further reading on the Bonferroni curve, its statistical inference, and the methodological advancements in the reliability framework, refer to Pundir et al. (2005).

1.5.5 Atkinson index

Atkinson (1970) proposed an inequality measure based on a social welfare function known as the Atkinson index or Atkinson measure. This index lies between 0 and 1 and helps to identify which end of the distribution contributes more significantly to the observed inequality. It is defined as

$$A_\epsilon = 1 - \frac{1}{E(X)}\left[\int_0^\infty x^{1-\epsilon} dF(x)\right]^{\frac{1}{1-\epsilon}}, \tag{1.5.22}$$

where $\epsilon > 0$ is a sensitivity parameter that assigns greater weight to lower incomes as it increases. ϵ is also known as the inequality aversion parameter. As ϵ approaches 1, the Atkinson index becomes more sensitive to variations at the lower end of the income distribution. On the other hand, as ϵ approaches 0, the Atkinson index becomes more receptive to changes in the upper end of the income distribution.

In quantile terms, the Atkinson index is defined as

$$A_\epsilon = 1 - \frac{1}{\int_0^1 Q(u)du} \left[\int_0^1 Q(u)^{1-\epsilon} du \right]^{\frac{1}{1-\epsilon}}. \tag{1.5.23}$$

1.5.6 Generalized entropy measures

The generalized entropy measures (Cowell and Kuga (1981)) represent another approach for measuring income inequality and are defined as follows

$$GE_\theta = \frac{1}{\theta(\theta - 1)} \int_0^\infty \left[\left(\frac{x}{E(X)} \right)^\theta - 1 \right] dF(x), \tag{1.5.24}$$

where θ is a sensitivity parameter, $\theta \in \mathbb{R}$ and $\theta \neq 0 \,\&\, 1$. Here the parameter θ focuses the lower tail for values of $\theta < 0$ and the upper tail for values of $\theta > 0$. Also, we obtain the squared coefficient of variation for $\theta = 2$.

In quantile terms, the generalized entropy measure is given as

$$GE_\theta = \frac{1}{\theta(\theta - 1)} \int_0^1 \left[\left(\frac{Q(u)}{\int_0^1 Q(u)du} \right)^\theta - 1 \right] du. \tag{1.5.25}$$

When $\theta \longrightarrow 0, 1$ the generalized entropy measures given in (1.5.24) takes the form

$$T_1 = GE_1 = \int_0^\infty \frac{x}{E(X)} \log\left(\frac{x}{E(X)}\right) dF(x), \tag{1.5.26}$$

$$T_2 = GE_0 = \int_0^\infty \log\left(\frac{E(X)}{x}\right) dF(x). \tag{1.5.27}$$

These indices are known as Theil coefficients (Theil (1967)). T_1 is commonly known as the Theil coefficient, whereas T_2 is the mean logarithmic deviation or Theil's second measure.

The Theil coefficients are represented in quantile terms as follows

$$T_1 = \int_0^1 \frac{Q(u)}{Q(u_0)} \log \frac{Q(u)}{Q(u_0)}\, du, \tag{1.5.28}$$

$$T_2 = \int_0^1 \log \frac{Q(u_0)}{Q(u)}\, du. \tag{1.5.29}$$

1.5.7 Frigyes measures

Instead of one, Frigyes (1965) introduced three measures with clear economic interpretation and are

$$\rho = \frac{m}{m_1},\ \nu = \frac{m_2}{m_1},\ \eta = \frac{m_2}{m}, \tag{1.5.30}$$

where $m = E(X)$, $m_1 = E(X|X < m)$, $m_2 = E(X|X \geqslant m)$. The measure ν evaluates the inequality level across the entire income distribution range. On the other hand, the measures ρ and η reflect inequality in the two regions of the distribution below and above the mean, respectively. The measures ρ, ν, and η take values in the interval $(1, \infty)$, however, they are easily convertible to the standard range of $(0, 1)$. The properties and uses of these three measures have been covered by Eltetö and Frigyes (1968).

In quantile terms, these three measures can be expressed as

$$\rho = \frac{u_0\, Q(u_0)}{\int_0^{u_0} Q(u)du}, \tag{1.5.31}$$

$$\nu = \frac{u_0}{(1-u_0)} \frac{\int_{u_0}^1 Q(u)du}{\int_0^{u_0} Q(u)du}, \tag{1.5.32}$$

$$\eta = \frac{\int_{u_0}^1 Q(u)du}{(1-u_0)\, Q(u_0)}. \tag{1.5.33}$$

1.5.8 Zenga curve

The Zenga curve proposed by Zenga (1984) is formulated using the first-moment distribution, hence, it requires the expected value of X to be finite. For $0 < u < 1$ the Zenga curve is defined as follows

$$\begin{aligned} Z(u) &= \frac{F_{(1)}^{-1}(u) - F^{-1}(u)}{F_{(1)}^{-1}(u)}, \\ &= 1 - \frac{F^{-1}(u)}{F_{(1)}^{-1}(u)}. \end{aligned} \tag{1.5.34}$$

Also, the Zenga concentration curve is the set $\{(u, Z(u)); u \in (0, 1)\}$. More information regarding the Zenga curve and associated index are available in Kleiber and Kotz (2003).

A new and more practical inequality measure based on the conditional expectations of the relevant distribution was put forward by Zenga (2007). Let X be a continuous and non-negative random variable having a finite positive mean μ, with a distribution function $F(x)$ and a density function $f(x)$ that is strictly positive on the interval $0 \leq a < b \leq \infty$, then Zenga curve is given as

$$A(x) = 1 - \frac{\mu^-(x)}{\mu^+(x)}, \tag{1.5.35}$$

where $\mu^-(x) = E(X|X \leq x)$ and $\mu^+(x) = E(X|X > x)$. The key feature of this measure of inequality is its approach of comparing segments of the population, where the segments being compared are always two mutually exclusive and adjacent groups. Moreover, the comparison focuses on the ratio between the arithmetic mean of these two groups. The corresponding inequality index I is given as

$$I = \int_a^b A(x) f(x)\, dx \tag{1.5.36}$$

The Zenga curve in (1.5.35) can be expressed in quantile terms as follows

$$\begin{aligned} I(u) &= A(Q(u)) \\ &= 1 - \frac{(1-u)}{u} \frac{\int_0^u Q(p)\, dp}{\int_u^1 Q(p)\, dp} \\ &= 1 - \frac{(1-u)}{u} \frac{\int_0^u Q(p)\, dp}{\mu - \int_0^u Q(p)\, dp}. \end{aligned} \tag{1.5.37}$$

The inequality index I is defined as the area beneath the curve $I(u)$ and is represented as

$$I = \int_0^1 I(u)\, du. \tag{1.5.38}$$

The relationship between the Zenga curve ($I(u)$) and the Lorenz curve is

$$I(u) = \frac{u - L(u)}{u(1 - L(u))}. \tag{1.5.39}$$

For more details on the characteristics of the Zenga measure in terms of quantiles, refer to Sreelakshmi and Nair (2014). Throughout this thesis, we have used the Zenga curve, proposed in 2007.

1.5.9 Leimkuhler curve

The Leimkuhler curve (Burrell (1991, 2005), Rousseau (1987)) in the field of informetrics and the Lorenz curve in economics are closely related. They illustrate the relationship between the cumulative proportion of total productivity versus the cumulative proportion of resources. The key difference between the construction of these two curves is that the Lorenz curve arranges the sources in ascending order of productivity, whereas the Leimkuhler curve arranges the sources in descending order.

Let $\mathcal{L}$ denote the class of non-negative random variables with a positive finite mean, and let $X \in \mathcal{L}$ have a distribution function $F(x)$, and a quantile function $Q(u)$. Then the Leimkuhler curve $K(u)$ associated with random variable X is given as

$$K(u) = \frac{1}{\mu}\int_{1-u}^{1} Q(p)\,dp, \quad 0 \le u \le 1. \tag{1.5.40}$$

The Leimkuhler curve defined in (1.5.40) satisfies the following properties.

1. $K(u)$ is continuous in the interval $0 \le u \le 1$.
2. $K(u)$ is differentiable almost everywhere within the interval $[0, 1]$.
3. $K(0)$ is equal to 0 and $K(1)$ is equal to 1.
4. $K(u)$ is non-decreasing function.
5. $K(u)$ is concave function.

The Leimkuhler curve is related to the Lorenz curve through the equation given below

$$K(u) = 1 - L(1-u). \tag{1.5.41}$$

Nair and Vineshkumar (2022) utilized quantile functions to develop new properties of Leimkuhler curves and studied applications in the field of informetrics. Also, see Sarabia (2008), Sarabia et al. (2010), Sarabia and Sarabia (2008) for further details on the Leimkuhler curve.

1.6 Other inequality measures

The Kolkata index, or the k-index proposed by Ghosh et al. (2014) is defined as the solution to the equation

$$k_F + L(k_F) = 1, \tag{1.6.1}$$

where $L(.)$ is the Lorenz function. k-index is an income inequality index that takes values in the interval $[1/2, 1]$, which makes it different from other inequality measures like the Gini coefficient or the Pietra index. It is also used as a summary measure. A simple normalization of the k-index is given by

$$K_F = 2\,k_F - 1. \tag{1.6.2}$$

This normalized k-index K_F ranges from 0 to 1, with $K_F = 0$ denoting complete equality and $K_F = 1$ denoting complete inequality. Refer to the works of Chatterjee et al. (2017), Banerjee et al. (2020a), and Banerjee et al. (2020b) for more information on the Kolkata index.

A broad category of inequality measures denoted by $\mathcal{T}_h(X)$, where h is a convex function in the interval $(0, \infty)$ was examined by Ord et al. (1981). It is defined as

$$\mathcal{T}_h(X) = E\left[\frac{h(X)}{E(X)}\right]. \tag{1.6.3}$$

Specifically, when $h(x) = \frac{(x^{\lambda+1}-1)}{\lambda(\lambda+1)}$ and $\lambda = -2, -1, 0,$ and 1, these correspond to various indices: the arithmetic mean to the harmonic mean ratio, the arithmetic mean to the geometric mean ratio, the Theil index, and the Herfindahl index (which is the squared coefficient of variation), respectively.

Additionally, Ord et al. (1981) proposed the entropy measures which function as inequality measures and is given as

$$e_\lambda(X) = \frac{1}{\lambda}\int_0^\infty f(x)(1 - f^\lambda(x))dx, \quad -1 < \lambda < \infty. \tag{1.6.4}$$

Analogously, the quantile form of the above equation is

$$e_\lambda(u) = \frac{1}{\lambda}\int_0^1 (1 - q^{-\lambda}(u))du, \quad -1 < \lambda < \infty. \tag{1.6.5}$$

Using a linear averaging technique, Mehran (1976) introduced a class of linear income inequality measures by

$$\Phi = \frac{1}{\mu}\int_0^1 (Q(u) - \mu)\, W(u)\, du, \tag{1.6.6}$$

where $W(u)$ represents a score function that is assumed to be independent of the shape of distribution with $\int_0^1 W(u)du = 0$. Each score function provides a specific linear inequality measure.

1.7 Poverty measures

Poverty measures are quantitative tools used to evaluate and describe the economic status of individuals or populations, determining whether they live in poverty and the extent of their deprivation. These measures are crucial for understanding the magnitude and nature of poverty, designing targeted interventions, and monitoring progress towards poverty alleviation. Let $D(z, y)$ be the function that represents the degree of deprivation faced by an individual whose income y is below the poverty line z.

Thus

$$\begin{aligned} P &= E_y \left[D\left(z, y\right) I\left(y < z\right) \right] \\ &= \int_0^z D\left(z, y\right) f(y) dy, \end{aligned} \tag{1.7.1}$$

where $I\left(y < z\right)$ denotes an indicator function that takes value 1 when $y < z$ and 0 in other cases, and $f(y)$ denotes the probability density function. To get a comprehensive understanding of poverty measurements, refer to Kakwani (1980a), Duclos and Araar (2006), and Morduch (2008). Furthermore, Chotikapanich et al. (2013) derived poverty measures based on the generalized beta distribution and analyzed shifts in poverty levels across countries in south and southeast Asia.

1.7.1 Headcount ratio

The most basic and popular measure to quantify poverty is the headcount ratio. It indicates the proportion of a population whose income is below the poverty line. This measure tracks the most direct aspect of the human scale of poverty by literally counting heads, which is useful for researchers and policymakers. The headcount ratio H stands for the proportion of the population living in poverty and is defined as

$$H = \frac{N_p}{N}, \tag{1.7.2}$$

where N_p denotes the number of individuals below the poverty line and N denotes total population (or sample) size.

Headcount ratio can also be represented by setting $D(z, y) = 1$ in (1.7.1) and is

$$H = F(z). \tag{1.7.3}$$

When we set $D(z, y) = 1$, we give every individual whose income below the poverty line the same weight, regardless of whether their income is marginally below or significantly below the poverty line. Therefore, the headcount ratio provides an estimate of the prevalence of poverty but fails to consider the severity of poverty faced by the poor.

1.7.2 Poverty gap ratio

The poverty gap ratio is a straightforward extension among the several alternatives to the headcount ratio. It considers both the proportion of the poor population and the severity of poverty among the poor. The poverty gap ratio measures the shortfall in income for each relative to the poverty line. By setting $D(z, y) = \left(\frac{z-y}{z}\right)$ in (1.7.1), we can derive poverty gap ratio. Thus

$$\begin{aligned} PG &= \int_0^z D(z, y) f(y) dy \\ &= \int_0^z \left(\frac{z-y}{z}\right) f(y) dy. \end{aligned} \tag{1.7.4}$$

1.7.3 Foster-Greer-Thorbecke measure

The Foster-Greer-Thorbecke (FGT) measure suggested by Foster et al. (1984) is a generalization of the poverty gap ratio. It is obtained by setting $D(z, y) = \left(\frac{z-y}{z}\right)^\alpha$ in (1.7.1) and is given as

$$FGT(\alpha) = \int_0^z \left(\frac{z-y}{z}\right)^\alpha f(y) dy, \tag{1.7.5}$$

where $\alpha \geq 0$. The parameter α influences how responsive the measure is to the intensity of poverty among those living below the poverty line. The FGT measure reduces to the headcount ratio when $\alpha = 0$ and reduces to the poverty gap ratio when $\alpha = 1$. Also, when $\alpha = 2$, it is commonly known as the squared poverty gap.

1.7.4 Watts index

Watts index proposed by Watts (1968) is the first distribution-sensitive poverty measure. It satisfies the poverty axioms like focus, monotonicity, transfer, and decomposability. Zheng (1993) identified a large set of axiomatic properties for the Watts index. Watts index is obtained by setting $D(z,y) = (\ln z - \ln y)$ in (1.7.1) and is given as

$$W = \int_0^z (\ln z - \ln y) f(y) dy. \tag{1.7.6}$$

In quantile terms, the Watts index can be written as

$$W = \int_0^u \ln\left(\frac{Q(u)}{Q(p)}\right) dp. \tag{1.7.7}$$

A measure introduced by Kakwani (1999) which is closely connected to the Watts index is $K^* = 1 - e^{-W}$.

1.7.5 Sen index

Sen (1976) introduced an index that aims to integrate the impact of the number of poor individuals, the severity of their poverty, and the distribution of poverty within the group. Sen index weights the poverty gap according to a person's position among the impoverished. It is defined as

$$S = 2\int_0^z \left(\frac{z-y}{z}\right)\left(\frac{F(z)-F(y)}{F(z)}\right) f(y) dy. \tag{1.7.8}$$

The Sen index is represented in quantile terms as follows

$$S(u) = u\left(\frac{u\,\omega_1^{'}(u) + \omega_2(u)}{u\,\omega_1^{'}(u) + \omega_1(u)}\right), \tag{1.7.9}$$

where

$$\omega_1(u) = \frac{1}{u}\int_0^u Q(p)dp,$$

and

$$\omega_2(u) = \frac{1}{u^2}\int_0^u (2p-u)\,Q(p)dp.$$

Here $\omega_1^{'}(u)$ denotes derivative of ω_1 with respect to u.

1.8 Objectives

The main objectives of this study are as follows:

- To utilize the advantages of quantile functions in income modeling.
- To analyze income distributions from the literature, with a particular focus on those with explicit quantile functions.
- To explore other distributions with explicit quantile functions and evaluate their potential for income modeling. Additionally, to derive income inequality measures for all identified distributions.
- To introduce a distribution based on quantile functions for modeling income.

Subsequently, to derive income inequality and poverty measures for the proposed model.

- To model the per capita household income data of Indian states using parametric models with closed-form quantile functions and to compute income inequality measures for each state.

1.9 Outline of the thesis

The thesis includes six chapters as outlined below.

Chapter 1 introduces quantile functions and their fundamental concepts. It also provides an overview of income inequality and poverty measures.

Chapter 2 contains a detailed literature review on income distributions, income inequality measures, and quantile functions. The utility of quantile functions for modeling reliability and income data is discussed along with distribution functions. Hence in this chapter, we explore some quantile functions mentioned in the literature and their effectiveness in modeling income data. We derive the Lorenz curve, Gini index, Pietra index, Bonferroni curve, Bonferroni index, and Zenga curve for five different distributions and evaluate the model adequacy of three distributions by applying them to real income data.

We conduct a comprehensive quantile-based income study of the Power-Pareto (PP) distribution in Chapter 3. We derive significant income inequality measures and analyze the Lorenz ordering associated with the PP distribution. Through a simulation study, we assess the efficacy of three estimation techniques, finding that the L-moments method performs better than the other two. The model is applied to two real income datasets, and we compute both empirical and estimated income inequality indices.

In Chapter 4, we propose the Singh Maddala Dagum (SMD) distribution, defined as the sum of the quantile functions of the Singh-Maddala (SM) and Dagum distributions. We derive its distributional properties, income inequality measures, and poverty measures. The poverty gap ratio and FGT measures are transformed into quantile form. The model is applied to a real dataset, and the parameters of the proposed distribution are estimated using the L-moments method.

In Chapter 5, six parametric models with closed-form quantile functions are employed to model per capita household income data across Indian states, using data from the India Human Development Survey-II (IHDS-II). The models analyzed are the Weibull, PP, SM, Dagum, SMD, and modified lambda family (MLF). Parameter estimation for these models in each state is carried out using the L-moments technique. We used the Q-Q plot, chi-square test, histogram, and correlation coefficient to identify the most suitable model for each state. Subsequently, for every state, we derive both empirical and theoretical measures of income inequality, including the Gini, Pietra, Atkinson, generalized entropy, Bonferroni, and Frigyes measures. The theoretical values are obtained from the best-fitting model for each state.

In Chapter 6, the key findings of the thesis are summarized, and recommendations are presented. This is followed by an extensive bibliography that includes all the referenced sources.

CHAPTER 2

A STUDY ON QUANTILE FUNCTIONS, INCOME DISTRIBUTIONS, AND MEASURES OF INCOME INEQUALITY

2.1 Introduction

[1] The Pareto model developed by Vilfredo Pareto in 1897 is credited for sparking the beginning of studies on income distribution. Following this, numerous models like lognormal, Dagum, SM, and others emerged in the literature for analyzing income data. Additionally, the literature features various measures of income inequality. Kakwani (1980a) conducted an extensive examination of income distribu-

[1] A part of this chapter has been published as an article titled "A Review on Quantile Functions, Income Distributions, and Income Inequality Measures" in the journal " Reliability: Theory & Applications" (See Ashlin and Haritha (2023)).

tions, inequalities in income, the impact of government policies on personal income distributions, and the assessment of poverty. Dagum (1980) presented the concept of the economic distance ratio, a tool for assessing the extent of income inequality between two groups. For a comprehensive study of income distributions and measures of income inequality from the beginning, refer to Haritha and Nair (2005). This chapter aims to consolidate all the discussions on income distributions and measures of income inequality found in the literature.

Analyzing probability distribution in practical situations can be achieved via two methodologies: one involves defining the distribution function, while the other uses the quantile function. Gilchrist (2000) offered an extensive discussion on the quantile function and model-building techniques. Most income distribution studies have utilized the distribution function method, with a minimal number adopting the quantile function method (Tarsitano (2004), Haritha et al. (2007), Sreelakshmi and Nair (2014)). Hence in this study, we have explored different quantile functions as potential models for income.

Following the introduction, in Section 2.2, we examine various income distributions that feature distribution functions in closed form. In Section 2.3, we explore certain quantile functions suitable as models for income. Section 2.4 focuses on quantile functions with applications in the field of reliability. The discussion in Section 2.5 centers on income inequality measures found in the literature, and we derive inequality measures for five distributions that have quantile functions in closed form. The estimation and real data analysis are performed in Section 2.6, and the summary of the study is given in Section 2.7.

2.2 Income distributions

Huang and Oluyede (2014) proposed the exponential Kumaraswamy-Dagum (EKD) distribution for lifetime and income data analysis. Its cumulative distribution function is

$$F(x) = \left\{1 - \left[1 - \left(1 + \lambda x^{-\delta}\right)^{-\alpha}\right]^{\phi}\right\}^{\theta}, \quad x > 0, \quad \alpha,\, \lambda,\, \delta,\, \phi,\, \theta > 0. \tag{2.2.1}$$

The fundamental statistical characteristics, including moments, mean and median deviations, reliability measures such as hazard functions, kth order statistics, Renyi entropy, and income inequality measures like Lorenz and Bonferroni curves, are determined. The parameters of the EKD are estimated using maximum likelihood estimation, and the model is applied to a real dataset.

Clementi et al. (2016) provided a survey of findings associated with the κ-generalized distribution, a statistical model used to describe the distribution of income and wealth. They explored the κ-generalized model (Clementi et al. (2008, 2007, 2009)) for income distribution, along with a mixed κ-generalized model for wealth distribution, and the extended versions of the κ-generalized distributions of both first and second kinds. They discussed the fundamental characteristics of these distributions, and their connections with other distribution models, and derived the income inequality measures like the Lorenz curve and Gini index. Their findings suggest that κ-generalized models provide a good fit for the distribution of income and wealth. Subsequently, Clementi (2023) reviewed and compiled all the results on the κ-generalized distribution into a single source.

The Exponential Arc Tan (EAT) model and the composite EAT–Lognormal model, proposed by Calderín-Ojeda et al. (2016), are two extensions of the exponential distribution. These models can describe income distributions, including cases with zero income. Their effectiveness is evaluated using Australian income data from 2001 to 2012, and the results showed that these models provide a better fit for the data when compared to the exponential and gamma distributions.

Bourguignon et al. (2016) introduced and analyzed a new Pareto-type (NP) distribution, which serves as a generalization of the Pareto distribution and is defined by

$$F(x) = 1 - \frac{2\,\beta^{\alpha}}{x^{\alpha} + \beta^{\alpha}}, \quad x \geq \beta > 0,\, \alpha > 0. \tag{2.2.2}$$

Later on Abd Raof et al. (2022) employed the NP distribution to model the income of the Malaysian upper-class group. Their analysis reveals that both the Pareto type 1 and the NP models fit the top income data well across all years considered. However, the NP model offers greater flexibility, covering a broader range of incomes in the upper tail of the distribution compared to the Pareto type 1 model. They also measured income inequality among Malaysian top earners using the Lorenz curve, Gini, and Theil indices based on the fitted NP model.

Mirzaei et al. (2018) introduced the odd Weibull (OW) distribution and examined its applicability in analyzing income inequality. Its CDF is

$$F(x) = \frac{\left[1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}\right]^{\nu}}{\left[1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}\right]^{\nu} + \left[e^{-\left(\frac{x}{\beta}\right)^{\alpha}}\right]^{\nu}}, \quad x > 0, \quad \alpha, \beta, \nu > 0. \tag{2.2.3}$$

They investigated various mathematical and statistical properties of the OW distribution, deriving expressions for the hazard function, quantile function, first incomplete moment, and several measures of income inequality. The study also explores Lorenz and Zenga ordering based on the OW distribution. Parameter estimation is carried out using moments and maximum likelihood approaches. Finally, the effectiveness of the OW distribution is demonstrated by fitting it to real income datasets.

The suitability of the log Student's t model for fitting and describing income distributions across various European Union countries during three different economic phases: 2006, 2011, and 2016 was done by Barroso et al. (2020). The study shows that, in most cases, the log Student's t distribution provides the best fit.

The article by Bakar and Dharini (2020) introduced six mixture models based on the Weibull distribution, namely, Weibull-Paralogistic, Weibull-Weibull, Weibull-Fisk, Weibull-Gamma, Weibull-Burr, and Weibull-Dagum, for analyzing income data. The maximum likelihood method is used to estimate the parameters of Weibull mixtures and examine their performance based on average income per tax unit data from ten countries through information-based criteria approaches and graphical observations. Furthermore, they present the application of these models to inequality measures such as the Gini, Bonferroni, and generalized entropy index. They also derived analytical expressions for poverty measures, including the head-count ratio and poverty-gap ratio.

Safari et al. (2020) conducted a comparative analysis of the income distribution of low-income households in Malaysia. The study examined several models, including the reverse Pareto (Kleiber and Kotz (2003), Safari et al. (2019)), shifted reverse exponential, shifted reverse stretched exponential (Brzezinski (2015)), and shifted reverse lognormal. For each of the four models, they obtained expressions for the Lorenz curve and Gini index. Based on the outcomes of the Kolmogorov-Smirnov (KS) test and the R^2 coefficient, it was concluded that the reverse Pareto distribution provides an adequate representation of income distribution in low-income households. Also, the Lorenz curve and Gini index, derived from the reverse Pareto models, indicate that income is distributed evenly among poor households. The reverse Pareto distribution is often termed the inverse Pareto or power function distribution. On the other hand, Safari et al. (2021a) observed that the best model to explain Malaysian urban household incomes is the generalized beta 2 (GB2) distribution.

The review by Hlasny (2021) encompassed historical evidence, empirical characteristics, associations among distribution functions, and the process of selecting models for top-income distributions. He concluded that the generalized Pareto and the generalized beta family, which consists of the Dagum, SM, and GB2 distributions are consistently successful in income modeling.

Trzcińska (2022) used the Zenga model (Zenga (2010)) for analyzing income distributions within the Czech Republic and Poland. The study shows that the Zenga distribution is effective in modeling income distribution. It demonstrates its applicability in both the examination of income inequality and the comparison of income distributions across the Czech Republic and Poland.

Majid et al. (2023) introduced a three-part composite Pareto model to analyze income data in Malaysia. This model segments the population into three distinct groups, each characterized by a different distribution model. Specifically, the lower segment is modeled using the inverse Pareto distribution, the upper segment with the Pareto distribution, and the middle segment is described by a different, unspecified distribution model. Applying a Gaussian mixture distribution for analyzing income data is recommended for the middle segment, resulting in the inverse Pareto-Gaussian mixture-Pareto distribution model.

The four-parameter logistic Burr XII (LBXII) distribution is introduced by Guerra et al. (2023) and its CDF is given as

$$F(x) = \left\{ 1 + \left[\alpha \log \left(1 + \left(\frac{x}{\beta} \right)^{\delta} \right) \right]^{-\lambda} \right\}^{-1}, \quad x > 0, \quad \alpha, \beta, \delta, \lambda > 0. \tag{2.2.4}$$

They demonstrated that the distribution can exhibit decreasing and upside-down-bathtub hazard functions. It is shown that the density function of the LBXII distribution consists of an infinite linear combination of Burr XII density functions. The mathematical characteristics of the LBXII distribution, including ordinary and incomplete moments, the generating function, and the quantile function, are explored. The model's parameters are estimated using the maximum likelihood method, and a simulation study is conducted via Monte Carlo experiments. Additionally, two practical examples are provided to showcase the effectiveness of the LBXII distribution in modeling income data.

2.3 Income distributions based on quantile functions

Sodomova et al. (2005) conducted the statistical analysis and modeling on the annual net income of 1,566 households in the Slovak Republic for the year 2002. The results indicate that the Weibull distribution, with parameters determined through maximum likelihood estimation, offered the best fit across all income intervals except for those at the extreme lower and upper ends. Two distributions from the Weibull-Pareto class, derived using quantile properties, provide a good fit for the lowest and highest income intervals.

Hankin and Lee (2006) developed and studied the PP distribution, defined by the quantile function

$$Q(u) = \frac{C\,u^{\lambda_1}}{(1-u)^{\lambda_2}}, \quad 0 \le u \le 1, \quad C,\, \lambda_1,\, \lambda_2 > 0. \tag{2.3.1}$$

It is also known as the 'Hankin and Lee' distribution and the 'Davies' distribution.

They analyzed the characteristics, shape, and connections with other distributions and derived the moments of the PP distribution. The maximum likelihood and log regression methods are utilized to estimate the parameters of this distribution and assess the efficiency of these two estimation techniques through simulations. They found that the regression method is suitable for small sample sizes and when parameters are nearly equal, whereas the maximum likelihood method is preferable in other cases. Moreover, they applied this distribution to model the release of toxic gases. The quantile function of the PP distribution is obtained by combining the quantile functions from both the Pareto and Power distributions, indicating its potential usefulness in modeling income distributions.

Haritha et al. (2007) explored the MLF, which is defined through the quantile function

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2}\left[\frac{u^{\lambda_3} - 1}{\lambda_3} - \frac{(1-u)^{\lambda_4} - 1}{\lambda_4}\right], \tag{2.3.2}$$

where $0 \leq u \leq 1$, $\lambda_1, \lambda_2, \lambda_3,$ and λ_4 are real. They conducted a thorough analysis of the distributional properties of the MLF, showing that many major income distributions can be derived as special cases, limiting cases, or approximations from this family. They expressed commonly used income inequality measures in quantile terms and derived those measures for the MLF. A novel estimation procedure, known as percentile method II, is developed by utilizing quantile-based measures of location, dispersion, skewness, and kurtosis. Their simulation study indicated that this estimation method outperforms traditional methods based on percentiles and moments.

The parameters of the MLF distribution are estimated for a real data using this new estimation method. Moreover, they used the truncated Gini index and the income gap ratio to characterize income distributions across various ranges of poverty and affluence.

Pérez and Alaiz (2011) used the Dagum type I model for analyzing the evolution of personal income in Spain from 1995 to 2005, and its quantile function is

$$Q(u) = \lambda^{\frac{1}{\delta}} \left[\left(\frac{1}{u} \right)^{\frac{1}{\beta}} - 1 \right]^{-\frac{1}{\delta}}, \quad 0 \leq u \leq 1, \quad \alpha, \beta, \delta > 0. \tag{2.3.3}$$

The model, as shown in equation (2.3.3), accurately represented the empirical income distribution in Spain. The study also delved into the economic significance of the parameters in the Dagum model. Data taken from the European Community Household Panel (ECHP) and the European Union Statistics on Income and Living Conditions (EU-SILC) are utilized to examine how changes in the model's parameters impacted income inequality growth in Spain and on various income percentiles.

Sreelakshmi and Nair (2014) carried out a quantile-based analysis of income data and explored the potential of the Zenga curve as an alternative measure of income inequality. They conducted an in-depth examination of the Zenga curve's properties and used stochastic orders derived from this curve to substantiate various results. The study identified connections between the Zenga curve and other inequality measures, such as the Bonferroni and Leimkuhler curves. Additionally, they found relationships between the Zenga curve and reliability measures, including the mean residual quantile function and the reversed mean residual quantile function. The research also utilized quantile-based models, such as the Govindarajulu distribution, a linear hazard quantile model, and the PP distribution for income modeling. They also derived various income inequality measures, such as the Lorenz and Bonferroni curves, corresponding to these models. Moreover, their work explored the link between L-moments and income inequality indices and extended the study to bivariate reliability concepts using copulas.

Sánchez et al. (2021) developed quantile regression models using the Birnbaum-Saunders distribution and its diagnostics, specifically for application to economic data. Saulo et al. (2023) provided parametric quantile regressions based on the SM and Dagum income distributions.

2.4 Quantile functions in reliability analysis

Nair et al. (2012) explored the technique of developing quantile functions for lifetime models by using the connection between the hazard quantile function and the Parzen score function (Parzen (1979)). The paper presented three models based on the score function, highlighting that many well-known distributions are special cases of these models. Additionally, various reliability properties of the Parzen score function are analyzed.

The concept of reliability in terms of quantiles is extensively discussed by Nair et al. (2013a). They also examined distributions with closed-form quantile functions, aging concepts, total time on test transforms, L-moments of residual life, hazard quantile function, stochastic orders, and the application of quantile-based modeling.

Bijamma et al. (2014) presented a software reliability model with a quantile function

$$Q(u) = k\,\beta_u(a+1, b+1) \tag{2.4.1}$$

where $0 \leq u \leq 1$, $k > 0$, a, b are real and $\beta_u(.,.)$ denotes an incomplete beta function. The study examined several distributional and reliability aspects of this quantile function. They managed to approximate (2.4.1) to two popularly recognized distributions namely the inverse Gaussian and the Weibull. The parameter estimation for this model is conducted using the L-moments technique, and the model is applied to real data.

The study by Nair et al. (2018) focused on examining the reliability aspects of the quantile-based proportional hazard model (PHM). They derived and showcased the aging properties and characterizations of the PHM with examples. This work also explored significant stochastic orders relevant to the PHM. Additionally, the dynamic cumulative residual Kullback-Leibler divergence based on quantiles is studied in the framework of PHM.

Vineshkumar and Nair (2019) introduced a new definition of bivariate quantile function appropriate for reliability modeling and demonstrated its applications. However, this work did not address the general properties of bivariate quantile functions. Later in 2021, Nair and Vineshkumar (2021) conducted an initial study into the properties of bivariate distributions expressed through quantile functions.

Some new quantile function models introduced by Dileep Kumar and Sankaran (2019) are useful in lifetime data analysis. In particular, the models are the half logistic-exponential geometric quantile function, the Pareto-Weibull quantile function, a new class of distributions that extends the class of distributions with a linear mean residual quantile function and the class of distributions with quadratic mean residual quantile functions. They thoroughly examined the distributional characteristics and reliability measures based on the quantiles of the proposed models. The estimation of parameters utilized the L-moments method and the percentile method. Real-life datasets are employed to validate the practical uses of these models. Driven by the special characteristics of quantile functions, they studied the properties and uses of the proportional odds model within a quantile framework. In this work, they discussed the use of quantile functions in the modeling and analysis of competing risks data. They also introduced and analyzed the cause-specific hazard quantile function. Furthermore, a definition of the revelation transform based on quantiles is proposed, along with an investigation into its features in the realm of lifetime data analysis. Additionally, the study delved into the reliability aspects of the proportional hazards revelation transform.

Aswin et al. (2020) proposed a class of distributions with a quadratic hazard quantile function and is given as

$$Q(u) = \frac{1}{2(\beta-\alpha)}\log(1+u) - \frac{1}{2(\beta+\alpha)}\log(1-u) - \frac{\alpha}{\beta^2-\alpha^2}\log\left(\frac{\beta+\alpha\, u}{\beta}\right) \quad (2.4.2)$$

where $0 \leq u \leq 1$, $\beta > 0$, $\beta > |\alpha|$, and $\beta \neq \alpha$. The authors explored various distributional properties, reliability features, and the characterizations of this model. They employed the least square method for estimation and illustrated the practical usefulness of the above model with real data.

Jeena and Asisha (2021) introduced the power exponential geometric distribution, which is obtained as the sum of the quantile functions of the power and exponential geometric distributions. Its quantile function is

$$Q(u) = \alpha\, u^{\frac{1}{\beta}} + \frac{1}{\lambda} \log\left(\frac{1-\delta\, u}{1-u}\right), \tag{2.4.3}$$

where $0 \leq u \leq 1$, α, β, $\lambda > 0$ and $0 < \delta < 1$. The study explored its distributional and reliability properties. Additionally, a simulation study is conducted, and the model is applied to real data for validation.

Deepthy et al. (2023) introduced the Burr III-Weibull (BIIIW) quantile function, which combines the quantile functions of both the Burr III and Weibull distributions and is given as

$$Q(u) = \sigma\left(-\log(1-u)\right)^{\frac{1}{\lambda}} + \left(u^{-\frac{1}{k}} - 1\right)^{-\frac{1}{c}}, \quad 0 \leq u \leq 1, \quad c,\, k,\, \sigma,\, \lambda > 0. \tag{2.4.4}$$

They discussed various distributional and reliability characteristics of this new model. Parameter estimation for the model is carried out using the least squares method and the L-moments method. The application of the BIIIW model is examined through the two real-life datasets. Due to the flexibility of the hazard quantile function, the model is efficient in modeling various types of lifetime data.

2.5 Income inequality measures

Some of the references related to income inequality measures discussed in Section 1.5 are excluded from this section.

The book edited by Chotikapanich (2008) focuses on modeling income distributions and Lorenz curves. The first section compiles five major papers in this area. The second section features four survey papers that discuss Lorenz functions, along with generalizations and extensions of certain income distributions. The final section contains eight papers highlighting recent research and developments in this field.

For thirty-five continuous distributions, Giorgi and Nadarajah (2010) has provided explicit expressions for the Bonferroni curve and Bonferroni index, in addition to the Lorenz curve and Gini index.

Fellman (2012) investigated two optimal scenarios in which the transformed variable either Lorenz dominates the original variable or is dominated by it. The first scenario is more applicable in practical terms than the second, as it leads to policies that reduce inequality. This study examines additional characteristics of the transformed Lorenz curves, particularly focusing on their limits.

The study by Nair et al. (2013b) examined the distributional and geometric characteristics of first and second-order partial moments when expressed in quantile terms. The rth order partial moment in quantile terms is given as

$$P_r(u) = \int_u^1 (Q(p) - Q(u))^r \; dp. \tag{2.5.1}$$

This paper primarily explores the concept of the quantile-based stop-loss transform. They also developed relationships between the scaled stop-loss transform curve and the Gini, Lorenz, Leimkuhler, and Bonferroni curves.

Sarabia and Jordá (2014) obtained explicit formulas for the Pietra index for different distributions like the generalized gamma distribution (GG), the generalized beta of the first kind (GB1), the generalized beta of the second kind (GB2), and its special cases. Additionally, they derived the Pietra index for other significant income models, including finite distribution mixtures and the κ-generalized distribution. The paper concludes with two practical examples using actual income data.

Utilizing a semiparametric approach, Pillai et al. (2014) estimated the Lorenz curve and Gini index specifically for the exponential distribution. These estimations were performed under type 1, type 2, and interval censoring. A Monte Carlo simulation study is utilized to compare the performance of the estimators based on their mean square error (MSE).

Porro (2015) examined the Lorenz ordering and orderings associated with the $I(u)$ curve for the Zenga distribution. Arnold (2015) explored the relationships, scale invariance, and inequality partial orders of the Lorenz, Bonferroni, Zenga, and $I(u)$ curves. The Bonferroni and Zenga curves are fundamentally similar to the Lorenz curve, and they identify the parent distribution up to a scale factor. Although $I(u)$ is scale-invariant, various distributions might have the same $I(u)$ curve, making it unable to identify the parent distribution up to a scale factor.

The book by Kämpke and Radermacher (2015) targets experts and informed lay readers interested in the mathematical foundations of Lorenz curves. It is divided into two parts: the first part delves into quantile functions, Lorenz curves, their mathematical properties, and partial ordering in detail. It also covers the extension of Pigou–Dalton transfers to general probability distributions using weak convergence and discusses the Atkinson theorem. The second part of the book focuses on specific Lorenz curves, explores methods for deriving Lorenz curves, and discusses a computational model that utilizes Lorenz curves.

Gómez-Déniz (2016) introduced a new family of parametric Lorenz curves using the arctan function by incorporating an extra parameter into an initial Lorenz curve $L_0(u)$. Specific cases of this family are derived using the initial Lorenz curve $L_0(u)$ as egalitarian, the Aggarwal and Singh Lorenz curve, and the Pareto Lorenz curve. The paper outlines the process for deriving distribution functions, calculating Gini and Pietra inequality measures, and suggests a technique for determining the Gini index through the inverse Lorenz curve. The utility of this family of curves is demonstrated through its application to two datasets yielding satisfactory outcomes through statistical techniques like least squares and maximum likelihood, proving its capability in precisely modeling income data.

Sarabia et al. (2017) offered a brief review of Theil indices and presented general formulas for these indices. This paper includes closed-form expressions for the Theil indices of several distributions, including the GG, GB1, GB2, Champernowne distribution, the κ-generalized income distribution, and various Pareto and gamma-type distributions. Additionally, they discuss the decomposition of generalized entropy measures and propose extending the Theil indices to higher dimensions.

Fellman (2018) thoroughly discussed the Lorenz curve and succinctly explained the Gini and Pietra indices. This work explored how the Lorenz curves, along with the Gini and Pietra indices, vary according to the model parameters for the Pareto (Rasche et al. (1980)), simplified Rao Tam (Rao and Tam (1987)), and Chotikapanich (Chotikapanich (1993)) distributions. It is noted that when these three models have an identical Gini index, the Lorenz curves of the simplified Rao Tam and Chotikapanich models appear quite alike, in contrast to the distinct Lorenz curve associated with Pareto's model.

Behdani and Borzadaran (2019) redefined some measures of income inequality through the use of quantiles. They characterized probability distributions by exploring the connection between the Lorenz curve and concepts from reliability theory, such as the mean residual quantile function and the reversed mean residual quantile function. Additionally, they used the quantile function and the Lorenz curve to study aging concepts.

Arora and Jain (2006) developed a test to evaluate generalized Lorenz dominance within a specific range but a test for assessing multisample generalized Lorenz dominance had not been explored. To address this, Rattan et al. (2021) introduced a test for this assessment using two distinct strategies: one based on specified intervals and another on confidence intervals. They also conducted a simulation study to evaluate the effectiveness of the proposed test.

A non-parametric method for estimating the Gini index in cases where the sample includes right-censored observations has been presented by Sudheesh et al. (2021). The proposed estimator is consistent and follows an asymptotic normal distribution. They used Monte Carlo simulations to evaluate the estimator's effectiveness. The result of the simulation study indicates that the confidence interval for the Gini index, derived from the proposed estimator, has a high coverage probability and is easy to implement.

Mdingi and Ho (2021) conducted a comprehensive literature review exploring the relationship between income inequality and economic growth. Their study thoroughly examines both theoretical and empirical literature to understand the relationship between income inequality and growth.

Outliers at both ends of an income distribution can impact the estimation of income inequality. Safari et al. (2021b) introduced a semi-parametric technique, which is a combination of inverse Pareto, empirical, and Pareto distributions to evaluate income inequality. This method employs a robust statistical estimation, based on the probability integral transform, to calculate the shape parameters for the inverse-Pareto and Pareto models, accommodating outliers in both tails of the income distribution. Using the semi-parametric model, they derived the Lorenz curve, Gini coefficient, generalized entropy index, and Atkinson index. A simulation study shows that the semi-parametric approach performs better than the traditional non-parametric methods for measuring income inequality, particularly in the presence of outliers. As an application, this semi-parametric method is applied to calculate income inequality in Malaysia using data from a household income survey.

Gómez-Déniz et al. (2022) developed parametric Lorenz curves based on the system of beta distributions, including the beta, Kumaraswamy, and generalized beta distributions. They established the Lorenz ordering based on the parameters of these models and derived explicit expressions for the Gini indices corresponding to each Lorenz curve. An empirical study using individual income data from the US indicates that the suggested models provide a precise fit for income data.

Sudheesh et al. (2022) proposed straightforward non-parametric estimators for the Gini covariance, Gini correlation, and Gini regression coefficients. The proposed estimators are consistent and follow an asymptotic normal distribution, with explicit expressions for the asymptotic variances. An empirical study assessed the effectiveness of confidence intervals generated using normal approximation and bootstrap-t. The finite-sample performance of the estimators is further evaluated through a Monte Carlo simulation, focusing on a bivariate Pareto distribution. Finally, simulations demonstrated that the jackknifed versions of the proposed estimators are preferable for small sample sizes, as they successfully reduce the bias present in the original estimators.

In skewed and heavily tailed income distributions, the mean is not an ideal measure of central tendency. Thus Brazauskas et al. (2024) discussed three median-based income inequality indices and their related equality curves, all of which are based on percentiles. These indices Ψ_1, Ψ_2 and Ψ_3 compare the median income of the poor to the median income of the entire population, the non-poor, and an equivalent proportion of the rich, respectively. They analyzed these indices both analytically and numerically using different income distribution models. Real data applications are demonstrated using capital incomes taken from surveys conducted in fifteen European countries in 2001 and 2018. They also studied the impact of income transfers on these indices.

Table 2.1 includes the income inequality curves and measures of the five distributions discussed in the above sections, namely the Bijamma Thomas distribution, the shifted reverse exponential distribution, the κ-generalized distribution, the shifted reverse stretched exponential distribution, and the Weibull-Paretovo distribution. Here income inequality measures are derived for the distributions having explicit quantile functions.

In Table 2.1, LC, GI, PI, BC, BI, and ZC stand for the Lorenz curve, Gini index, Pietra index, Bonferroni curve, Bonferroni index, and Zenga curve, respectively. The value u_0 in the context of the Pietra index is determined by solving for u in the equation $\mu = Q(u)$, where μ represents the mean of the corresponding distribution.

Table 2.1: Income inequality curves and inequality measures

Quantile functions	**Income inequality curves / inequality measures**	
Bijamma Thomas distribution $Q\left(u\right)=k\beta_{u}\left(a+1,b+1\right)$	LC	$\dfrac{[u\beta_u(a+1,b+1)-\beta_u(a+2,b+1)]}{\beta(a+1,b+2)}$
	GI	$(a+1)(a+b+3)^{-1}$
	PI	$\dfrac{\beta_{u_0}(a+2,b+1)}{\beta(a+1,b+2)}$
	BC	$\dfrac{1}{\beta(a+1,b+2)}\left[\beta_u(a+1,b+1)-\frac{1}{u}\beta_u(a+2,b+1)\right]$
	BI	$\dfrac{1}{\beta(a+1,b+2)}\left\{\dfrac{\pi\csc(b\pi)\Gamma(a+2)\left[\text{Har.no}(a+1)-\text{Har.no}(a+b+2)\right]}{\Gamma(-b)\Gamma(a+b+3)}\right\}$
	ZC	$\dfrac{\beta(a+1,b+2)-\beta_u(a+1,b+1)+\frac{1}{u}\beta_u(a+2,b+1)}{\beta(a+1,b+2)-u\beta_u(a+1,b+1)+\beta_u(a+2,b+1)}$
Shifted reverse exponential distribution $Q(u)=x_0+\dfrac{1}{\lambda}\log(u)$	LC	$\dfrac{u(\log u+\lambda x_0-1)}{\lambda x_0-1}$
	GI	$\dfrac{1}{2(\lambda x_0-1)}$
	PI	$\dfrac{u_0}{\lambda x_0-1}$
	BC	$\dfrac{\log u+\lambda x_0-1}{\lambda x_0-1}$
	BI	$\dfrac{1}{\lambda x_0-1}$
	ZC	$\dfrac{\log u}{1+u(\log u-1)-\lambda x_0(1-u)}$
κ-generalized distribution $Q(u)=\beta\left[\ln_{\kappa}\left(\frac{1}{1-u}\right)\right]^{\frac{1}{\alpha}}$	LC	$I_{1-(1-u)^{2\kappa}}\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)$
	GI	$1-\dfrac{2\beta\left(\frac{1}{\alpha}+1,\frac{1}{\kappa}-\frac{1}{2\alpha}\right)}{\beta\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}$
	PI	$\dfrac{2\kappa u_0(1-u_0)^{-\kappa/\alpha}\left(1-(1-u_0)^{2\kappa}\right)^{1/\alpha}-\beta_{1-(1-u_0)^{2\kappa}}\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}{\beta\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}$
	BC	$\dfrac{1}{u}I_{1-(1-u)^{2\kappa}}\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)$
	BI	$1-\displaystyle\int_0^1\frac{1}{u}I_{1-(1-u)^{2\kappa}}\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)du$
	ZC	$\dfrac{u\beta\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)-\beta_{1-(1-u)^{2\kappa}}\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}{u\left[\beta\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)-\beta_{1-(1-u)^{2\kappa}}\left(\frac{1}{\alpha}+1,\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)\right]}$

Table 2.1: (continued)

Quantile functions	Income inequality curves / inequality measures	
Shifted reverse stretched exponential distribution $Q(u) = x_0 - \lambda(-\log u)^{\frac{1}{\rho}}$	LC	$\dfrac{x_0 u - \lambda\,\Gamma\left(1+\frac{1}{\rho}, -\log u\right)}{x_0 - \lambda\,\Gamma\left(1+\frac{1}{\rho}\right)}$
	GI	$(1 - 2^{\frac{-1}{\rho}})\dfrac{\lambda\,\Gamma\left(1+\frac{1}{\rho}\right)}{\left[x_0 - \lambda\,\Gamma\left(1+\frac{1}{\rho}\right)\right]}$
	PI	$\dfrac{\lambda\,\Gamma\left(1+\frac{1}{\rho}, -\log u_0\right) - \lambda u_0(-\log u_0)^{\frac{1}{\rho}}}{x_0 - \lambda\,\Gamma\left(1+\frac{1}{\rho}\right)}$
	BC	$\dfrac{x_0 - \lambda u^{-1}\Gamma\left(1+\frac{1}{\rho}, -\log u\right)}{x_0 - \lambda\,\Gamma\left(1+\frac{1}{\rho}\right)}$
	BI	$\dfrac{\lambda\,\Gamma\left(1+\frac{1}{\rho}\right)}{x_0\rho - \lambda\,\Gamma\left(\frac{1}{\rho}\right)}$
	ZC	$\dfrac{\lambda\left[\Gamma\left(1+\frac{1}{\rho}, -\log u\right) - u\Gamma\left(1+\frac{1}{\rho}\right)\right]}{u\left\{x_0(1-u) + \lambda\left[\Gamma\left(1+\frac{1}{\rho}, -\log u\right) - \Gamma\left(1+\frac{1}{\rho}\right)\right]\right\}}$
Weibull-Paretovo distribution $Q(u) = \lambda + \eta\left[(1-u)\,(-\log(1-u))^{\beta} + \dfrac{u}{(1-u)^{\gamma}}\right]$	LC	$\frac{1}{\mu}\left\{(\gamma-1)^{-1}(\gamma-2)^{-1}(1-u)^{-\gamma}\left[(1-u)^{\gamma} + u(u+\gamma-u\gamma) - 1\right]\eta + \lambda u + 2^{-(\beta+1)}\eta\left[\Gamma(\beta+1) - \Gamma(\beta+1, -2\log(1-u))\right]\right\}$
	GI	$\dfrac{-3^{-(\beta+1)}\eta\left[6^{\beta+1}(\gamma+1) + \left(2^{\beta+2} - 3^{\beta+1}\right)(\gamma-1)(\gamma-2)(\gamma-3)\Gamma(\beta+1)\right]}{(\gamma-3)\left\{2^{\beta+1}[\eta + (\gamma-1)(\gamma-2)\lambda] + \eta(\gamma-1)(\gamma-2)\Gamma(\beta+1)\right\}}$
	PI	$\frac{\eta}{\mu}\{u_0\,(1-u_0)\,[-\log(1-u_0)]^{\beta} + (\gamma-1)^{-1}(\gamma-2)^{-1}\,(1-u_0)^{-\gamma}\left[u_0^2(\gamma-1)^2 - u_0\gamma + 1 - (1-u_0)^{\gamma}\right] - 2^{-(\beta+1)}\left[\Gamma(\beta+1) - \Gamma\left(\beta+1, -2\log(1-u_0)\right)\right]\}$
	BC	$\frac{1}{\mu}\left\{(\gamma-1)^{-1}(\gamma-2)^{-1}u^{-1}(1-u)^{-\gamma}\left[(1-u)^{\gamma} + u(u+\gamma-u\gamma) - 1\right]\eta + \lambda + 2^{-(\beta+1)}\,\eta\,u^{-1}[\Gamma(\beta+1) - \Gamma(\beta+1, -2\log(1-u))]\right\}$
	BI	$1 - \int_0^1 \frac{1}{\mu}\left\{(\gamma-1)^{-1}(\gamma-2)^{-1}u^{-1}(1-u)^{-\gamma}\left[(1-u)^{\gamma} + u(u+\gamma-u\gamma) - 1\right]\eta + \lambda + 2^{-(\beta+1)}\eta\,u^{-1}[\Gamma(\beta+1) - \Gamma(\beta+1, -2\log(1-u))]\right\}du$
	ZC	$\frac{\eta}{2}.\frac{k_1}{k_2}$

The symbol $\beta_u(.,.)$ represents the incomplete beta function, while csc(.) denotes the cosecant function. The notation Har.no(n) represents the n^{th} harmonic number, defined as Har.no$(n) = \sum_{k=1}^{n} \frac{1}{k}$. The function $\Gamma(.)$ stands for the gamma function, while $\Gamma(.,.)$ represents the incomplete gamma function. The term, $I_u(.,.)$ denotes the regularized incomplete beta function, defined as $I_u(.,.) = \frac{\beta_u(.,.)}{\beta(.,.)}$. For the Weibull-Paretovo distribution the mean, $\mu = \lambda + \frac{\eta}{(\gamma-1)(\gamma-2)} + 2^{-1-\beta}\eta\,\Gamma(\beta+1)$. Finally, $k_1 = \frac{2(1-u)^{-\gamma+1}[(1-u)^{\gamma}+\gamma u-u-1]}{(\gamma-1)(\gamma-2)} + 2^{-\beta}[\beta\Gamma(\beta) - u\Gamma(\beta+1) - \Gamma(\beta+1, -2\log(1-u))]$, and $k_2 = u(1-u)\left[\frac{(1-u)^{-\gamma}(\gamma u-u-1)\eta}{(\gamma-1)(\gamma-2)} - \lambda\right] - 2^{-(\beta+1)}\eta u\Gamma(\beta+1, -2\log(1-u))$.

2.6 Estimation and applications

This section uses the per capita personal income data to evaluate the suitability of the shifted reverse exponential, Bijamma Thomas, and Weibull-Paretovo distributions for modeling income. The parameters for these three distributions are estimated using the method of percentiles I.

2.6.1 Method of percentiles I

The percentile method I is one of several techniques used to estimate the parameters of a quantile function. We apply the percentile method I discussed in Karian and Dudewicz (1999) to estimate the parameters.

Consider a dataset $X_1, X_2, \ldots, X_n$ and define $\tilde{\pi}_p$ as the $(100\,p)^{th}$ percentile of the data. The percentile $\tilde{\pi}_p$ can be calculated by initially expressing $(n+1)p$ as $r + \frac{a}{b}$, where r is a positive integer, and the fraction $(\frac{a}{b})$ takes value in the interval $[0, 1]$. Then, $\tilde{\pi}_p$ is determined by

$$\tilde{\pi}_p = X_{(r)} + \frac{a}{b}(X_{(r+1)} - X_{(r)}) \tag{2.6.1}$$

where $X_{(1)}, X_{(2)} \ldots X_{(n)}$ represent the ordered dataset.

Now we examine the four sample percentiles, $\tilde{\upsilon}_1, \tilde{\upsilon}_2, \tilde{\upsilon}_3$, and $\tilde{\upsilon}_4$, which is given as

$$\tilde{\upsilon}_1 = \tilde{\pi}_{0.5}, \tag{2.6.2}$$

$$\tilde{\upsilon}_2 = \tilde{\pi}_{0.9} - \tilde{\pi}_{0.1}, \tag{2.6.3}$$

$$\tilde{\upsilon}_3 = \frac{\tilde{\pi}_{0.5} - \tilde{\pi}_{0.1}}{\tilde{\pi}_{0.9} - \tilde{\pi}_{0.5}}, \tag{2.6.4}$$

$$\tilde{\upsilon}_4 = \frac{\tilde{\pi}_{0.75} - \tilde{\pi}_{0.25}}{\tilde{\upsilon}_2}. \tag{2.6.5}$$

Here $\tilde{\upsilon}_1$, $\tilde{\upsilon}_2$, $\tilde{\upsilon}_3$ and $\tilde{\upsilon}_4$ represent the sample median, inter-decile range, the ratio of the left tail-weight to the right tail-weight and ratio of the inter-quartile range to the inter-decile range respectively. In this method, the parameter estimates are derived by matching the sample percentiles with their respective population percentiles i.e., $\tilde{\upsilon}_i = \upsilon_i$ for $i = 1, 2, 3, 4$.

2.6.2 Applications

To explore the potential of the shifted reverse exponential, Bijamma Thomas, and Weibull-Paretovo distributions for modeling income data, we use the 2021 per capita personal income (in dollars) from 64 counties in Colorado, US. This data is sourced from `https://www.bea.gov`. The parameters for these three distributions are estimated using the percentile method.

The Chi-square test and Q-Q plot play a crucial role in evaluating the adequacy

Table 2.2: Parameter estimates, chi-square statistic, and p-value

Distribution	Parameter estimates	Chi-square statistic	p-value
Shifted reverse exponential	$\hat{x}_0 = 7.11434 \times 10^4$ $\hat{\lambda} = 4.50376 \times 10^{-5}$	21.09340	0.02045
Bijamma Thomas	$\hat{k} = 9.00356 \times 10^2$ $\hat{a} = -1.95688$ $\hat{b} = -2.66202$	7.86247	0.64227
Weibull-Paretovo	$\hat{\lambda} = 3.26130 \times 10^4$ $\hat{\eta} = 2.05187 \times 10^4$ $\hat{\beta} = 4.03255 \times 10^{-1}$ $\hat{\gamma} = 4.78090 \times 10^{-1}$	5.00374	0.89093

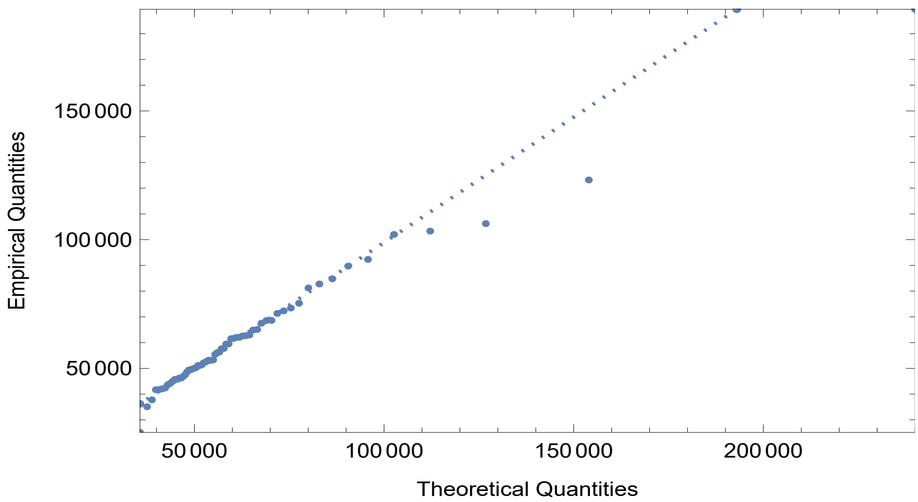


Figure 2.1: Q-Q plot for the per capita personal income of counties in Colorado state in 2021

of the model. Table 2.2 presents the parameter estimates, test statistics, and p-values for the shifted reverse exponential, Bijamma Thomas, and Weibull-Paretovo distributions. Although the p-value for the Bijamma Thomas distribution exceeds 0.05, the Q-Q plot indicates a poor fit. Therefore, based on Table 2.2 and the Q-Q plot in Figure 2.1, we conclude that the Weibull-Paretovo distribution provides the best fit for the real data under consideration.

The income inequality measures such as Gini, Pietra, and Bonferroni indices of the Weibull–Paretovo distribution are presented in Table 2.1. These indices take the values 0.21783, 0.15519, and 0.27319, respectively, indicating a low level of income inequality in the above dataset.

2.7 Summary

This chapter has examined recent studies on income distributions, income inequality measures, and quantile functions from the literature. The study is organized into five sections: income distributions, quantile-based income models, new quantile functions for reliability assessments, income inequality measures, and applications. For the five distributions, we derived the Lorenz curve, Gini index, Pietra index, Bonferroni curve, Bonferroni index, and Zenga curve. Three distributions are applied to the per capita personal income data of 64 counties in Colorado, and the Weibull-Paretovo distribution is found to provide the best fit.

CHAPTER 3

A QUANTILE-BASED INCOME ANALYSIS OF POWER-PARETO DISTRIBUTION

3.1 Introduction

Income modeling using distribution functions dates back over a century, while the quantile-based approach was more modestly introduced by Tarsitano in 2004. A key property of quantiles is that the sum of two quantile functions produces another quantile function, and their product also results in a quantile function when restricted to non-negative values. The Power-Pareto (PP) distribution (Gilchrist (2000), Hankin and Lee (2006)) is derived from the product of the quantile functions of the Power and Pareto distributions. This chapter utilizes the PP distribution within a quantile-based framework to model income.

The remaining portion of the chapter is structured as follows. Section 3.2 discusses the origin and properties of the PP distribution. Section 3.3 derives the key quantile-based income inequality measures associated with the PP distribution and study Lorenz ordering. In Section 3.4, we discuss three quantile-based estimation methods for this distribution. Section 3.5 presents a simulation study using the PP distribution to evaluate the capabilities of the three estimating approaches used in this study. In Section 3.6, we apply the above model to real income datasets and calculate both empirical and estimated income inequality measures. Finally, Section 3.7 concludes the study.

3.2 The PP distribution

The quantile function of the PP distribution is

$$Q(u) = \frac{C\, u^{\lambda_1}}{(1-u)^{\lambda_2}}, \quad 0 \leq u \leq 1, \quad C > 0,\, \lambda_1 > 0,\, \lambda_2 > 0. \tag{3.2.1}$$

Here, C denotes the scale parameter, while λ_1 and λ_2 are shape parameters that determine the behavior of the left and right tails, respectively. Also, it is possible that either λ_1 or λ_2 to be zero, but not both simultaneously. This distribution is also known as the 'Hankin and Lee' distribution and the 'Davies' distribution.

The quantile density function of the PP distribution is

$$q(u) = \frac{C\, u^{\lambda_1}}{(1-u)^{\lambda_2}} \left[\frac{\lambda_1}{u} + \frac{\lambda_2}{(1-u)} \right]. \tag{3.2.2}$$

Hankin and Lee (2006) analyzed the properties of the PP distribution and compared it with other distributions, employing both maximum likelihood estimation and the least squares technique for parameter estimation. Sankaran and Preetha Kumari (2010) utilized the PP distribution to develop a parameter regression model and assessed its attributes within a reliability context. Nair and Vineshkumar (2010) showcased the application of the PP distribution in reliability studies and validated their findings with real data. Nair et al. (2013a) studied the basic characteristics of the PP distribution, including L-moments, percentile-based measures, and certain reliability-based characterizations. Sreelakshmi and Nair (2014) briefly examined and derived economic measures such as the Lorenz curve and income share elasticity for this distribution. Recently, Caeiro and Norouzirad (2024) compared the performance of various estimation methods for the PP distribution through a simulation study and real data analysis. Research on income studies using the PP distribution has been limited, so this work focuses on a quantile-based income analysis of this distribution.

3.3 Income inequality measures

Income inequality describes the extent of uneven distribution of income across a population. It continues to be a significant topic of discussion even today, owing to the considerable income disparities that exist in the population.

The income inequality measures of the PP distribution discussed in this section are derived using quantile-based equations.

The Lorenz curve of the PP distribution, based on (1.5.4) is

$$\begin{aligned} L(u) &= \frac{\beta_u(\lambda_1+1, 1-\lambda_2)}{\beta(\lambda_1+1, 1-\lambda_2)} \\ &= I_u(\lambda_1+1, 1-\lambda_2), \quad \lambda_2 < 1, \end{aligned} \tag{3.3.1}$$

where $\beta(.,.)$ and $\beta_u(.,.)$ represent a complete beta function and incomplete beta function respectively. The regularized incomplete beta function is represented as $I_u(.,.)$.

The Gini index of the PP distribution is derived using (1.5.7) and is

$$G = \frac{\lambda_1+\lambda_2}{\lambda_1-\lambda_2+2}, \quad \lambda_2 < 1. \tag{3.3.2}$$

Now, the generalized Gini index based on (1.5.10) is

$$G_\delta = 1 - \frac{\delta\, \beta(\lambda_1+1, \delta-\lambda_2)}{\beta(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1. \tag{3.3.3}$$

The relative mean deviation and the Pietra index of the PP distribution obtained using (1.5.15) and (1.5.16) are

$$\tau = \frac{2\left[u_0^{\lambda_1+1}(1-u_0)^{-\lambda_2} - \beta_{u_0}(\lambda_1+1, 1-\lambda_2)\right]}{\beta(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1, \tag{3.3.4}$$

$$P = \frac{u_0^{\lambda_1+1}(1-u_0)^{-\lambda_2} - \beta_{u_0}(\lambda_1+1, 1-\lambda_2)}{\beta(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1. \tag{3.3.5}$$

Here u_0 is determined by solving the equation $\mu = Q(u)$ for u, where the mean μ of the PP distribution is given by $C\beta(\lambda_1+1, 1-\lambda_2)$, and the corresponding $Q(u)$ is provided in (3.2.1).

The Bonferroni curve and the Bonferroni index of the above distribution, based on (1.5.20) and (1.5.21) are

$$B(u) = \frac{1}{u} I_u(\lambda_1+1, 1-\lambda_2), \quad \lambda_2 < 1, \tag{3.3.6}$$

$$B = 1 - \frac{{}_3F_2(\lambda_1+1, \lambda_1+1, \lambda_2; \lambda_1+2, \lambda_1+2; 1)}{(\lambda_1+1)^2\beta(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1, \tag{3.3.7}$$

where ${}_3F_2(\lambda_1+1, \lambda_1+1, \lambda_2; \lambda_1+2, \lambda_1+2; 1) = \sum_{k=0}^{\infty} \frac{(\lambda_1+1)_k(\lambda_1+1)_k(\lambda_2)_k}{(\lambda_1+2)_k(\lambda_1+2)_k}\left(\frac{1}{k!}\right)$ is a generalized hypergeometric function and $(\lambda_1+1)_k$ represents an ascending factorial.

The Lorenz curve, Gini index, Bonferroni curve, and Bonferroni index of the PP distribution are discussed in Giorgi and Nadarajah (2010).

The Atkinson index and generalized entropy measure of the PP distribution based on (1.5.23) and (1.5.25) are

$$A_\epsilon = 1 - \frac{\left[\beta\left(\lambda_1(1-\epsilon)+1, 1-\lambda_2(1-\epsilon)\right)\right]^{\frac{1}{1-\epsilon}}}{\beta(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1. \tag{3.3.8}$$

$$GE_\theta = \frac{1}{\theta(\theta-1)}\left[\frac{\beta(\lambda_1\theta+1, 1-\lambda_2\theta)}{\left(\beta(\lambda_1+1, 1-\lambda_2)\right)^\theta} - 1\right], \quad \lambda_2 < 1. \tag{3.3.9}$$

The Theil coefficients T_1 and T_2 of the PP distribution derived using (1.5.28) and (1.5.29) are

$$\begin{aligned} T_1 = {} & \frac{(1-u_0)^{\lambda_2}}{u_0^{\lambda_1}}\beta\left(\lambda_1+1, 1-\lambda_2\right)\Bigg\{\log C + \lambda_1\left[\Psi\left(\lambda_1+1\right) - \Psi\left(\lambda_1+2-\lambda_2\right)\right] \\ & - \lambda_2\left[\Psi\left(1-\lambda_2\right) - \Psi\left(\lambda_1+2-\lambda_2\right)\right] - \log\frac{Cu_0^{\lambda_1}}{(1-u_0)^{\lambda_2}}\Bigg\}, \end{aligned} \tag{3.3.10}$$

$$T_2 = \lambda_1\left(\log u_0 + 1\right) - \lambda_2\left(\log\left(1-u_0\right)+1\right), \tag{3.3.11}$$

where $\Psi(.)$ denotes digamma function.

The Frigyes measures ρ, ν and η of the PP using (1.5.31)-(1.5.33) are as follows

$$\begin{aligned} \rho &= \frac{u_0^{\lambda_1+1}}{(1-u_0)^{\lambda_2}}\frac{1}{\beta_{u_0}(\lambda_1+1, 1-\lambda_2)}, \quad \lambda_2 < 1. \\ \nu &= \frac{u_0}{(1-u_0)}\left[\frac{1}{I_{u_0}(\lambda_1+1, 1-\lambda_2)} - 1\right], \quad \lambda_2 < 1. \\ \eta &= \frac{(1-u_0)^{\lambda_2-1}}{u_0^{\lambda_1}}\left[\beta(\lambda_1+1, 1-\lambda_2) - \beta_{u_0}(\lambda_1+1, 1-\lambda_2)\right], \quad \lambda_2 < 1. \end{aligned} \tag{3.3.12}$$

The Zenga curve is related to the Lorenz curve and is represented by (1.5.39). Accordingly, the Zenga curve of the PP distribution is given as follows

$$I(u) = \frac{u\beta(\lambda_1+1, 1-\lambda_2) - \beta_u(\lambda_1+1, 1-\lambda_2)}{u\left[\beta(\lambda_1+1, 1-\lambda_2) - \beta_u(\lambda_1+1, 1-\lambda_2)\right]}, \quad \lambda_2 < 1. \tag{3.3.13}$$

The Leimkuhler curve of the PP distribution, also derived in Nair and Vineshkumar (2022), is given as

$$\begin{aligned} K(u) &= 1 - \frac{\beta_{1-u}(\lambda_1+1, 1-\lambda_2)}{\beta(\lambda_1+1, 1-\lambda_2)} \\ &= 1 - I_{1-u}(\lambda_1+1, 1-\lambda_2), \quad \lambda_2 < 1. \end{aligned} \tag{3.3.14}$$

The Lorenz, Bonferroni, and Zenga curves of PP distribution are given in Figure 3.1, 3.2, and 3.3 respectively.

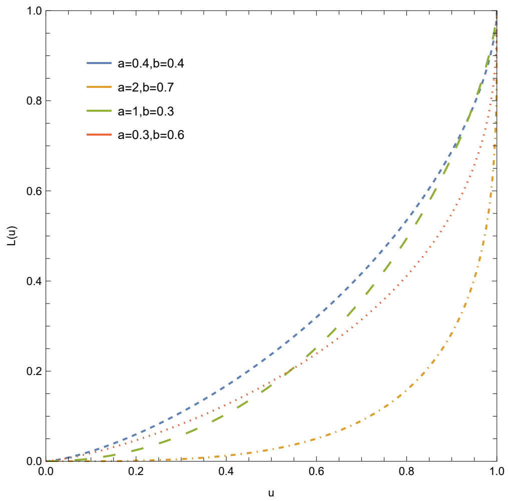


Figure 3.1: Lorenz curves of PP distribution

Most poverty or wealth indices derived from income data typically rely on the proportion of individuals within a given category, combined with their income dis-

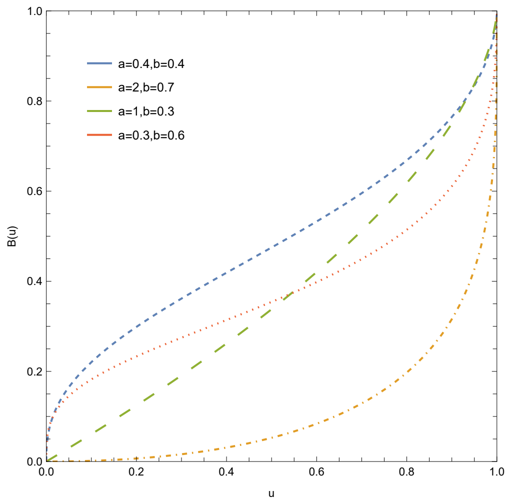


Figure 3.2: Bonferroni curves of PP distribution

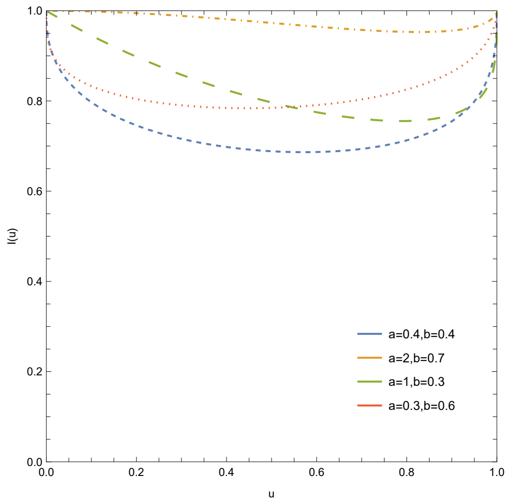


Figure 3.3: Zenga curves of PP distribution

tribution. This is often assessed using the income gap ratio and income inequality measures, such as the Gini index, truncated at a specific income threshold. Sen (1976), Takayama (1979), and Sen (1986) addressed these indices. In Nair et al. (2008), we find a detailed study of income gap ratios for both the poor and the rich and the truncated Gini indices for each group separately.

Let u and $(1 - u^*)$ represent the proportion of poor and rich individuals in the population, respectively. In quantile terms, the income gap ratio for both the poor and the rich is expressed as follows

$$\alpha(u) = 1 - \frac{\int_0^u Q(p)dp}{uQ(u)}. \tag{3.3.15}$$

$$\alpha^*(u^*) = 1 - \frac{(1-u^*)Q(u^*)}{\int_{u^*}^1 Q(p)dp}. \tag{3.3.16}$$

The truncated Gini index for the poor and the rich in quantile terms are already discussed in (1.5.11) and (1.5.12).

Now the income gap ratio and the truncated Gini index for poor and rich based on the PP distribution are given as

$$\alpha(u) = 1 - \frac{(1-u)^{\lambda_2}}{u^{\lambda_1+1}} \beta_u(\lambda_1+1, 1-\lambda_2), \tag{3.3.17}$$

$$\alpha^*(u^*) = 1 - \frac{u^{*\lambda_1}(1-u^*)^{1-\lambda_2}}{\beta(\lambda_1+1, 1-\lambda_2) - \beta_{u*}(\lambda_1+1, 1-\lambda_2)}, \tag{3.3.18}$$

$$G(u) = \left(\frac{2}{u}\right) \frac{\beta_u(\lambda_1+2, 1-\lambda_2)}{\beta_u(\lambda_1+1, 1-\lambda_2)} - 1, \tag{3.3.19}$$

$$G^*(u^*) = 1 - \left(\frac{2}{1-u^*}\right) \frac{\beta(\lambda_1+1, 2-\lambda_2) - \beta_{u*}(\lambda_1+1, 2-\lambda_2)}{\beta(\lambda_1+1, 1-\lambda_2) - \beta_{u*}(\lambda_1+1, 1-\lambda_2)}. \tag{3.3.20}$$

Here in equations (3.3.17)-(3.3.20) we have $\lambda_2 < 1$.

Remark 3.3.1. *For a PP distribution with parameters $(C, \lambda_1, 0)$, the income inequality curves and measures specified in (3.3.1)-(3.3.14) and (3.3.17)-(3.3.20) reduces to corresponding income inequality curves and measures of the Power distribution.*

Remark 3.3.2. *For a PP distribution with parameters $(C, 0, \lambda_2)$, the income inequal-*

ity curves and measures specified in (3.3.1)-(3.3.14) and (3.3.17)-(3.3.20) reduces to corresponding income inequality curves and measures of the Pareto distribution.

3.3.1 Lorenz ordering

A brief description of Lorenz ordering is provided in Section 1.5.1. When the quantile function is available in a closed form, we can employ star-shaped ordering to establish Lorenz ordering. Therefore, in this context, we can apply the Definition 1.5.1 to demonstrate the Lorenz ordering of the PP distribution.

Theorem 3.3.1. *The Lorenz ordering for a PP distribution is,*

$Case1$: $X \sim PP(1, \lambda_1, \lambda_2)$ *and* $Y \sim PP(1, \lambda_1, {\lambda_2}')$ *then* $Q_X(u)/Q_Y(u)$ *is non-increasing in* u *iff* ${\lambda_2}' > \lambda_2$.

$Case2$: $X \sim PP(1, \lambda_1, \lambda_2)$ *and* $Y \sim PP(1, {\lambda_1}', \lambda_2)$ *then* $Q_X(u)/Q_Y(u)$ *is non-increasing in* u *iff* ${\lambda_1}' > \lambda_1$.

$Case3$: $X \sim PP(1, \lambda_1, \lambda_2)$ *and* $Y \sim PP(1, {\lambda_1}', {\lambda_2}')$ *then* $Q_X(u)/Q_Y(u)$ *is non-increasing in* u *iff* ${\lambda_1}' > \lambda_1$ *and* ${\lambda_2}' > \lambda_2$.

In the three cases mentioned above, the scale parameter C is taken as 1, as the Lorenz ordering remains unaffected by scale changes.

3.4 Methods of estimation

In this section, we discuss three quantile-based estimation techniques, namely the L-moment method and two variations of the percentile method, that can be used for estimating the parameters of the PP distribution. In the next section, a simulation study is conducted to evaluate the performance of these three estimation techniques.

3.4.1 Method of L-moments

L-moments estimation involves equating sample L-moments with the corresponding population L-moments to obtain the parameter estimates. Hosking (1990) developed a comprehensive framework and conducted an in-depth study of L-moments. These moments tend to have lower sample variances and are more resistant to the influence of outliers.

Consider a random sample $X_1, X_2, \cdots, X_n$ of size n, then the r^{th} population and sample L-moments are given as

$$L_r = \int_0^1 \sum_{k=0}^{r-1} (-1)^{r-1-k} \binom{r-1}{k} \binom{r-1+k}{k} u^k \, Q(u) du, \tag{3.4.1}$$

and

$$l_r = \frac{1}{r} \sum_{i=1}^{n} \frac{\sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \binom{i-1}{r-k-1} \binom{n-i}{k}}{\binom{n}{r}} x_{i:n}, \tag{3.4.2}$$

where $x_{i:n}$ is the i^{th} order statistic. By equating, $L_r = l_r$, $r = 1, 2, ...$ we get estimates of the parameters. In this case, the number of equations to consider matches the number of parameters in the model.

Nair et al. (2013a) provided the first four L-moments of the PP distribution and are

$$L_1 = C\beta(\lambda_1 + 1, 1 - \lambda_2), \tag{3.4.3}$$

$$L_2 = \frac{C(\lambda_1 + \lambda_2)}{\lambda_1 - \lambda_2 + 2}\beta(\lambda_1 + 1, 1 - \lambda_2), \tag{3.4.4}$$

$$L_3 = \frac{C(\lambda_1^2 + \lambda_2^2 + 4\lambda_1\lambda_2 + \lambda_2 - \lambda_1)}{(\lambda_1 - \lambda_2 + 2)_{(2)}}\beta(\lambda_1 + 1, 1 - \lambda_2), \tag{3.4.5}$$

$$L_4 = \frac{C(\lambda_1 + \lambda_2)(\lambda_1^2 + \lambda_2^2 + 8\lambda_1\lambda_2 - 3\lambda_1 + 3\lambda_2 + 2)}{(\lambda_1 - \lambda_2 + 2)_{(3)}}\beta(\lambda_1 + 1, 1 - \lambda_2). \tag{3.4.6}$$

3.4.2 Method of percentiles I

The method of percentiles I involves equating appropriate sample percentiles with their corresponding population percentiles to obtain the estimates. This estimation method is explained in Section 2.6.1.

For the PP distribution the population percentiles υ_1, υ_2, and υ_3 are given as follows

$$\upsilon_1 = C\,2^{\lambda_2 - \lambda_1}, \tag{3.4.7}$$

$$\upsilon_2 = \frac{C\,(9^{\lambda_1} - 9^{-\lambda_2})}{10^{\lambda_1 - \lambda_2}}, \tag{3.4.8}$$

$$\upsilon_3 = \frac{5^{\lambda_1 - \lambda_2} - 9^{-\lambda_2}}{9^{\lambda_1} - 5^{\lambda_1 - \lambda_2}}. \tag{3.4.9}$$

The three sample statistics used here are the sample median ($\tilde{\upsilon}_1$), the inter-decile range ($\tilde{\upsilon}_2$), and the ratio of left to right tail weight ($\tilde{\upsilon}_3$), respectively. By equating $\tilde{\upsilon}_1 = \upsilon_1$, $\tilde{\upsilon}_2 = \upsilon_2$, and $\tilde{\upsilon}_3 = \upsilon_3$, we can estimate the parameters C, λ_1, and λ_2 of the PP distribution.

3.4.3 Method of percentiles II

The method of percentiles II introduced by Haritha et al. (2008) utilizes quantile-based measures of location, dispersion, skewness, and kurtosis. In this approach, the population's median, quartile deviation, Galton's coefficient of skewness, and Moor's kurtosis, as given in (1.4.1), (1.4.3), (1.4.4) and (1.4.5), are equated to the corresponding sample statistics to derive the estimators.

These measures for the PP distribution are also provided in Nair et al. (2013a) and are

$$M = C\,2^{\lambda_2-\lambda_1}, \tag{3.4.10}$$

$$QD = \frac{C\,4^{\lambda_2-\lambda_1}}{2}\left(3^{\lambda_1} - 3^{-\lambda_2}\right), \tag{3.4.11}$$

$$S = \frac{3^{\lambda_1} + 3^{-\lambda_2} - 2^{\lambda_2-\lambda_1+1}}{3^{\lambda_1} - 3^{-\lambda_2}}, \tag{3.4.12}$$

$$T = \frac{2^{\lambda_2-\lambda_1}\left(7^{\lambda_1} - 5^{\lambda_1}3^{-\lambda_2} + 3^{\lambda_1}5^{-\lambda_2} - 7^{-\lambda_2}\right)}{3^{\lambda_1} - 3^{-\lambda_2}}. \tag{3.4.13}$$

Hence the sample median, quartile deviation, and measure of skewness be equated to (3.4.10), (3.4.11), and (3.4.12), respectively, to obtain the estimates C, λ_1, and λ_2

of the PP distribution.

3.5 Simulation study to compare three estimation techniques

A simulation study is conducted to assess the effectiveness of the method of L-moments, as well as the method of percentiles I and II. The performance of these three competing estimation techniques is evaluated using data generated through the inversion method.

Samples with sizes of 100, 500, 1000, and 3000 are generated from the PP distribution, having parameters $C = 8.326$, $\lambda_1 = 0.319$, and $\lambda_2 = 0.181$. For every sample size, the L-moments, method of percentiles I, and method of percentiles II are used to determine the bias and mean square error (MSE), repeating this procedure 1000 times. The entire simulation is conducted using the R statistical software.

This simulation study shows that as the sample size increases, the absolute bias and mean square error (MSE) in the L-moment and percentile method I tend to decrease. From Table 3.1, it is evident that as the sample size increases, the MSE decreases, while the absolute bias shows a fluctuation at one point for the parameter λ_1 in the method of percentiles II. When comparing the numerical values of bias, none of the estimation methods consistently shows lower bias. However, the method of L-moments generally exhibits a lower MSE at all points compared to the other

Table 3.1: 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

two methods. As a result, the L-moments method provides better estimates for the PP distribution.

3.6 Applications

To illustrate the application of the PP distribution, we use two real datasets. As mentioned in Section 3.5, the L-moment method yields the most accurate estimates for the PP distribution. Therefore, we apply the L-moment estimation method to determine the parameters of the PP distribution.

Table 3.2: Per capita personal income (in dollars) by industries in Indiana

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	

The first dataset is taken from Census Bureau (2007) and is given in Table 3.2. It shows the per capita personal income (in dollars) by various industries in the state of Indiana, United States, for the year 2005. We apply the L-moment estimation method to determine the distribution parameters. Here the sample L-moments are

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

The parameter estimates are obtained by matching the sample L-moments with the corresponding population L-moments given in (3.4.3), (3.4.4), and (3.4.5). The resulting system of three nonlinear equations is solved using the Newton-Raphson method, and the estimates obtained are presented in Table 3.3. The PP distribution is obtained by taking the product of the quantile functions of the Power and Pareto distributions. Hence we fitted the above data to these distributions, and the results are also given in Table 3.3.

Table 3.3: Parameter estimates, chi-square statistic, and p-value of the fitted models for dataset I

Distribution	Parameter estimates	Chi-square statistic	p-value
PP	$\hat{C}$=26104.50 $\hat{\lambda}_1$=0.03399 $\hat{\lambda}_2$=0.10744	9.58381	0.65242
Power	$\hat{\alpha}$=32846.75 $\hat{\lambda}_1$=0.15846	40.71903	< 0.001
Pareto	$\hat{\sigma}$=24475.25 $\hat{\lambda}_2$=0.13679	22.98863	0.02782

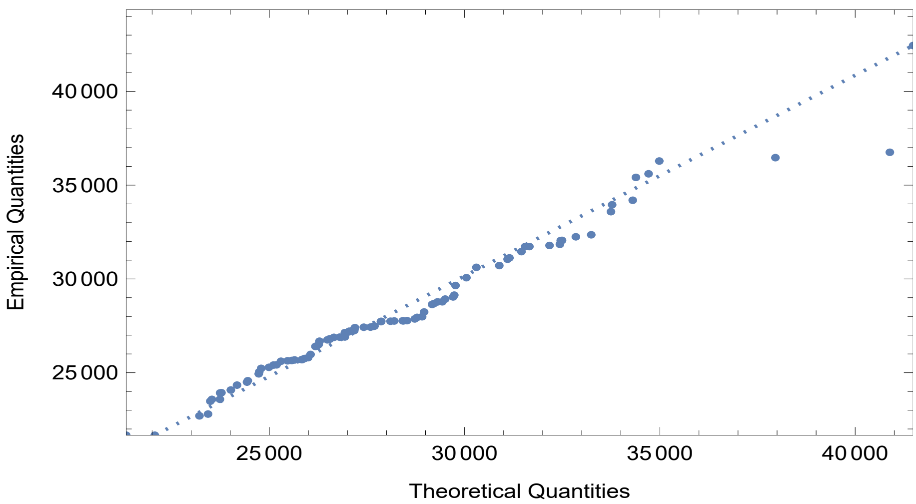


Figure 3.4: Q-Q plot of the per capita personal income by various industries in Indiana state in 2005

Table 3.4: Income inequality indices for the per capita personal income by various industries in Indiana state

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
Theil coefficients	T_1	0.00931	0.00974	0.00043
	T_2	0.00891	0.00920	0.00029
Frigyes measures	ρ	1.08597	1.09389	0.00792
	ν	1.24716	1.23756	0.00960
	η	1.14843	1.13134	0.01709

Two methods are used to evaluate the adequacy of the model. The first is the Q-Q plot, shown in Figure 3.4, demonstrating that the data aligns well with the PP distribution. Additionally, the chi-square test is employed to assess the goodness of fit. From Table 3.3 it is clear that the PP distribution fits the given dataset reasonably well, whereas Power and Pareto do not.

The empirical values and theoretical estimates of the Gini, Bonferroni, Pietra, Atkinson indices, along with generalized entropy, Theil coefficients, and Frigyes measures for the per capita personal income by various industries in Indiana state, are shown in Table 3.4. The empirical values of these inequality measures are computed using R software; for further details, refer to Section 5.2.2. The theoretical estimates of these income inequality measures are calculated for the PP distribution using a quantile-based formula outlined in Section 3.3. To calculate the empirical and theoretical values of the Atkinson index and generalized entropy, we use $\epsilon = 0.5$ and

$\theta = 0.5$, respectively. Table 3.4 shows that the PP distribution provides the closest estimates to the empirical values, with a maximum absolute difference (Abs .diff) of 0.01709. The term 'Abs. diff' in Table 3.4 and 3.7 denotes the absolute difference between the estimated and the empirical value.

The second data is taken from `https://open.dosm.gov.my` and is given in Table 3.5. It contains the mean gross monthly household income of the 40 administrative districts in Sarawak, Malaysia, for the year 2022. We use the method of L-moments to estimate the parameters of the PP distribution. The sample L-moments corresponding to the second dataset are

$$l_1 = 4925.4 \quad l_2 = 709.6526 \quad \text{and} \quad l_3 = 208.8994.$$

Table 3.5: Mean gross monthly household income of the districts in Sarawak

4557	5621	4239	4873	4422	9645	3509	4841	4368	3548
4002	3711	4053	7588	4476	5148	4129	4935	4117	5373
4362	7932	5962	3936	3395	7123	4621	5440	5120	4128
4988	6645	3917	3477	4881	6390	4017	5617	3218	4692

Here, also the parameter estimates are derived by equating the sample L-moments with the corresponding population L-moments. The resulting three nonlinear equations are solved using the Newton-Raphson method. The parameter estimates of the PP distribution along with the Power and Pareto distributions are given in Table 3.6. The Q-Q plot and the chi-square goodness of fit test evaluate how well the model fits the data. The Q-Q plot for the second dataset is shown in Figure 3.5, which indicates

Table 3.6: Parameter estimates, chi-square statistic, and p-value of the fitted models for dataset II

Distribution	Parameter estimates	Chi-square statistic	p-value
PP	$\hat{C}$=4189.68 $\hat{\lambda}_1$=0.07140 $\hat{\lambda}_2$=0.19846	4.24811	0.83407
Power	$\hat{\alpha}$=6583.62 $\hat{\lambda}_1$=0.33667	14.47853	0.07011
Pareto	$\hat{\sigma}$=3684.84 $\hat{\lambda}_2$=0.25187	9.43194	0.30718

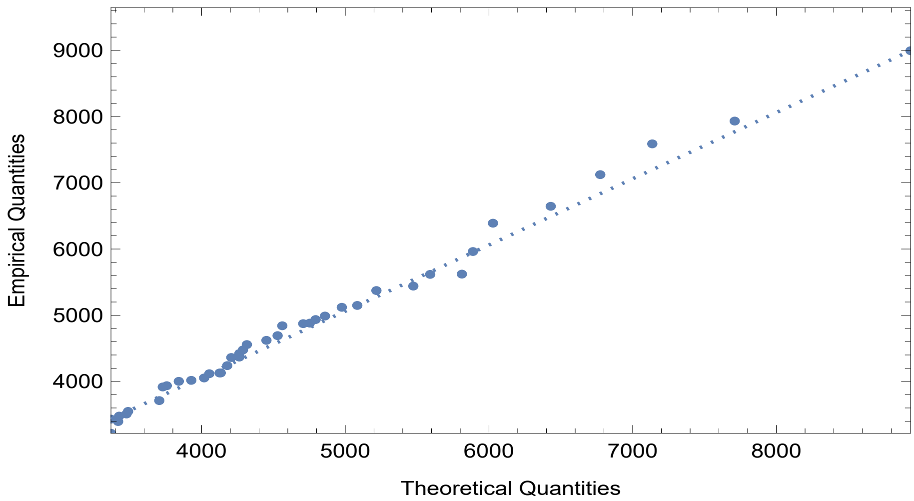


Figure 3.5: Q-Q plot of the mean gross monthly household income of the districts in Sarawak state in 2022

that the data closely follow the PP distribution. In addition, the chi-square value and p-value from Table 3.6 indicate that the PP distribution provides a good fit for the dataset. Figure 3.6 shows the histogram of the data together with the density functions of the Power, Pareto, and PP distributions. This figure clearly shows that the PP distribution gives a better fit to the dataset than the other two distributions.

Table 3.7 presents empirical values and the theoretical estimates for various inequality measures, including the Gini, Bonferroni, Pietra, and Atkinson indices, as

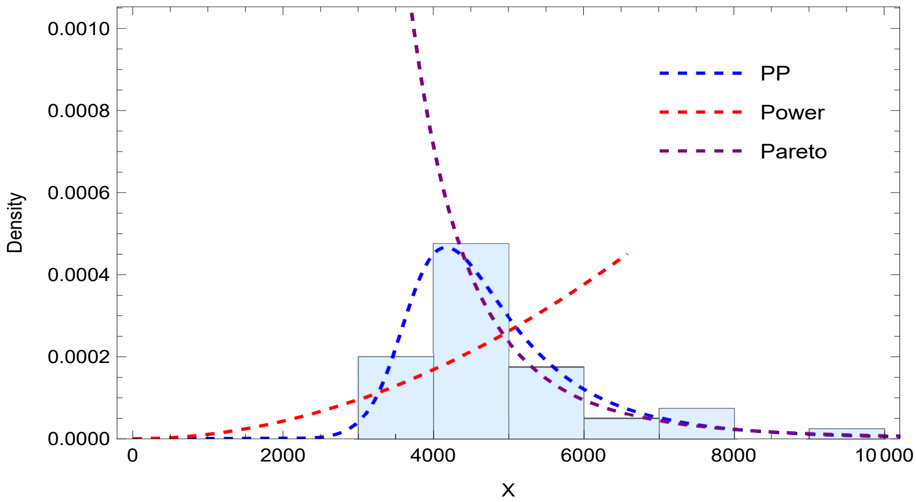


Figure 3.6: The densities of the Power, Pareto, and PP distributions fitted to the mean gross monthly household income of the districts in Sarawak state in 2022

Table 3.7: Income inequality indices for the mean gross monthly household income of Sarawak state

Income inequality		Empirical	Estimated	Abs. diff
Gini index		0.14048	0.14408	0.00360
Bonferroni index		0.18587	0.19707	0.01120
Pietra index		0.09972	0.10193	0.00221
Atkinson index		0.01617	0.01810	0.00193
Generalized entropy		0.03248	0.03637	0.00389
Theil coefficients	T_1	0.03366	0.03866	0.00500
	T_2	0.03157	0.03472	0.00315
Frigyes measures	ρ	1.18984	1.19447	0.00463
	ν	1.50623	1.52006	0.01383
	η	1.26591	1.27259	0.00668

well as generalized entropy, Theil coefficients, and Frigyes measures, based on the average gross monthly household income of districts in Sarawak state. From Table 3.7, it is clear that the PP distribution provides the closest approximations to the empirical values, with a maximum absolute difference (Abs. diff) of 0.01383.

3.7 Summary

In this chapter, we conducted a quantile-based income analysis of the PP distribution. We derived significant income inequality curves and measures of the PP distribution and examined its Lorenz ordering. Through a simulation study, we assessed the efficiency of three different estimation techniques, finding that the L-moments method performed better than the others. Additionally, we applied the model to two income datasets and computed both empirical and theoretical income inequality indices.

CHAPTER 4

SINGH MADDALA DAGUM DISTRIBUTION

4.1 Introduction

[1] The distribution and quantile functions are equivalent methods for modeling and analyzing statistical data. However, the quantile function has certain properties that the distribution function lacks. Notably, the sum of two quantile functions and the product of two positive quantile functions results in another valid quantile function.

This chapter introduces a new quantile function that is particularly useful for analyzing income data. Given the flexibility and widespread use of the SM and Dagum distributions in income modeling, we propose the Singh Maddala Dagum (SMD) distribution, which is formed by adding the quantile functions of these two

[1]A part of this chapter has been published as an article titled "Singh Maddala Dagum Distribution with Application to Income Data" in the journal "Statistics and Applications" (See Ashlin and Haritha (2024)).

distributions.

The SM distribution initially introduced by Singh and Maddala in 1975 and later refined in 1976, has gathered special interest within the field of income distributions. This distribution is a specific case of the GB2 distribution and is also recognized as the Burr XII, or simply the Burr distribution. For an in-depth analysis of the SM distribution, see the works of Kleiber and Kotz (2003), Shahzad and Asghar (2013b), and Kumar (2017). The distribution and quantile functions of the SM distribution are specified as

$$G(x) = 1 - \left[1 + \left(\frac{x}{b}\right)^{a}\right]^{-q}, \quad x > 0, \tag{4.1.1}$$

and

$$Q_1(u) = b\left[(1-u)^{-\frac{1}{q}} - 1\right]^{\frac{1}{a}}, \quad 0 < u < 1, \tag{4.1.2}$$

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

The Dagum distribution, introduced by Dagum (1977), also belongs to the class of GB2 distributions and is known as the Burr III distribution. It finds broad applications across various fields, including reliability, meteorology, quality control, insurance, business failure data, and income modeling. For further insights into the Dagum distribution, see Kleiber and Kotz (2003), Zenga et al. (2012), and Shahzad and Asghar (2013a). The distribution and quantile functions of the Dagum distribution are specified as

$$H(x) = \left[1 + \left(\frac{x}{b}\right)^{-a}\right]^{-p}, \quad x > 0, \tag{4.1.3}$$

and

$$Q_2(u) = b\left[u^{-\frac{1}{p}} - 1\right]^{-\frac{1}{a}}, \quad 0 < u < 1, \tag{4.1.4}$$

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

The rest of the chapter is organized as follows. Section 4.2 introduces the SMD distribution along with its fundamental characteristics. Section 4.3 explores several well-known distributions within the proposed class or derived through relevant transformations on the proposed quantile function. The distributional properties, including skewness, kurtosis, L-moments, and order statistics are discussed in Section 4.4. Section 4.5 focuses on key income inequality and poverty measures associated with the proposed class. Section 4.6 illustrates the application of the SMD distribution to a real income dataset. Finally, the overall conclusions of the study are summarized in Section 4.7.

4.2 SMD quantile function

Motivated by the property that the sum of two quantile functions results in another valid quantile function, we introduce a class of distributions, with the following quantile function

$$Q(u) = b\left[\left((1-u)^{-\frac{1}{q}} - 1\right)^{\frac{1}{a}} + \left(u^{-\frac{1}{p}} - 1\right)^{-\frac{1}{a}}\right], \quad 0 < u < 1, \tag{4.2.1}$$

where a, b, p, $q > 0$. It is obtained by taking the sum of quantile functions of SM and Dagum distributions given in (4.1.2) and (4.1.4) respectively. 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 given as

$$\begin{aligned} q(u) &= \frac{d\,Q(u)}{du} \\ &= b\left[\frac{(1-u)^{-\frac{1}{q}-1}\left((1-u)^{-\frac{1}{q}}-1\right)^{\frac{1}{a}-1}}{aq} + \frac{u^{-\frac{1}{p}-1}\left(u^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}-1}}{ap}\right]. \end{aligned} \tag{4.2.2}$$

The density and distribution functions for the class of distributions defined in (4.2.1) are not available in closed form. However, they can be obtained through numerical inversion of the quantile function. The density function $f(x)$ for the SMD distribution is expressed in terms of the distribution function as follows

$$f(x) = \frac{1}{b}\left[\frac{apq\;\;F(x)^{\frac{1}{p}+1}(1-F(x))^{\frac{1}{q}+1}}{pF(x)^{\frac{1}{p}+1}[(1-F(x))^{-\frac{1}{q}}-1]^{\frac{1}{a}-1} + q(1-F(x))^{\frac{1}{q}+1}(F(x)^{-\frac{1}{p}}-1)^{-\frac{1}{a}-1}}\right]. \tag{4.2.3}$$

Figure 4.1 shows the density function plot for different parameter combinations. The plot demonstrates that, depending on the parameter values, the family encompasses decreasing, unimodal, positively skewed, and negatively skewed models.

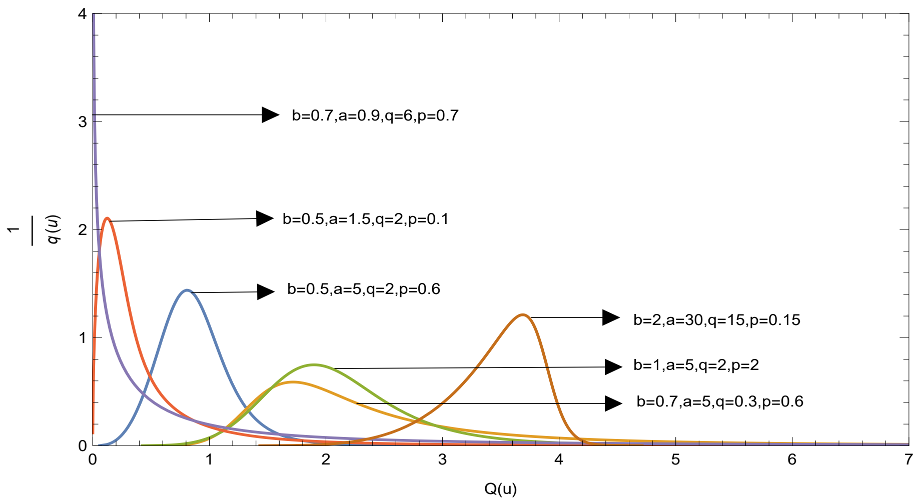


Figure 4.1: Plots of density function for different values of parameters

4.3 Members of the family

Several popular distributions can be derived from the proposed class given in (4.2.1), for various parameter values and by applying specific transformations outlined in Gilchrist (2000).

Case 1. $b > 0,\ q > 0,\ a = 1$ and $p \to 0$

The quantile function of the proposed class tends to the Lomax distribution and is expressed as

$$Q(u) = b\left[(1-u)^{-\frac{1}{q}} - 1\right]. \tag{4.3.1}$$

Case 2. $b > 0,\ a > 0,\ q = 1$ and $p \to 0$

The quantile function of the proposed class tends to the Fisk distribution and is expressed as

$$Q(u) = b\left[(1-u)^{-1} - 1\right]^{\frac{1}{a}}. \tag{4.3.2}$$

Case 3. $b > 0, a = q$ and $p \to 0$

The quantile function of the proposed class tends to the Paralogistic distribution and is expressed as

$$Q(u) = b\left[(1-u)^{-\frac{1}{a}} - 1\right]^{\frac{1}{a}}. \tag{4.3.3}$$

By implementing a reciprocal transformation to (4.3.3), we obtain the inverse Paralogistic distribution with the corresponding quantile function

$$Q(u) = \frac{1}{Q(1-u)} = k\left(u^{-\frac{1}{a}} - 1\right)^{-\frac{1}{a}}, \tag{4.3.4}$$

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

The theorems below establish the relationships between the random variables corresponding to the SM, SMD, and Dagum distributions.

Theorem 4.3.1. *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. Suppose $Q^*(u) = Q_S(u) + Q_R(u),$ 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).

Consider that $V \sim SM(a,b,q)$ and $W \sim Dagum(a,b,p),$ then $V + Q_W(F_V(V))$

has $SMD(a, b, p, q)$ distribution by the above result.

We have, $Q_W(u) = b\left(u^{-\frac{1}{p}} - 1\right)^{-\frac{1}{a}}$ and $F_V(V) = 1 - \left[1 + \left(\frac{V}{b}\right)^a\right]^{-q}$.

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

Theorem 4.3.2. *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 (4.3.1). □

4.4 Distributional characteristics

Quantile functions simplify the process of describing a distribution through its moments. As a result, quantile-based measures are commonly used in statistical analysis to assess distributional characteristics such as location, dispersion, skewness, and kurtosis. These measures can then be applied to estimate model parameters by equating the population's characteristics with the corresponding sample characteristics.

4.4.1 Measures of location, spread and shape

The mean of the SMD distribution is

$$\mu = b\left[\frac{\Gamma\left(1 + \frac{1}{a}\right)\Gamma\left(q - \frac{1}{a}\right)}{\Gamma(q)} + \frac{\Gamma\left(p + \frac{1}{a}\right)\Gamma\left(1 - \frac{1}{a}\right)}{\Gamma(p)}\right]. \tag{4.4.1}$$

The median of the SMD distribution is

$$\begin{aligned} M &= Q(0.5) \\ &= b\left[\left(2^{\frac{1}{q}}-1\right)^{\frac{1}{a}}+\left(2^{\frac{1}{p}}-1\right)^{-\frac{1}{a}}\right]. \end{aligned} \tag{4.4.2}$$

The interquartile range is

$$\begin{aligned} IQR &= Q\left(0.75\right)-Q\left(0.25\right) \\ &= b\Bigg\{\left[(0.25)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}-\left[(0.75)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}} \\ &+\left[(0.75)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}-\left[(0.25)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}\Bigg\}. \end{aligned} \tag{4.4.3}$$

The Galton's skewness and Moors kurtosis measures are expressed in (4.4.4) and (4.4.5) respectively.

$$\begin{aligned} S &= \frac{Q\left(0.25\right)+Q\left(0.75\right)-2M}{IQR} \\ &= \frac{S_1+S_2}{\left[(0.25)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}-\left[(0.75)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}+\left[(0.75)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}-\left[(0.25)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}}, \end{aligned} \tag{4.4.4}$$

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

$$
\begin{aligned}
T &= \frac{Q\left(0.875\right) - Q\left(0.625\right) + Q\left(0.375\right) - Q\left(0.125\right)}{IQR} \\
&= \frac{T_1 + T_2}{\left[(0.25)^{-\frac{1}{q}} - 1\right]^{\frac{1}{a}} - \left[(0.75)^{-\frac{1}{q}} - 1\right]^{\frac{1}{a}} + \left[(0.75)^{-\frac{1}{p}} - 1\right]^{-\frac{1}{a}} - \left[(0.25)^{-\frac{1}{p}} - 1\right]^{-\frac{1}{a}}},
\end{aligned} \tag{4.4.5}
$$

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

4.4.2 L-moments

The definitions and formulas for calculating L-moments are discussed in Section 1.4.4. Hence the first four L-moments of the SMD distribution are as follows

$$L_1 = b\left[A_1 O_1 + A_2 R_1\right], \tag{4.4.6}$$

$$L_2 = b\left[A_1\left(O_1 - O_2\right) - A_2\left(R_1 - R_2\right)\right], \tag{4.4.7}$$

$$L_3 = b\left[A_1\left(O_1 - 3O_2 + 2O_3\right) + A_2\left(R_1 - 3R_2 + 2R_3\right)\right], \tag{4.4.8}$$

$$L_4 = b\left[A_1\left(O_1 - 6O_2 + 10O_3 - 5O_4\right) - A_2\left(R_1 - 6R_2 + 10R_3 - 5R_4\right)\right], \tag{4.4.9}$$

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 $i = 1, 2, 3, 4$.

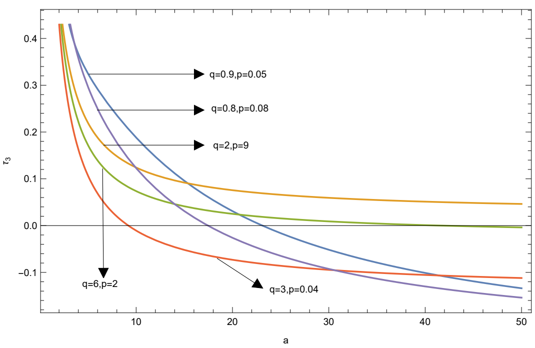
The L-coefficient of variation (τ_2), skewness (τ_3), and kurtosis (τ_4) for the SMD distribution are presented in (4.4.10), (4.4.11), and (4.4.12), respectively.

$$
\begin{aligned}
\tau_2 &= \frac{L2}{L1} \\
&= \frac{A_1\left(O_1 - O_2\right) - A_2\left(R_1 - R_2\right)}{A_1 O_1 + A_2 R_1}.
\end{aligned}
\tag{4.4.10}
$$

$$
\begin{aligned}
\tau_3 &= \frac{L3}{L2} \\
&= \frac{A_1\left(O_1 - 3O_2 + 2O_3\right) + A_2\left(R_1 - 3R_2 + 2R_3\right)}{A_1\left(O_1 - O_2\right) - A_2\left(R_1 - R_2\right)}.
\end{aligned}
\tag{4.4.11}
$$

$$
\begin{aligned}
\tau_4 &= \frac{L4}{L2} \\
&= \frac{A_1\left(O_1 - 6O_2 + 10O_3 - 5O_4\right) - A_2\left(R_1 - 6R_2 + 10R_3 - 5R_4\right)}{A_1\left(O_1 - O_2\right) - A_2\left(R_1 - R_2\right)}.
\end{aligned}
\tag{4.4.12}
$$

Figures 4.2, 4.3, and 4.4 show the plots of the L-coefficients of skewness (τ_3) and kurtosis (τ_4) for various parameter values. In Figure 4.2, the curve for τ_3 decreases with increasing a when q and p are fixed, while the curve for τ_4 also decreases with a for fixed q and p, provided that $p > 1$.Figure 4.3 illustrates that the curves for both τ_3 and τ_4 increase with p when a and q are kept constant, given that $q \geq 1$. Figure 4.4 represents the curves of τ_3 and τ_4 for fixed values of a and p, when q varies.



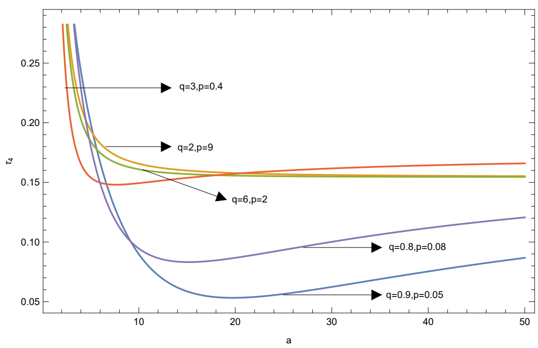


Figure 4.2: Plot of L-coefficients of skewness and kurtosis for particular values of q and p as a function of the parameter a

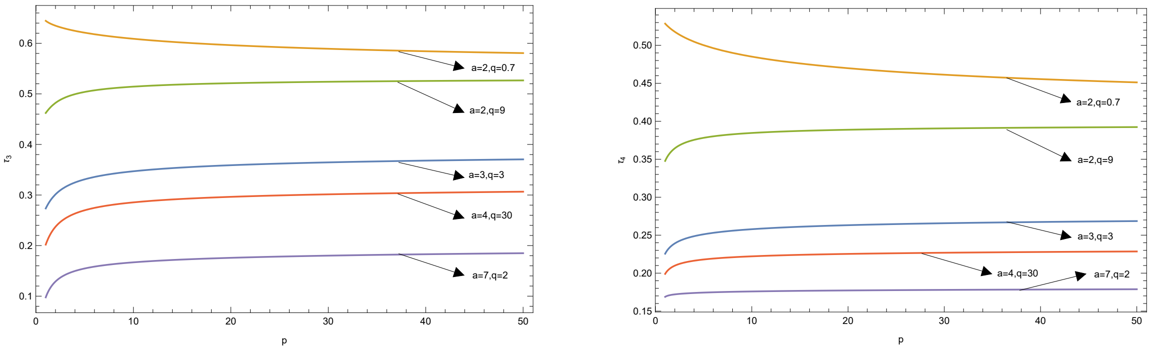


Figure 4.3: Plot of L-coefficients of skewness and kurtosis for particular values of a and q as a function of the parameter p

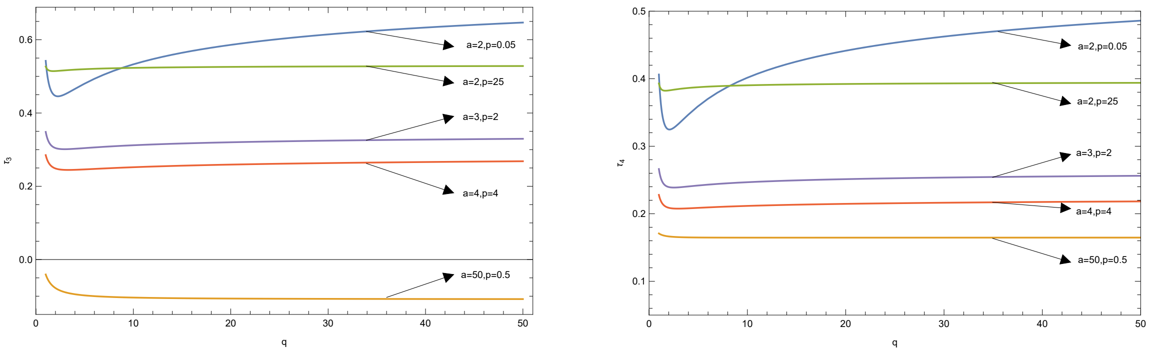


Figure 4.4: Plot of L-coefficients of skewness and kurtosis for particular values of a and p as a function of the parameter q

4.4.3 Order statistics

The density function of the r^{th} order statistic $X_{r:n}$ is expressed as

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

From (4.2.3), we get

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

Hence

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

The above expression can be represented in terms of quantiles as follows

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

For SMD distribution, the quantile function of the first-order statistic $X_{1:n}$ takes the form

$$\begin{aligned} Q_{(1)}(u) &= Q\left[1-(1-u)^{\frac{1}{n}}\right] \\ &= b\left\{\left[(1-u)^{-\frac{1}{nq}}-1\right]^{\frac{1}{a}} + \left[\left(1-(1-u)^{\frac{1}{n}}\right)^{-\frac{1}{p}}-1\right]^{-\frac{1}{a}}\right\}, \qquad (4.4.13) \end{aligned}$$

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

$$\begin{aligned}
Q_{(n)}(u) &= Q\left(u^{\frac{1}{n}}\right) \\
&= b\left\{\left[\left(1-u^{\frac{1}{n}}\right)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}+\left[u^{-\frac{1}{np}}-1\right]^{-\frac{1}{a}}\right\}.
\end{aligned} \tag{4.4.14}$$

4.5 Income inequality and poverty measures

Income inequality and poverty measures are widely discussed topics in statistical and economic literature. An income inequality measure aims to provide an index that quantifies the disparities in income within or between groups, while a poverty measure assesses the extent of poverty experienced by individuals whose income falls below a specified poverty threshold.

4.5.1 Income inequality measures

The Lorenz curve of the SMD distribution based on (1.5.4) is

$$L(u) = \frac{q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a}, q-\frac{1}{a}\right)+p\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}{q\beta\left(1+\frac{1}{a}, q-\frac{1}{a}\right)+p\beta\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)}, \tag{4.5.1}$$

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

The Gini index of the SMD distribution obtained using (1.5.7) is

$$G = 1-2\int_0^1 L(u)du$$

$$= 1 - 2\left[\frac{q\beta\left(1+\frac{1}{a},2q-\frac{1}{a}\right)+p\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)-p\beta\left(2p+\frac{1}{a},1-\frac{1}{a}\right)}{q\beta\left(1+\frac{1}{a},q-\frac{1}{a}\right)+p\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)}\right]. \quad (4.5.2)$$

The Pietra index of the SMD distribution derived using (1.5.15) is

$$P = \frac{u_0 Q\left(u_0\right) - b\left[q\beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a},q-\frac{1}{a}\right)+p\beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a},1-\frac{1}{a}\right)\right]}{\mu}, \quad (4.5.3)$$

where $Q(u_0) = b\left[\left((1-u_0)^{-\frac{1}{q}}-1\right)^{\frac{1}{a}}+\left(u_0^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}\right]$ and mean μ is given in (4.4.1).

The Bonferroni curve of the SMD distribution based on (1.5.20) is

$$B(u) = \frac{q\beta_{1-(1-u)^{\frac{1}{q}}}\left(1+\frac{1}{a},q-\frac{1}{a}\right)+p\beta_{u^{\frac{1}{p}}}\left(p+\frac{1}{a},1-\frac{1}{a}\right)}{u\left[q\beta\left(1+\frac{1}{a},q-\frac{1}{a}\right)+p\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)\right]}. \quad (4.5.4)$$

The Frigyes measures ρ, ν and η of the SMD distribution obtained using (1.5.31)-(1.5.33) are as follows.

$$\rho = \frac{u_0\left[\left((1-u_0)^{-\frac{1}{q}}-1\right)^{\frac{1}{a}}+\left(u_0^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}\right]}{q\beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a},q-\frac{1}{a}\right)+p\beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a},1-\frac{1}{a}\right)}, \quad (4.5.5)$$

$$\nu = \frac{u_0}{(1-u_0)}\left[\frac{q\beta\left(1+\frac{1}{a},q-\frac{1}{a}\right)+p\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)}{q\beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a},q-\frac{1}{a}\right)+p\beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a},1-\frac{1}{a}\right)}-1\right], \quad (4.5.6)$$

$$\eta = \frac{q\,\mathfrak{w}_1 + p\,\mathfrak{w}_2}{(1-u_0)\left[\left((1-u_0)^{-\frac{1}{q}}-1\right)^{\frac{1}{a}}+\left(u_0^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}\right]}, \quad (4.5.7)$$

where

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

The Zenga curve of the SMD distribution based on (1.5.39) is given as

$$I(u) = \frac{q\,\mathfrak{z}_1 + p\,\mathfrak{z}_2}{q\,\mathfrak{z}_3 + p\,\mathfrak{z}_4}, \tag{4.5.8}$$

where

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

Our proposed model SMD distribution does not have closed-form expressions for the Bonferroni index, Atkinson index, and generalized entropy measures. Figures 4.5, 4.6, and 4.7 display the Lorenz, Bonferroni, and Zenga curves of the SMD distribution.

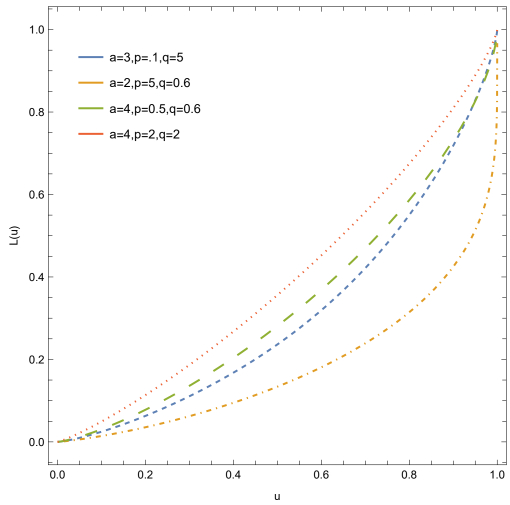


Figure 4.5: Graph of SMD Lorenz curve

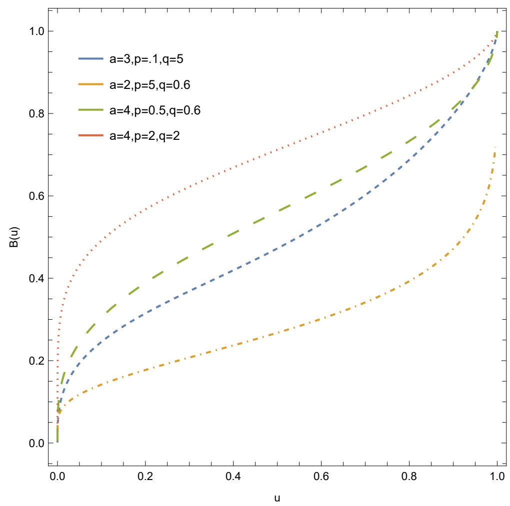


Figure 4.6: Graph of SMD Bonferroni curve

4.5.2 Poverty measures

Poverty measures are mainly utilized to monitor socioeconomic progress and establish benchmarks for achievement or setbacks. Most poverty measures can be described as the average level of deprivation experienced by the poor. Section 1.7 outlines the key poverty measures and provides their formulas.

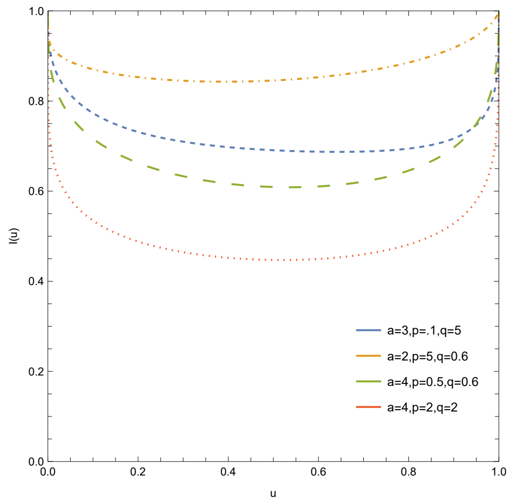


Figure 4.7: Graph of SMD Zenga curve

In this section, we derive quantile-based formulas for the poverty gap ratio and FGT measure. By applying the transformation $F(z) = u$ and $F(y) = p$, where $0 < u < 1$ and $0 < p < 1$, in (1.7.4), the poverty gap ratio can be expressed in quantile form as follows

$$PG = \int_0^u \left(\frac{Q(u) - Q(p)}{Q(u)} \right) dp. \tag{4.5.9}$$

The poverty gap ratio is also called the income gap ratio of the poor in Nair et al. (2008). It can be expressed using the reversed mean residual quantile function as follows

$$PG = \frac{uR(u)}{Q(u)}, \tag{4.5.10}$$

where $R(u) = u^{-1} \int_0^u (Q(u) - Q(p))\, dp$.

The FGT measure defined in (1.7.5) can be expressed in quantile form by applying the same transformation used for the poverty gap ratio. It is obtained as

$$FGT(\alpha) = \int_0^u \left(\frac{Q(u) - Q(p)}{Q(u)} \right)^{\alpha} dp. \tag{4.5.11}$$

where $\alpha \geq 0$ represents the inequality aversion parameter.

Now the poverty gap ratio for SMD distribution is derived using (4.5.9) and is

$$PG = u - \frac{q\beta_{1-(1-u)^{\frac{1}{q}}} \left(1 + \frac{1}{a}, q - \frac{1}{a}\right) + p\beta_{u^{\frac{1}{p}}} \left(p + \frac{1}{a}, 1 - \frac{1}{a}\right)}{\left((1-u)^{-\frac{1}{q}} - 1\right)^{\frac{1}{a}} + \left(u^{-\frac{1}{p}} - 1\right)^{-\frac{1}{a}}}. \tag{4.5.12}$$

Furthermore, the FGT measure lacks a closed-form expression for the SMD distribution.

The Sen index proposed by Sen (1976) is expressed in quantile form as

$$S(u) = u \left(\frac{u\,\omega_1'(u) + \omega_2(u)}{u\,\omega_1'(u) + \omega_1(u)} \right), \tag{4.5.13}$$

where $\omega_1(u) = \frac{1}{u}\int_0^u Q(p)dp$, and $\omega_2(u) = \frac{1}{u^2}\int_0^u (2p - u)\, Q(p)dp$. Here $\omega_1'(u)$ indicates the derivative of ω_1 with respect to u. For SMD distribution $\omega_1(u)$ and $\omega_2(u)$ are given as

$$\begin{aligned}
\omega_1(u) &= \frac{b}{u}\left[q\beta_{1-(1-u)^{\frac{1}{q}}} \left(1 + \frac{1}{a}, q - \frac{1}{a}\right) + p\beta_{u^{\frac{1}{p}}} \left(p + \frac{1}{a}, 1 - \frac{1}{a}\right) \right], \\
\omega_2(u) &= \frac{b}{u^2}\Bigg[(2-u)q\beta_{1-(1-u)^{\frac{1}{q}}} \left(1 + \frac{1}{a}, q - \frac{1}{a}\right) - 2q\beta_{1-(1-u)^{\frac{1}{q}}} \left(1 + \frac{1}{a}, 2q - \frac{1}{a}\right) \\
&\quad + 2p\beta_{u^{\frac{1}{p}}} \left(2p + \frac{1}{a}, 1 - \frac{1}{a}\right) - up\beta_{u^{\frac{1}{p}}} \left(p + \frac{1}{a}, 1 - \frac{1}{a}\right) \Bigg].
\end{aligned}$$

The truncated Gini index for the poor in (1.5.11) can also be written as

$$G(u) = 1 - \frac{2}{\omega_1(u)} \int_0^u Q(p) \left(\frac{u-p}{u^2} \right) dp$$
$$= \frac{\omega_2(u)}{\omega_1(u)}. \tag{4.5.14}$$

Now, for SMD distribution, the truncated Gini index for the poor is obtained as

$$G(u) = \frac{2}{u} \times \frac{A}{B} - 1, \tag{4.5.15}$$

where

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

The Watts index, defined in (1.7.7), and a measure introduced by Kakwani (1999), expressed as $K^* = 1 - e^{-W}$, do not have simple algebraic expressions for the SMD distribution.

4.6 Applications

In this section, we utilize a real-life dataset to estimate the parameters of the class of distribution described in (4.2.1) and evaluate the model's performance and practical applications. The L-moments estimation method described in Section 3.4.1 is used to estimate the parameters of this distribution.

The data, sourced from `https://ihds.umd.edu`, contains per-capita household income data for Himachal Pradesh from the IHDS-II. The SMD distribution is obtained by summing the quantile functions of the SM and Dagum distributions. Accordingly, the dataset is fitted to the SM, SMD, and Dagum distributions, with the parameter estimates provided in Table 4.1.

Table 4.1: Parameter estimates, chi-square statistic, and p-value of the fitted models for the dataset

Distribution	Parameter estimates	Chi-square statistic	p-value
SMD	$\hat{b} = 61372.5$ $\hat{a} = 2.05570$ $\hat{p} = 0.11765$ $\hat{q} = 6.03782$	14.77930	0.99933
SM	$\hat{b} = 33195.5$ $\hat{a} = 1.67809$ $\hat{q} = 1.27762$	34.60771	0.53479
Dagum	$\hat{b} = 34745.4$ $\hat{a} = 2.04751$ $\hat{p} = 0.72442$	38.84004	0.34294

Here, the Q-Q plot and the chi-square goodness of fit test are used to examine the adequacy of the model. The Q-Q plot in Figure 4.8 shows that the SMD distribution fits the data well. Additionally, the chi-square statistics and their corresponding p-

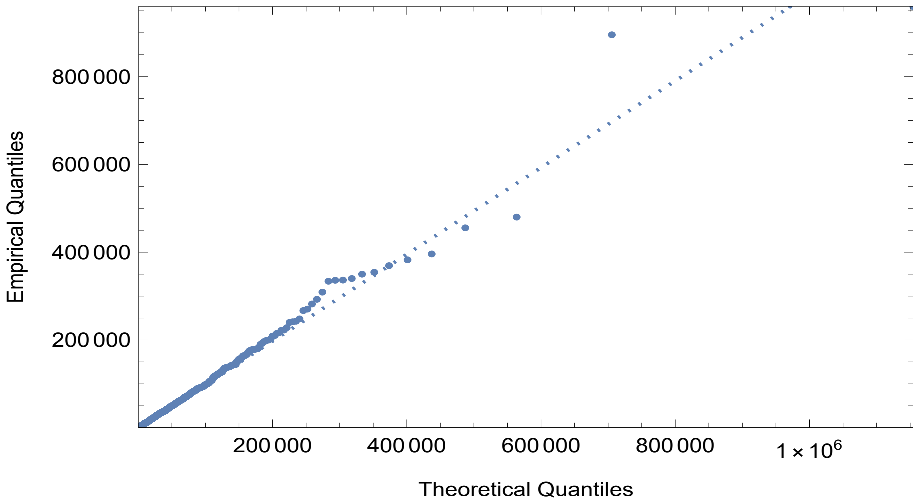


Figure 4.8: Q-Q plot of per-capita household income in Himachal Pradesh based on IHDS-II

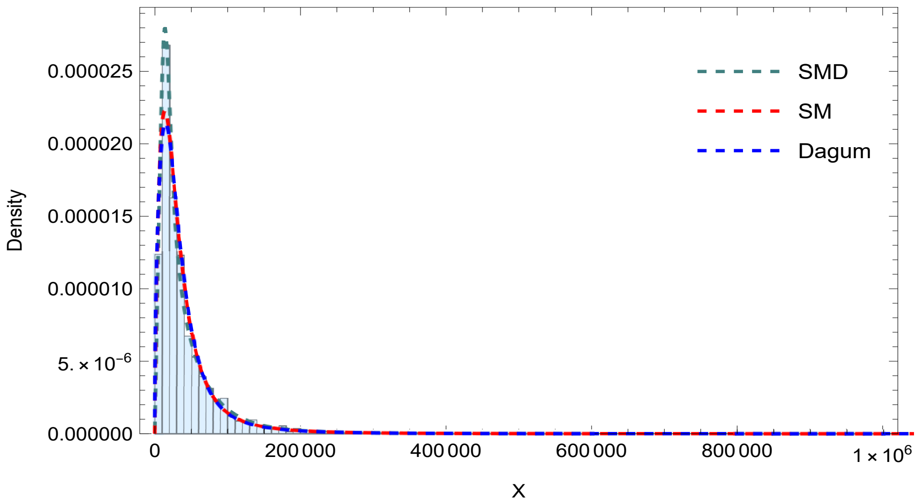


Figure 4.9: The densities of the SM, SMD, and Dagum distributions fitted to the per-capita household income of Himachal Pradesh from IHDS-II

values in Table 4.1 indicate that the SMD distribution offers a good fit to the data. Similarly, the SM and Dagum distributions also provide an acceptable fit. Figure 4.9 displays the histogram of the data along with the density functions for the SM, SMD, and Dagum distributions. This figure clearly shows that the SMD distribution fits the dataset more accurately than the other two models.

4.7 Summary

We introduce the SMD distribution by adding the quantile functions of the SM and Dagum distributions. The proposed class includes several well-known distributions. We studied the distributional properties as well as the major income inequality, and poverty measures of this class. We also derived quantile-based formulas for poverty measures, such as the poverty gap ratio and the FGT measure. Parameter estimation of the model is performed using the L-moments method. The proposed distribution is applied to a real income dataset and demonstrates a superior fit compared to the SM and Dagum distributions.

CHAPTER 5

A COMPARATIVE ANALYSIS OF INCOME INEQUALITY ACROSS INDIAN STATES USING QUANTILE FUNCTIONS

5.1 Introduction

[1] The modeling of income data began with the pioneering work of Pareto (1897), and since then, income distribution has remained a widely discussed topic. Research has shown that while the Pareto distribution effectively models high incomes above a certain threshold, it performs poorly for lower income levels. The lognormal distribution, initially studied in the Gibrat (1931) thesis and further examined by Aitchison

[1] A part of this chapter has been published as an article titled "Comparison of Income Inequality Among Indian States Using Quantile Functions" in the journal "Computational Economics" (See Ashlin and Haritha (2025)).

and Brown (1957), is more suitable for modeling lower incomes.

March (1898) primarily applied the gamma distribution to model income data. Amoroso (1925) applied a four-parameter generalized gamma distribution to model the Prussian incomes of the year 1912. Similarly, Bartels (1977) used a three-parameter generalized gamma distribution to analyze the fiscal incomes of three regions in the Netherlands. In another study, Vartia and Vartia (1980) modeled the Finnish taxed income of 1967 using a four-parameter shifted beta 2 distribution. For income data modeling, McDonald (1984), Butler and McDonald (1989), and others used the GB2 distribution.

The significant contributions to income data modeling and analysis are well documented in the influential works by Arnold (1983) and Kleiber and Kotz (2003). Recent income models based on distribution and quantile functions are reviewed in Section 2.2 and 2.3 respectively.

Most studies on economic inequality in India are based on consumption expenditure data rather than income data. In developing countries, obtaining data on household or individual incomes is always challenging. However, the India Human Development Survey (IHDS) (Desai and Vanneman, 2018, Desai et al., 2018) has made income data publicly available. Despite this, little research has analyzed income inequality variations within Indian states over time. To address this gap, Tyagi (2023) used income data from IHDS-I and IHDS-II to estimate the empirical Gini coefficient, Atkinson index, and generalized entropy measures for Indian states at two different periods.

The literature lacks a comparative study of quantile-based income distribution models across countries. Additionally, some probability distributions used in income modeling do not have a closed-form distribution function but do have a quantile function. Therefore, in this study, we modeled per-capita household income data from Indian states using six parametric models with closed-form quantile functions. We then computed the empirical and theoretical income inequality measures for each state.

The rest of the chapter is organized as follows: Section 5.2 addresses the quantile-based income models and income inequality measures employed in this study. Section 5.3 explains the estimation method used. The applications to per-capita household income data are presented in Section 5.4. Finally, Section 5.5 wraps up the study with concluding remarks.

5.2 Quantile-based income models and income inequality measures

5.2.1 Quantile-based income models

This study examines six parametric distributions, such as Weibull, PP, SM, Dagum, SMD, and MLF, for modeling the per capita household income data of Indian states. While all these distributions possess closed-form quantile functions, only the Weibull, SM, and Dagum distributions have closed-form distribution functions.

Weibull distribution (Weibull (1939a,b)) was initially introduced to model the distribution of material breaking strengths. The Weibull distribution may be relatively lesser-known in economics, but its applicability for modeling income data and insurance losses has been highlighted by D'Addario (1974) and Hogg and Klugman (1983), respectively. Its quantile function is given as

$$Q(u) = \beta \left(-\log \left(1-u\right)\right)^{\frac{1}{\alpha}}, \quad 0 \leq u \leq 1, \tag{5.2.1}$$

where, $\alpha, \beta > 0$.

The quantile-based income analysis of PP distribution is discussed in Chapter 3, and its quantile function is outlined in 3.2.1. In Chapter 4, we explored the SM, Dagum, and SMD distributions, with their respective quantile functions presented in (4.1.2), (4.1.4), and (4.2.1). Freimer et al. (1988) introduced the MLF, and its application in income modeling is further explored by Haritha et al. (2008). Its quantile function is given in (2.3.2).

5.2.2 Income inequality measures

Income data can be presented either as individual observations or in groups. For samples with individual observations, income inequality can be assessed using parametric and non-parametric methods. In this chapter, we apply both approaches to calculate income inequality measures.

In the non-parametric approach, the empirical formulas for calculating income inequality measures such as the Gini index, Pietra index, Bonferroni index, Atkinson index, generalized entropy, and Frigyes measures are provided in (5.2.2)-(5.2.8). For more information, see Kleiber and Kotz (2003), Bakar and Dharini (2020), and Tyagi (2023). Here, x_i represents the income of the i^{th} unit, $\bar{x}$ stands for the mean income, and n is the total number of observations. Additionally, the empirical Bonferroni index is determined after arranging the observations in ascending order. If the data consists of non-zero, positive, and ascendingly sorted observations, the Gini coefficient can be easily computed using (5.2.3); otherwise, (5.2.2) should be used.

$$\widehat{G} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2\, \bar{x}} \tag{5.2.2}$$

$$\widehat{G} = \frac{2}{n^2 \bar{x}} \sum_{i=1}^{n} i\,(x_i - \bar{x}) \tag{5.2.3}$$

$$\widehat{P} = \frac{\sum_{i=1}^{n} |x_{i-}\bar{x}|}{2n\bar{x}} \tag{5.2.4}$$

$$\widehat{B} = \frac{\sum_{i=1}^{n} \left(x_i - \sum_{j=1}^{i} \frac{x_j}{i} \right)}{\sum_{i=1}^{n} x_i} \tag{5.2.5}$$

$$\widehat{A}_\epsilon = 1 - \frac{1}{\bar{x}} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \tag{5.2.6}$$

where $0 \leq \epsilon \neq 1$.

$$\widehat{GE}_\theta = \frac{1}{n\,\theta\,(\theta - 1)} \sum_{i=1}^{n} \left[\left(\frac{x_i}{\bar{x}} \right)^{\theta} - 1 \right] \tag{5.2.7}$$

where $\theta \neq 0\,\&\,1$.

The three Frigyes measures are given as follows

$$\begin{aligned}
\widehat{\rho} &= \frac{E\left(X\right)}{E(X|X<m)} = \left(\frac{m-1}{n}\right)\frac{\sum_{i=1}^{n}x_i}{\sum_{i=1}^{m-1}x_i},\\
\widehat{\nu} &= \frac{E\left(X|X\geq m\right)}{E(X|X<m)} = \left(\frac{m-1}{n-m}\right)\frac{\sum_{i=m}^{n}x_i}{\sum_{i=1}^{m-1}x_i},\\
\widehat{\eta} &= \frac{E\left(X|X\geq m\right)}{E(X)} = \left(\frac{n}{n-m}\right)\frac{\sum_{i=m}^{n}x_i}{\sum_{i=1}^{n}x_i}.
\end{aligned} \tag{5.2.8}$$

In contrast, the parametric approach requires theoretical formulas to calculate income inequality measures. This approach involves estimating the distributional parameters and substituting them into the equations for the theoretical measures. Table 5.1 presents quantile-based income inequality measures for the Weibull, SM, and Dagum distributions. The income inequality measures of the PP and SMD distributions are discussed in Section 3.3 and 4.5.1, respectively, while the measures for MLF can be found in Haritha et al. (2008). The MLF lacks closed-form expressions for the Atkinson index and generalized entropy while the SMD distribution lacks closed-form equations for the Bonferroni index, Atkinson index, and generalized entropy. Consequently, numerical integration is performed in Mathematica to compute theoretical inequality measures that do not have explicit formulas.

In Table 5.1, GI, PI, BI, AI, and GE represent the Gini index, Pietra index, Bonferroni index, Atkinson index, and generalized entropy measures, respectively. Additionally, FM_1, FM_2 and FM_3 denote the three Frigyes measures. The value of

Table 5.1: Income inequality measures

Distributions	Income inequality measures	
Weibull	GI	$1-2^{-\frac{1}{\alpha}}$
	PI	$\frac{2\left[u_0(-\log(1-u_0))^{\frac{1}{\alpha}}-\Upsilon\right]+\Gamma\left(1+\frac{1}{\alpha}\right)-(-\log(1-u_0))^{\frac{1}{\alpha}}}{2\,\Gamma\left(1+\frac{1}{\alpha}\right)}$
	BI	$1-\sum_{i=0}^{\infty}\frac{(i+2)^{-\frac{1}{\alpha}-1}}{i+1}$
	AI	$1-\frac{1}{\Gamma\left(1+\frac{1}{\alpha}\right)}\left[\Gamma\left(\frac{1-\epsilon}{\alpha}+1\right)\right]^{\frac{1}{1-\epsilon}}$
	GE	$\frac{1}{\theta(\theta-1)}\left[\frac{\Gamma\left(1+\frac{\theta}{\alpha}\right)}{\left(\Gamma\left(1+\frac{1}{\alpha}\right)\right)^{\theta}}-1\right]$
	FM_1	$\frac{u_0(-\log(1-u_0))^{\frac{1}{\alpha}}}{\Upsilon}$
	FM_2	$\left(\frac{u_0}{1-u_0}\right)\left[\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\Upsilon}-1\right]$
	FM_3	$\frac{\Gamma\left(1+\frac{1}{\alpha}\right)-\Upsilon}{(1-u_0)(-\log(1-u_0))^{\frac{1}{\alpha}}}$
SM	GI	$1-\frac{\Gamma(q)\Gamma(2q-\frac{1}{a})}{\Gamma(q-\frac{1}{a})\Gamma(2q)}$
	PI	$\frac{(2u_0-1)\left[(1-u_0)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}-2q\beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a},q-\frac{1}{a}\right)+q\beta\left(1+\frac{1}{a},q-\frac{1}{a}\right)}{2q\beta\left(1+\frac{1}{a},q-\frac{1}{a}\right)}$
	BI	$1-\frac{a\,\Gamma\left(2+\frac{1}{a}\right)}{(a+1)\,\beta\left(q-\frac{1}{a},1+\frac{1}{a}\right)}\sum_{i=0}^{\infty}\frac{\Gamma\left(q\,i+2q-\frac{1}{a}\right)}{(i+1)\,\Gamma(q\,i+2q+1)}$
	AI	$1-\frac{q^{\frac{\epsilon}{1-\epsilon}}\left[\beta\left(1+\frac{1-\epsilon}{a},q-\frac{1-\epsilon}{a}\right)\right]^{\frac{1}{1-\epsilon}}}{\beta\left(1+\frac{1}{a},q-\frac{1}{a}\right)}$
	GE	$\frac{1}{\theta(\theta-1)}\left[\frac{q^{1-\theta}\beta\left(1+\frac{\theta}{a},q-\frac{\theta}{a}\right)}{\left(\beta\left(1+\frac{1}{a},q-\frac{1}{a}\right)\right)^{\theta}}-1\right]$

Table 5.1: (continued)

Distributions	Income inequality measures	
SM	FM_1	$\frac{u_0\left[(1-u_0)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}}{q\,\beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a},q-\frac{1}{a}\right)}$
	FM_2	$\frac{u_0}{1-u_0}\left[\frac{1}{I_{1-(1-u_0)^{\frac{1}{q}}}\left(1+\frac{1}{a},q-\frac{1}{a}\right)}-1\right]$
	FM_3	$\frac{q\,\mathcal{B}_1}{(1-u_0)\left[(1-u_0)^{-\frac{1}{q}}-1\right]^{\frac{1}{a}}}$
Dagum	GI	$\frac{\Gamma(p)\Gamma(2p+\frac{1}{a})}{\Gamma(2p)\Gamma(p+\frac{1}{a})}-1$
	PI	$\frac{(2u_0-1)\left(u_0^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}-2p\beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a},1-\frac{1}{a}\right)+p\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)}{2p\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)}$
	BI	$1-\frac{a^2p\times {}_3F_2\left(p+\frac{1}{a},p+\frac{1}{a},\frac{1}{a};p+\frac{1}{a}+1,p+\frac{1}{a}+1;1\right)}{(ap+1)^2\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)}$
	AI	$1-\frac{p^{\frac{\epsilon}{1-\epsilon}}\left[\beta\left(p+\frac{1-\epsilon}{a},1-\frac{1-\epsilon}{a}\right)\right]^{\frac{1}{1-\epsilon}}}{\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)}$
	GE	$\frac{1}{\theta(\theta-1)}\left[p^{1-\theta}\frac{\beta\left(p+\frac{\theta}{a},1-\frac{\theta}{a}\right)}{\left(\beta\left(p+\frac{1}{a},1-\frac{1}{a}\right)\right)^{\theta}}-1\right]$
	FM_1	$\frac{u_0\left(u_0^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}}{p\,\beta_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a},1-\frac{1}{a}\right)}$
	FM_2	$\frac{u_0}{1-u_0}\left[\frac{1}{I_{u_0^{\frac{1}{p}}}\left(p+\frac{1}{a},1-\frac{1}{a}\right)}-1\right]$
	FM_3	$\frac{p\,\mathcal{B}_2}{(1-u_0)\left(u_0^{-\frac{1}{p}}-1\right)^{-\frac{1}{a}}}$

u_0 in these inequalities is obtained by solving for u in the equation $\mu = Q(u)$, where μ denotes the mean and $Q(u)$ represents the quantile function of the respective distribution. Here, $\Upsilon = \gamma\left(1 + \frac{1}{\alpha}, -\log\left(1 - u_0\right)\right)$, where $\gamma\left(a, x\right) = a^{-1}x^{a}\,{}_1F_1\left(a; a+1; -x\right)$ and $\Gamma\left(a, x\right) + \gamma\left(a, x\right) = \Gamma a$. The term ${}_3F_2$ denotes a generalized hypergeometric function. Furthurmore, $\mathcal{B}_1$ is defined as $\left[\beta\left(1 + \frac{1}{a}, q - \frac{1}{a}\right) - \beta_{1-(1-u_0)^{\frac{1}{q}}}\left(1 + \frac{1}{a}, q - \frac{1}{a}\right)\right]$ and $\mathcal{B}_2$ is defined as $\left[\beta\left(p + \frac{1}{a}, 1 - \frac{1}{a}\right) - \beta_{u_0^{\frac{1}{p}}}\left(p + \frac{1}{a}, 1 - \frac{1}{a}\right)\right]$.

5.3 Estimation

The main goal of estimation is to find values that closely approximate the true model parameters. Various criteria can ensure this proximity, leading to different estimation methods. One method involves matching the population characteristics of the fitted model with the corresponding sample properties of the data. This includes estimation methods such as probability-weighted moments, L-moments, conventional moments, and percentiles. A second class of estimation methods relies on optimality conditions, aiming to minimize the difference between the fitted model and observed data, or that yield the most probable estimates.

The L-moments estimation technique described in Section 3.4.1 is applied to estimate the parameters of the six models used in this study. The L-moments of the SM, Dagum, and MLF distributions can be found in Shahzad and Asghar (2013b), Shahzad and Asghar (2013a), and Haritha et al. (2008), respectively. Additionally, the L-moments of the PP distribution are presented in Section 3.4.1, and the L-moments of the SMD distribution are provided in Section 4.4.2. Also, we require the first two L-moments of the Weibull distribution to estimate the parameters, α and β in (5.2.1), and it is given as follows

$$L_1 = \beta\,\Gamma\left(1 + \frac{1}{\alpha}\right),$$
$$L_2 = \left(1 - 2^{-\frac{1}{\alpha}}\right)\beta\,\Gamma\left(1 + \frac{1}{\alpha}\right).$$

5.4 Data application

5.4.1 Data

The IHDS is a nationally representative panel survey covering multiple topics across Indian states and union territories, except Andaman & Nicobar, and Lakshadweep. IHDS data can be accessed from `https://ihds.umd.edu`. This analysis uses data from IHDS-II, which includes 42,152 families across 1,503 villages and 971 urban neighborhoods throughout India. Most of these families were re-interviewed from the initial IHDS survey conducted in 2005. IHDS-II employed a comprehensive questionnaire to capture all aspects of income. In this study, we use per-capita household income data from each state and model it with six parametric distributions. Observations with negative or zero income and outliers, comprising about 1.48%, are excluded from the analysis.

Table 5.2: Parameter estimates, chi-square statistic, and p-value for the distributions fitted to the per-capita household income data of Indian states.

State	Distribution	Parameter estimates	Chi-square value	p-value
Jammu & Kashmir	Weibull	$\hat{\alpha}$=1.08330, $\hat{\beta}$= 44100.20	53.45425	0.00177
	PP	$\hat{C}$ =33222.90, $\hat{\lambda}_1$ =0.59451, $\hat{\lambda}_2$ = 0.42898	9.80608	0.99900
	SM	$\hat{b}$ = 37685, $\hat{a}$= 1.77284, $\hat{q}$ = 1.39426	8.94463	0.99957
	Dagum	$\hat{b}$= 38585, $\hat{a}$ =2.29792, $\hat{p}$ = 0.66899	12.02829	0.99418
	SMD	$\hat{b}$ = 25051.90, $\hat{a}$ = 2.31170, $\hat{p}$ = 0.20392, $\hat{q}$ = 1.14204	7.43182	0.99993
	MLF	$\hat{\lambda}_1$ = 24782, $\hat{\lambda}_2$ = 0.000058458, $\hat{\lambda}_3$ = 0.81693, $\hat{\lambda}_4$ = -0.37580	11.19978	0.99678
Himachal Pradesh	Weibull	$\hat{\alpha}$ = 0.95122, $\hat{\beta}$ = 42715.5	153.89400	<0.001
	PP	$\hat{C}$ = 30271, $\hat{\lambda}_1$ = 0.62314, $\hat{\lambda}_2$ = 0.48386	36.77498	0.43281
	SM	$\hat{b}$ = 33195.50, $\hat{a}$ = 1.67809, $\hat{q}$ = 1.27762	34.60771	0.53479
	Dagum	$\hat{b}$ = 34745.40, $\hat{a}$ = 2.04751, $\hat{p}$ = 0.72442	38.84004	0.34294
	SMD	$\hat{b}$ = 61372.50, $\hat{a}$ = 2.05570, $\hat{p}$ = 0.11765, $\hat{q}$ = 6.03782	14.77930	0.99933
	MLF	$\hat{\lambda}_1$ =19887.70, $\hat{\lambda}_2$ = 0.0000517652, $\hat{\lambda}_3$ = 1.31279, $\hat{\lambda}_4$ = -0.39919	57.97155	0.01158
Punjab	Weibull	$\hat{\alpha}$ = 0.97146, $\hat{\beta}$= 42766.90	279.68594	<0.001
	PP	$\hat{C}$ = 25275.80, $\hat{\lambda}_1$ = 0.45353, $\hat{\lambda}_2$ = 0.52843	34.16930	0.68969
	SM	$\hat{b}$ = 24431.10, $\hat{a}$ = 2.13922, $\hat{q}$ = 0.87362	39.32874	0.45516
	Dagum	$\hat{b}$ = 22698.50, $\hat{a}$ = 1.90042, $\hat{p}$ = 1.24999	33.91232	0.70083
	SMD	$\hat{b}$ = 94021.40, $\hat{a}$ = 2.13786, $\hat{p}$ = 0.06378, $\hat{q}$ = 12.73190	37.37963	0.54389
	MLF	$\hat{\lambda}_1$ = 20784.10, $\hat{\lambda}_2$ = 0.0000647864, $\hat{\lambda}_3$ = 1.09591, $\hat{\lambda}_4$ = -0.48366	66.53972	0.00389
Uttarakhand	Weibull	$\hat{\alpha}$ = 0.94986, $\hat{\beta}$ = 29999.20	57.50761	<0.001
	PP	$\hat{C}$ = 21509.70, $\hat{\lambda}_1$ = 0.63446, $\hat{\lambda}_2$ = 0.48097	24.24828	0.39017
	SM	$\hat{b}$ = 23868.40, $\hat{a}$ = 1.65451, $\hat{q}$ = 1.31005	20.31885	0.62261
	Dagum	$\hat{b}$ = 24960.70, $\hat{a}$ = 2.05764, $\hat{p}$ = 0.70418	22.06330	0.51643
	SMD	$\hat{b}$ = 45790.80, $\hat{a}$ = 2.12072, $\hat{p}$ = 0.11800, $\hat{q}$ = 7.77030	17.72232	0.77237
	MLF	$\hat{\lambda}_1$ = 13475.20, $\hat{\lambda}_2$ =0.0000704259, $\hat{\lambda}_3$ = 1.45557, $\hat{\lambda}_4$ = -0.38293	32.60033	0.08829
Haryana	Weibull	$\hat{\alpha}$ = 0.93458, $\hat{\beta}$ = 33871.20	205.66002	<0.001
	PP	$\hat{C}$ = 21428.50, $\hat{\lambda}_1$ = 0.53382, $\hat{\lambda}_2$ = 0.52051	29.93999	0.85107
	SM	$\hat{b}$ = 21579.10, $\hat{a}$ = 1.88235, $\hat{q}$ = 1.02327	28.80009	0.88437
	Dagum	$\hat{b}$ = 21779.49, $\hat{a}$ = 1.91972, $\hat{p}$ = 0.96599	29.22564	0.87251
	SMD	$\hat{b}$ = 15255.40, $\hat{a}$ = 2.39289, $\hat{p}$ = 0.18244, $\hat{q}$ = 0.77850	33.34190	0.72509
	MLF	$\hat{\lambda}_1$ = 17569.50, $\hat{\lambda}_2$ = 0.0000845374, $\hat{\lambda}_3$ = 0.81413, $\hat{\lambda}_4$ = -0.50493	53.73273	0.05840

Table 5.2: (continued)

State	Distribution	Parameter estimates	Chi-square value	p-value
Rajasthan	Weibull	$\hat{\alpha} = 0.94538$, $\hat{\beta} = 26119.60$	367.23084	< 0.001
	PP	$\hat{C} = 15875.80$, $\hat{\lambda}_1 = 0.49254$, $\hat{\lambda}_2 = 0.52819$	28.71642	0.98366
	SM	$\hat{b} = 15604.30$, $\hat{a} = 2.00273$, $\hat{q} = 0.93940$	30.97418	0.96556
	Dagum	$\hat{b} = 15139.70$, $\hat{a} = 1.89705$, $\hat{p} = 1.10407$	29.18962	0.98069
	SMD	$\hat{b} = 15133.90$, $\hat{a} = 2.15392$, $\hat{p} = 0.07167$, $\hat{q} = 0.94985$	29.82734	0.97604
	MLF	$\hat{\lambda}_1 = 13462$, $\hat{\lambda}_2 = 0.000112926$, $\hat{\lambda}_3 = 0.83106$, $\hat{\lambda}_4 = -0.51259$	75.71728	0.00499
Uttar Pradesh	Weibull	$\hat{\alpha} = 0.91982$, $\hat{\beta} = 18113.50$	453.85872	< 0.001
	PP	$\hat{C} = 11605.20$, $\hat{\lambda}_1 = 0.55425$, $\hat{\lambda}_2 = 0.52164$	40.55744	0.89480
	SM	$\hat{b} = 11811.20$, $\hat{a} = 1.82522$, $\hat{q} = 1.05708$	41.31629	0.87796
	Dagum	$\hat{b} = 12062.30$, $\hat{a} = 1.91344$, $\hat{p} = 0.92105$	42.00321	0.86136
	SMD	$\hat{b} = 11432$, $\hat{a} = 1.98705$, $\hat{p} = 0.08836$, $\hat{q} = 1.09699$	49.90144	0.59557
	MLF	$\hat{\lambda}_1 = 9121.01$, $\hat{\lambda}_2 = 0.000149698$, $\hat{\lambda}_3 = 0.91310$, $\hat{\lambda}_4 = -0.49430$	95.44015	< 0.001
Bihar	Weibull	$\hat{\alpha}= 0.91046$, $\hat{\beta}= 16677.98$	227.93006	< 0.001
	PP	$\hat{C} = 9549.85$, $\hat{\lambda}_1 = 0.46198$, $\hat{\lambda}_2 = 0.55457$	23.31414	0.96117
	SM	$\hat{b} = 9178.86$, $\hat{a} = 2.08512$, $\hat{q} = 0.85314$	23.93380	0.95223
	Dagum	$\hat{b} = 8314.29$, $\hat{a} = 1.81152$, $\hat{p} = 1.31274$	24.41765	0.94429
	SMD	$\hat{b} = 4362.22$, $\hat{a} = 1.90102$, $\hat{p} = 1.38754$, $\hat{q} = 0.90252$	24.81848	0.93706
	MLF	$\hat{\lambda}_1 = 8797.96$, $\hat{\lambda}_2 = 0.000198268$, $\hat{\lambda}_3 = 0.66186$, $\hat{\lambda}_4 = -0.56804$	34.86590	0.56948
Sikkim	Weibull	$\hat{\alpha}= 1.17664$, $\hat{\beta} = 54778.50$	14.23904	0.28571
	PP	$\hat{C} = 34861.80$, $\hat{\lambda}_1 = 0.40916$, $\hat{\lambda}_2 = 0.45899$	5.03183	0.95691
	SM	$\hat{b} = 33983.90$, $\hat{a} = 2.39365$, $\hat{q} = 0.89976$	5.03183	0.95691
	Dagum	$\hat{b} = 32614.80$, $\hat{a} = 2.18721$, $\hat{p} = 1.17556$	4.81259	0.96395
	SMD	$\hat{b} = 93591.50$, $\hat{a} = 2.44477$, $\hat{p} = 0.06288$, $\hat{q} = 9.21887$	4.26661	0.97813
	MLF	$\hat{\lambda}_1 = 29943.90$, $\hat{\lambda}_2 = 0.0000529872$, $\hat{\lambda}_3 = 0.94261$, $\hat{\lambda}_4 = -0.40217$	6.43170	0.89278
Arunachal Pradesh	Weibull	$\hat{\alpha}= 0.91278$, $\hat{\beta} = 81812.90$	18.71183	0.22709
	PP	$\hat{C} = 89537.80$, $\hat{\lambda}_1 = 1.06191$, $\hat{\lambda}_2 = 0.37020$	22.52715	0.09471
	SM	$\hat{b} = 290884$, $\hat{a} = 1.07673$, $\hat{q} = 4.59039$	18.75544	0.22503
	Dagum	$\hat{b} = 126736$, $\hat{a} = 2.53118$, $\hat{p} = 0.32011$	23.46874	0.07468
	SMD	$\hat{b} = 36953.30$, $\hat{a} = 4.41210$, $\hat{p} = 2.78123$, $\hat{q} = 1.35143$	34.05024	0.00335
	MLF	$\hat{\lambda}_1 = 21124.52$, $\hat{\lambda}_2 = 0.0000609631$, $\hat{\lambda}_3 = 0.96982$, $\hat{\lambda}_4 = -0.51160$	36.74661	0.00138

Table 5.2: (continued)

State	Distribution	Parameter estimates	Chi-square value	p-value
Nagaland	Weibull	$\hat{\alpha} = 0.97621$, $\hat{\beta} = 56908$	15.65998	0.26798
	PP	$\hat{C} = 51382.70$, $\hat{\lambda}_1 = 0\ .82500$, $\hat{\lambda}_2 = 0.40518$	8.96721	0.77542
	SM	$\hat{b}$= 81760.90, $\hat{a} = 1.34359$, $\hat{q} = 2.28942$	10.82055	0.62585
	Dagum	$\hat{b} = 67847.8$, $\hat{a} = 2.37790$, $\hat{p} = 0.44415$	8.05677	0.83988
	SMD	$\hat{b} = 133086.82$, $\hat{a} = 1.26178$, $\hat{p} = 0.00355$, $\hat{q} = 3.66715$	9.79159	0.71088
	MLF	$\hat{\lambda}_1 = 32184$, $\hat{\lambda}_2 = 0.0000401895$, $\hat{\lambda}_3 = 0.76766$, $\hat{\lambda}_4 = -0.36848$	9.49195	0.73486
Manipur	Weibull	$\hat{\alpha} = 1.73906$, $\hat{\beta} = 48715.80$	1.76002	0.99918
	PP	$\hat{C} = 50621.90$, $\hat{\lambda}_1 = 0.57253$, $\hat{\lambda}_2 = 0.20556$	2.79271	0.99319
	SM	$\hat{b} = 118753.45$, $\hat{a} = 1.90743$, $\hat{q} = 6.18523$	1.66873	0.99936
	Dagum	$\hat{b} = 59589.10$, $\hat{a} = 4.41998$, $\hat{p} = 0.35268$	2.08779	0.99816
	SMD	$\hat{b} = 52123.70$, $\hat{a} = 4.39599$, $\hat{p} = 0.19806$, $\hat{q} = 141.95667$	3.06383	0.98986
	MLF	$\hat{\lambda}_1 = 36620.60$, $\hat{\lambda}_2 = 0.0000530083$, $\hat{\lambda}_3 = 0.55038$, $\hat{\lambda}_4 = -0.00451$	1.08326	0.99992
Mizoram	Weibull	$\hat{\alpha} = 1.24314$, $\hat{\beta} = 66450.40$	3.97321	0.97069
	PP	$\hat{C} = 56110.40$, $\hat{\lambda}_1 = 0.60385$, $\hat{\lambda}_2 = 0.35663$	4.24248	0.96228
	SM	$\hat{b} = 71962.40$, $\hat{a} = 1.78052$, $\hat{q} = 1.81579$	4.27742	0.96109
	Dagum	$\hat{b} = 67327.60$, $\hat{a} = 2.72196$, $\hat{p} = 0.54383$	4.64880	0.94698
	SMD	$\hat{b} = 75386.30$, $\hat{a} = 2.73965$, $\hat{p} = 0.14837$, $\hat{q} = 7.83161$	4.93871	0.93410
	MLF	$\hat{\lambda}_1 = 38393.40$, $\hat{\lambda}_2 = 0.0000328686$, $\hat{\lambda}_3 = 1.03166$, $\hat{\lambda}_4 = -0.21081$	8.55044	0.66331
Tripura	Weibull	$\hat{\alpha} = 1.25147$, $\hat{\beta} = 26308.90$	22.75483	0.15740
	PP	$\hat{C} = 18407.90$, $\hat{\lambda}_1 = 0.45000$, $\hat{\lambda}_2 = 0.41531$	7.88528	0.96894
	SM	$\hat{b} = 18816.60$, $\hat{a} = 2.25189$, $\hat{q} = 1.08115$	8.36108	0.95821
	Dagum	$\hat{b} = 19122.90$, $\hat{a} = 2.39986$, $\hat{p} = 0.90071$	7.88528	0.96894
	SMD	$\hat{b} = 21896.10$, $\hat{a} = 2.39870$, $\hat{p} = 0.07268$, $\hat{q} = 1.61425$	7.34345	0.97865
	MLF	$\hat{\lambda}_1 = 15545.70$, $\hat{\lambda}_2 = 0.000111951$, $\hat{\lambda}_3 = 0.75768$, $\hat{\lambda}_4 = -0.36349$	9.64003	0.91802
Meghalaya	Weibull	$\hat{\alpha} = 0.98753$, $\hat{\beta} = 38455.90$	11.88801	0.61530
	PP	$\hat{C} = 35000.50$, $\hat{\lambda}_1 = 0.82389$, $\hat{\lambda}_2 = 0.39908$	12.29989	0.58223
	SM	$\hat{b} = 56628.60$, $\hat{a} = 1.34618$, $\hat{q} = 2.34783$	12.71406	0.54915
	Dagum	$\hat{b} = 46182.20$, $\hat{a} = 2.41064$, $\hat{p} = 0.43866$	10.81431	0.70056
	SMD	$\hat{b} = 65896$, $\hat{a} = 2.67999$, $\hat{p} = 0.14630$, $\hat{q} = 52.24140$	9.33732	0.80885
	MLF	$\hat{\lambda}_1 = 15280.80$, $\hat{\lambda}_2 = 0.0000404445$, $\hat{\lambda}_3 = 2.09340$, $\hat{\lambda}_4 = -0.21197$	16.19671	0.30151

Table 5.2: (continued)

State	Distribution	Parameter estimates	Chi-square value	p-value
Assam	Weibull	$\hat{\alpha} = 0.99976$, $\hat{\beta} = 31763.70$	99.63165	< 0.001
	PP	$\hat{C} = 23320.80$, $\hat{\lambda}_1 = 0.62285$, $\hat{\lambda}_2 = 0.45917$	29.59604	0.53822
	SM	$\hat{b} = 26220.30$, $\hat{a} = 1.69045$, $\hat{q} = 1.35414$	28.60266	0.58992
	Dagum	$\hat{b} = 27182.40$, $\hat{a} = 2.15128$, $\hat{p} = 0.68317$	29.66946	0.53441
	SMD	$\hat{b} = 47771.40$, $\hat{a} = 2.25564$, $\hat{p} = 0.11745$, $\hat{q} = 8.84803$	18.25926	0.96616
	MLF	$\hat{\lambda}_1 = 14638.80$, $\hat{\lambda}_2 = 0.0000654578$, $\hat{\lambda}_3 = 1.46412$, $\hat{\lambda}_4 = -0.34512$	55.54066	0.00436
West Bengal	Weibull	$\hat{\alpha} = 0.90554$, $\hat{\beta} = 27292$	372.15192	< 0.001
	PP	$\hat{C} = 17435.30$, $\hat{\lambda}_1 = 0.56092$, $\hat{\lambda}_2 = 0.52698$	73.76096	0.00438
	SM	$\hat{b} = 17755.30$, $\hat{a} = 1.80429$, $\hat{q} = 1.05862$	72.82032	0.00540
	Dagum	$\hat{b} = 18151.20$, $\hat{a} = 1.89399$, $\hat{p} = 0.91877$	74.39073	0.00380
	SMD	$\hat{b} = 58652.70$, $\hat{a} = 2.16688$, $\hat{p} = 0.08357$, $\hat{q} = 15.76601$	32.16558	0.92452
	MLF	$\hat{\lambda}_1 = 10656$, $\hat{\lambda}_2 = 0.0000779768$, $\hat{\lambda}_3 = 1.92962$, $\hat{\lambda}_4 = -0.42600$	104.37460	< 0.001
Jharkhand	Weibull	$\hat{\alpha} = 0.86648$, $\hat{\beta} = 20126.30$	117.10535	< 0.001
	PP	$\hat{C} = 13483.70$, $\hat{\lambda}_1 = 0.63046$, $\hat{\lambda}_2 = 0.52752$	35.84448	0.17814
	SM	$\hat{b} = 14320.60$, $\hat{a} = 1.64059$, $\hat{q} = 1.18034$	35.70311	0.18238
	Dagum	$\hat{b} = 15064.90$, $\hat{a} = 1.88497$, $\hat{p} = 0.79095$	36.84274	0.15029
	SMD	$\hat{b} = 35072.10$, $\hat{a} = 1.95034$, $\hat{p} = 0.10650$, $\hat{q} = 7.97153$	41.22656	0.06579
	MLF	$\hat{\lambda}_1 = 8447.66$, $\hat{\lambda}_2 = 0.000108006$, $\hat{\lambda}_3 = 1.52862$, $\hat{\lambda}_4 = -0.45068$	71.54190	< 0.001
Orissa	Weibull	$\hat{\alpha} = 0.87145$, $\hat{\beta} = 18412.90$	410.89852	< 0.001
	PP	$\hat{C} = 10118.60$, $\hat{\lambda}_1 = 0.44968$, $\hat{\lambda}_2 = 0.57743$	32.14270	0.86414
	SM	$\hat{b} = 9621.99$, $\hat{a} = 2.10906$, $\hat{q} = 0.80790$	38.01212	0.64665
	Dagum	$\hat{b} = 8232.51$, $\hat{a} = 1.74169$, $\hat{p} = 1.47129$	30.13489	0.91418
	SMD	$\hat{b} = 52035.80$, $\hat{a} = 2.03343$, $\hat{p} = 0.05748$, $\hat{q} = 19.08230$	32.07433	0.86608
	MLF	$\hat{\lambda}_1 = 8341.41$, $\hat{\lambda}_2 = 0.000151836$, $\hat{\lambda}_3 = 1.18831$, $\hat{\lambda}_4 = -0.54264$	92.13101	< 0.001
Chhattisgarh	Weibull	$\hat{\alpha} = 0.76363$, $\hat{\beta} = 17146.90$	305.42667	< 0.001
	PP	$\hat{C} = 9354.93$, $\hat{\lambda}_1 = 0.51080$, $\hat{\lambda}_2 = 0.61822$	33.71838	0.52992
	SM	$\hat{b} = 8973.71$, $\hat{a} = 1.87713$, $\hat{q} = 0.85161$	38.27264	0.32320
	Dagum	$\hat{b} = 7914.84$, $\hat{a} = 1.62389$, $\hat{p} = 1.33793$	33.95700	0.51832
	SMD	$\hat{b} = 56979.10$, $\hat{a} = 1.93084$, $\hat{p} = 0.06201$, $\hat{q} = 24.45070$	23.79824	0.92423
	MLF	$\hat{\lambda}_1 = 6347.29$, $\hat{\lambda}_2 = 0.000141024$, $\hat{\lambda}_3 = 1.82247$, $\hat{\lambda}_4 = -0.56459$	65.66647	0.00128

Table 5.2: (continued)

State	Distribution	Parameter estimates	Chi-square value	p-value
Madhya Pradesh	Weibull	$\hat{\alpha} = 0.89310$, $\hat{\beta} = 17413.60$	447.05734	< 0.001
	PP	$\hat{C} = 10226.30$, $\hat{\lambda}_1 = 0.49448$, $\hat{\lambda}_2 = 0.55336$	28.84853	0.99035
	SM	$\hat{b} = 9957.72$, $\hat{a} = 1.97674$, $\hat{q} = 0.90596$	29.57439	0.98731
	Dagum	$\hat{b} = 9416.30$, $\hat{a} = 1.81248$, $\hat{p} = 1.17712$	29.73716	0.98653
	SMD	$\hat{b} = 33911.30$, $\hat{a} = 1.74572$, $\hat{p} = 0.06097$, $\hat{q} = 5.70088$	48.36476	0.49879
	MLF	$\hat{\lambda}_1 = 8990.78$, $\hat{\lambda}_2 = 0.000178377$, $\hat{\lambda}_3 = 0.73707$, $\hat{\lambda}_4 = -0.55637$	63.19845	0.08361
Gujarat	Weibull	$\hat{\alpha} = 0.83334$, $\hat{\beta} = 28776.70$	216.17278	< 0.001
	PP	$\hat{C} = 18881.40$, $\hat{\lambda}_1 = 0.63486$, $\hat{\lambda}_2 = 0.54521$	31.25132	0.83747
	SM	$\hat{b} = 19850.10$, $\hat{a} = 1.62248$, $\hat{q} = 1.14969$	33.93150	0.73910
	Dagum	$\hat{b} = 20858.50$, $\hat{a} = 1.82585$, $\hat{p} = 0.81650$	31.62483	0.82513
	SMD	$\hat{b} = 22909.40$, $\hat{a} = 1.65942$, $\hat{p} = 0.08174$, $\hat{q} = 1.44293$	27.75644	0.92836
	MLF	$\hat{\lambda}_1 = 13906.30$, $\hat{\lambda}_2 = 0.0000892538$, $\hat{\lambda}_3 = 0.94916$, $\hat{\lambda}_4 = -0.52421$	62.09572	0.01412
Maharashtra	Weibull	$\hat{\alpha} = 1.08139$, $\hat{\beta} = 30549.40$	239.31163	< 0.001
	PP	$\hat{C} = 24815.10$, $\hat{\lambda}_1 = 0.65990$, $\hat{\lambda}_2 = 0.40648$	108.91138	< 0.001
	SM	$\hat{b} = 31174.60$, $\hat{a} = 1.63050$, $\hat{q} = 1.68641$	100.18403	< 0.001
	Dagum	$\hat{b} = 30293.70$, $\hat{a} = 2.40239$, $\hat{p} = 0.56315$	126.64433	< 0.001
	SMD	$\hat{b} = 41194.60$, $\hat{a} = 2.48908$, $\hat{p} = 0.14149$, $\hat{q} = 10.78540$	56.32609	0.25022
	MLF	$\hat{\lambda}_1 = 15397.10$, $\hat{\lambda}_2 = 0.0000657236$, $\hat{\lambda}_3 = 1.31086$, $\hat{\lambda}_4 = -0.26972$	135.53774	< 0.001
Andhra Pradesh	Weibull	$\hat{\alpha} = 1.14639$, $\hat{\beta} = 24638.20$	160.25950	< 0.001
	PP	$\hat{C} = 17687.60$, $\hat{\lambda}_1 = 0.52024$, $\hat{\lambda}_2 = 0.42873$	43.32107	0.41477
	SM	$\hat{b} = 18827.50$, $\hat{a} = 1.98448$, $\hat{q} = 1.21072$	39.99545	0.55930
	Dagum	$\hat{b} = 19394.20$, $\hat{a} = 2.31362$, $\hat{p} = 0.78211$	40.49195	0.53725
	SMD	$\hat{b} = 27269.70$, $\hat{a} = 1.81675$, $\hat{p} = 0.02318$, $\hat{q} = 1.99551$	36.95966	0.69148
	MLF	$\hat{\lambda}_1 = 15074.40$, $\hat{\lambda}_2 = 0.000132962$, $\hat{\lambda}_3 = 0.48970$, $\hat{\lambda}_4 = -0.44057$	40.25369	0.54783
Karnataka	Weibull	$\hat{\alpha} =0.93914$, $\hat{\beta} = 28517.40$	458.13241	< 0.001
	PP	$\hat{C} = 17333.10$, $\hat{\lambda}_1 = 0.49627$, $\hat{\lambda}_2 = 0.53003$	39.81258	0.90981
	SM	$\hat{b} = 17052.20$, $\hat{a} = 1.98922$, $\hat{q} = 0.94285$	39.96478	0.90686
	Dagum	$\hat{b} = 16573.10$, $\hat{a} = 1.89023$, $\hat{p} = 1.09763$	37.95640	0.94081
	SMD	$\hat{b} = 11142.30$, $\hat{a} = 2.56581$, $\hat{p} = 0.24821$, $\hat{q} = 0.67946$	59.45113	0.25238
	MLF	$\hat{\lambda}_1 = 15257$, $\hat{\lambda}_2 = 0.000110607$, $\hat{\lambda}_3 = 0.67753$, $\hat{\lambda}_4 = -0.53591$	56.50516	0.34543

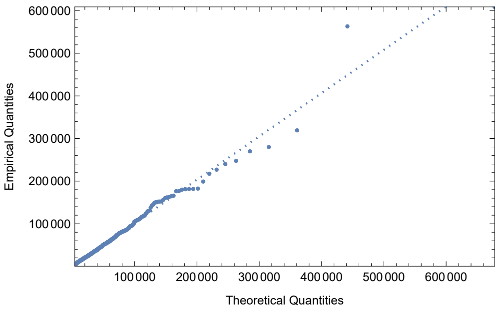
Table 5.2: (continued)

State	Distribution	Parameter estimates	Chi-square value	p-value
Goa	Weibull	$\hat{\alpha}$ = 1.39754, $\hat{\beta}$ = 39755.70	14.24673	0.58034
	PP	$\hat{C}$ = 37800.10, $\hat{\lambda}_1$ = 0.62785, $\hat{\lambda}_2$ = 0.28735	7.12698	0.97074
	SM	$\hat{b}$ = 61829.40, $\hat{a}$ = 1.73987, $\hat{q}$ = 2.84578	7.31651	0.96667
	Dagum	$\hat{b}$ = 45984.20, $\hat{a}$ = 3.28954, $\hat{p}$ = 0.42854	8.00638	0.94868
	SMD	$\hat{b}$ = 301035, $\hat{a}$ = 1.56546, $\hat{p}$ = 0.00146, $\hat{q}$ = 25.68720	8.63095	0.92783
	MLF	$\hat{\lambda}_1$ = 27866.70, $\hat{\lambda}_2$ = 0.000074885, $\hat{\lambda}_3$ = 0.40668, $\hat{\lambda}_4$ = -0.25276	7.14315	0.97040
Kerala	Weibull	$\hat{\alpha}$ = 1.24923, $\hat{\beta}$ = 48439.40	78.48560	<0.001
	PP	$\hat{C}$ = 37795.70, $\hat{\lambda}_1$ = 0.53720, $\hat{\lambda}_2$ = 0.38101	26.99429	0.88666
	SM	$\hat{b}$ = 43001.40, $\hat{a}$ = 1.95934, $\hat{q}$ = 1.43648	28.53141	0.83949
	Dagum	$\hat{b}$ = 43266.60, $\hat{a}$ = 2.57994, $\hat{p}$ = 0.66028	35.43710	0.54238
	SMD	$\hat{b}$ = 61373.90, $\hat{a}$ = 1.80666, $\hat{p}$ = 0.01238, $\hat{q}$ = 2.31413	24.50421	0.94278
	MLF	$\hat{\lambda}_1$ = 31167.10, $\hat{\lambda}_2$ = 0.0000668706, $\hat{\lambda}_3$ = 0.45094, $\hat{\lambda}_4$ = -0.38362	36.58450	0.48836
Tamil Nadu	Weibull	$\hat{\alpha}$ = 1.14495, $\hat{\beta}$ = 36739.70	172.35762	<0.001
	PP	$\hat{C}$ = 25008.10, $\hat{\lambda}_1$ = 0.47776, $\hat{\lambda}_2$ = 0.44528	27.12488	0.95294
	SM	$\hat{b}$ = 25503.70, $\hat{a}$ = 2.11868, $\hat{q}$ = 1.06960	25.49330	0.97240
	Dagum	$\hat{b}$ = 25940.30, $\hat{a}$ = 2.23977, $\hat{p}$ = 0.91125	25.35602	0.97371
	SMD	$\hat{b}$ = 45453.20, $\hat{a}$ = 1.88885, $\hat{p}$ = 0.03791, $\hat{q}$ = 2.51854	39.15850	0.55271
	MLF	$\hat{\lambda}_1$ = 22186.60, $\hat{\lambda}_2$ = 0.0000925338, $\hat{\lambda}_3$ = 0.49879, $\hat{\lambda}_4$ = -0.46058	38.33929	0.58953

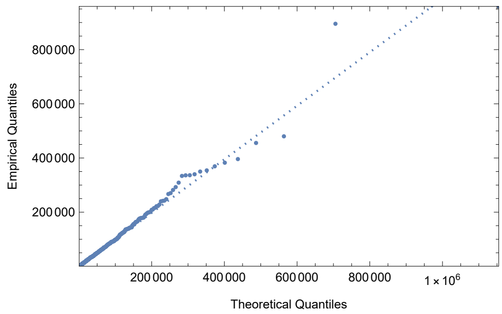
5.4.2 Fitting of distributions

The six distributions such as Weibull, PP, SM, SMD, Dagum, and MLF are fitted to the income data of each state. This study utilizes the L-moments method for parameter estimation. The adequacy of the model is assessed using the Q-Q plot and the chi-square goodness of fit approach. Models are accepted if their p-value exceeds 0.05 and their Q-Q plot demonstrates a satisfactory fit. Table 5.2 presents the parameter estimates, chi-square statistics, and p-values of the six models.

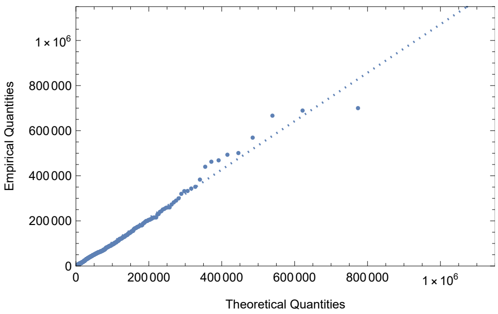
Figure 5.1 displays the Q-Q plot corresponding to the best model for each state.



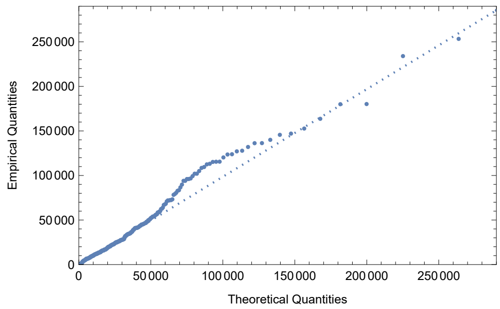
(1) Q-Q plot for Jammu & Kashmir dataset w.r.t. MLF.



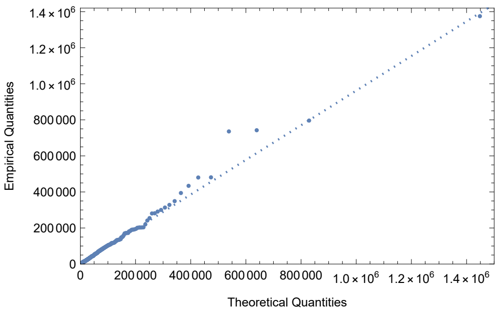
(2) Q-Q plot for Himachal Pradesh dataset w.r.t. SMD.



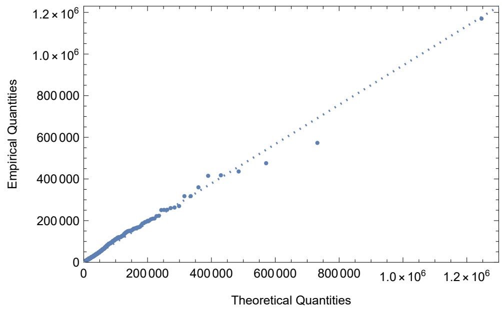
(3) Q-Q plot for Punjab dataset w.r.t. SMD.



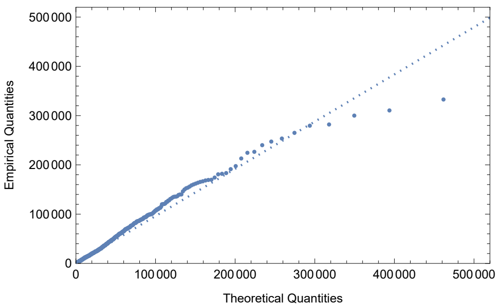
(4) Q-Q plot for Uttarakhand dataset w.r.t. SM.



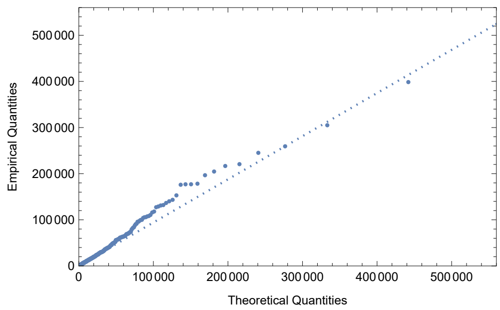
(5) Q-Q plot for Haryana dataset w.r.t. MLF.



(6) Q-Q plot for Rajasthan dataset w.r.t. SMD.

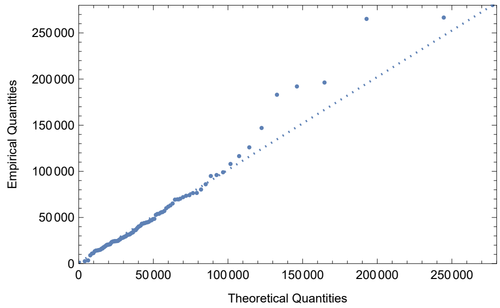


(7) Q-Q plot for Uttar Pradesh dataset w.r.t. SMD.

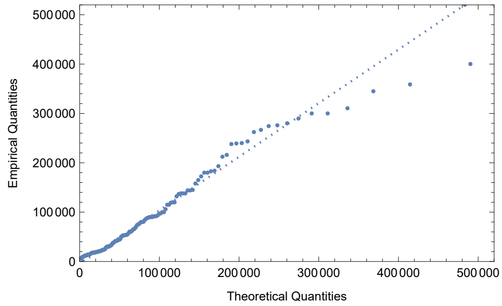


(8) Q-Q plot for Bihar dataset w.r.t. Dagum.

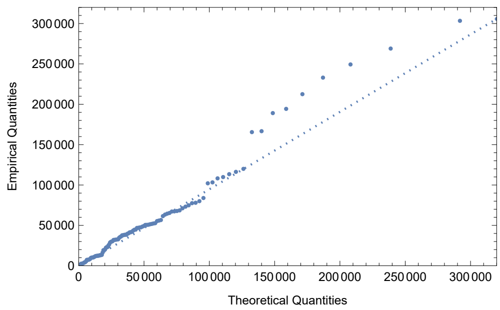
Figure 5.1: Q-Q plot of per-capita household income data of Indian states



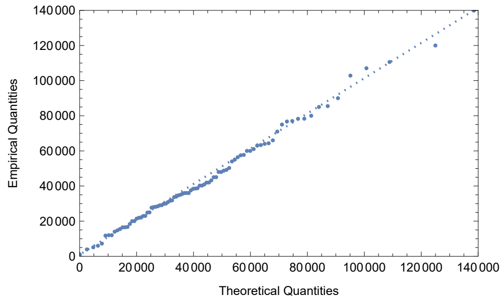
(9) Q-Q plot for Sikkim dataset w.r.t. Dagum.



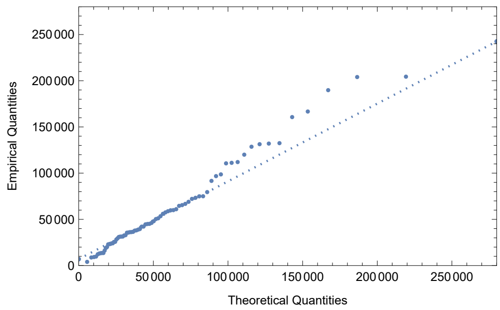
(10) Q-Q plot for Arunachal Pradesh dataset w.r.t. SM.



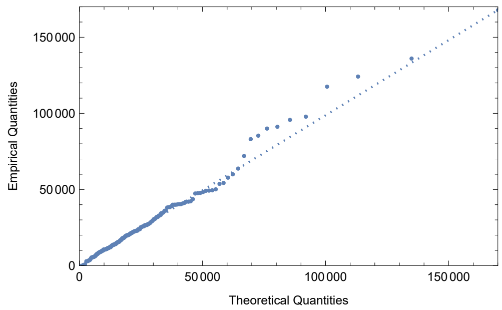
(11) Q-Q plot for Nagaland dataset w.r.t. SM.



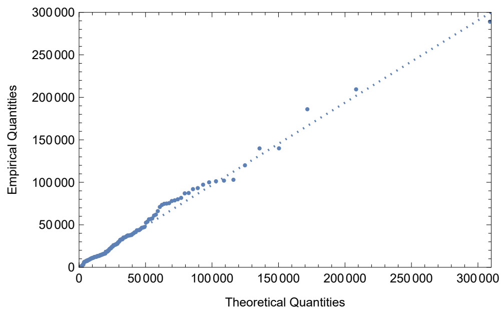
(12) Q-Q plot for Manipur dataset w.r.t. Weibull.



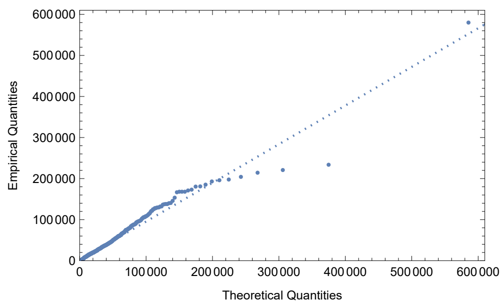
(13) Q-Q plot for Mizoram dataset w.r.t. SMD.



(14) Q-Q plot for Tripura dataset w.r.t. SMD.

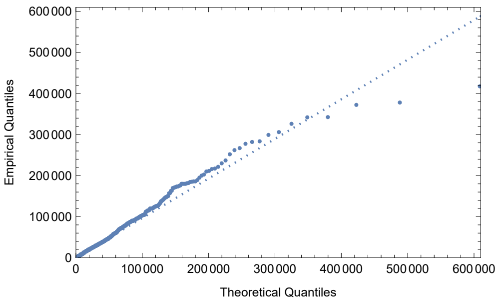


(15) Q-Q plot for Meghalaya dataset w.r.t. SM.

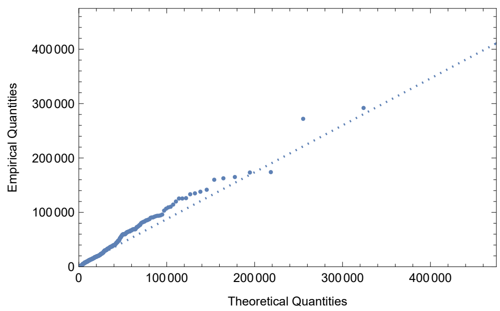


(16) Q-Q plot for Assam dataset w.r.t. SMD.

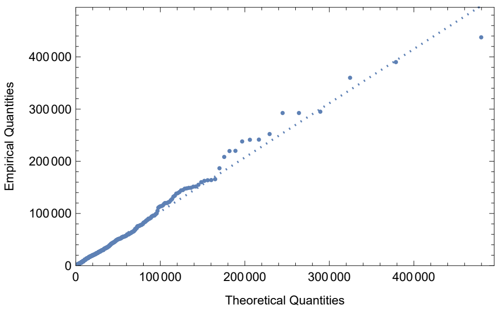
Figure 5.1: (continued)



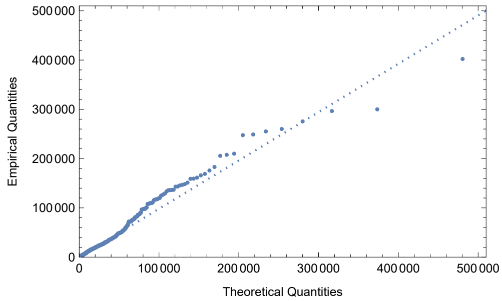
(17) Q-Q plot for West Bengal dataset w.r.t. SMD.



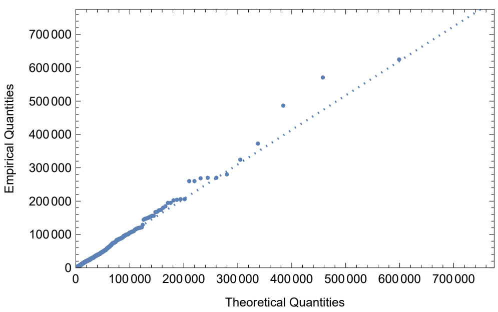
(18) Q-Q plot for Jharkhand dataset w.r.t. SMD.



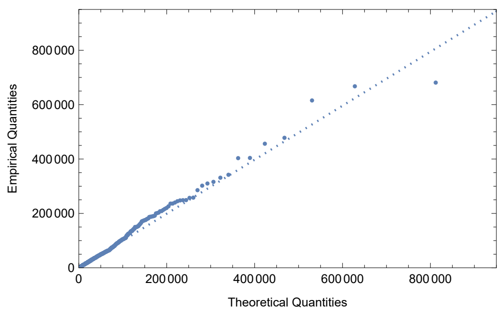
(19) Q-Q plot for Orissa dataset w.r.t. SMD.



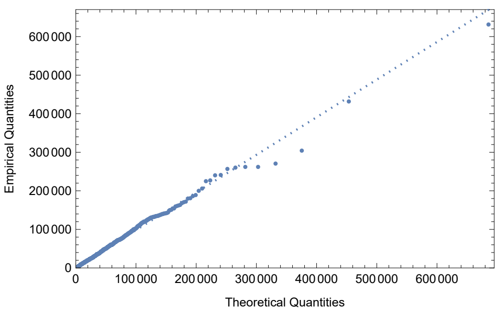
(20) Q-Q plot for Chhattisgarh dataset w.r.t. SMD.



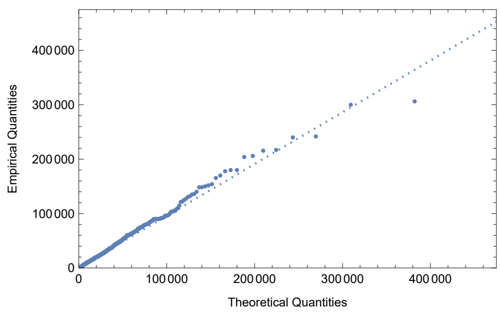
(21) Q-Q plot for Madhya Pradesh dataset w.r.t. SMD.



(22) Q-Q plot for Gujarat dataset w.r.t. SMD.

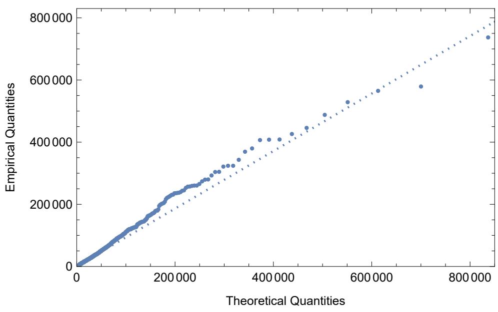


(23) Q-Q plot for Maharashtra dataset w.r.t. SMD.

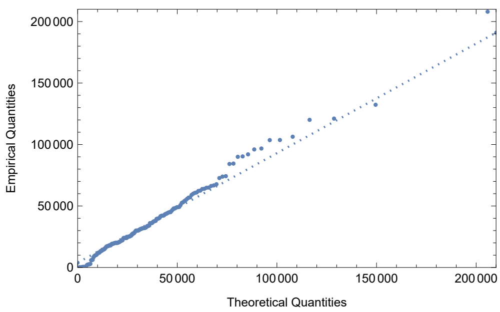


(24) Q-Q plot for Andhra Pradesh dataset w.r.t. SMD.

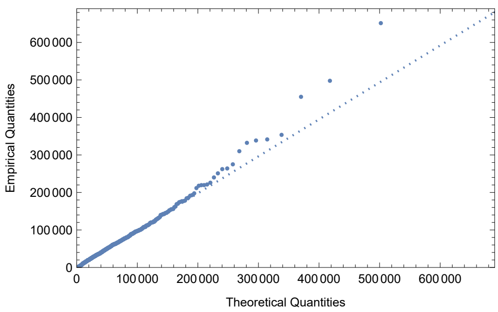
Figure 5.1: (continued)



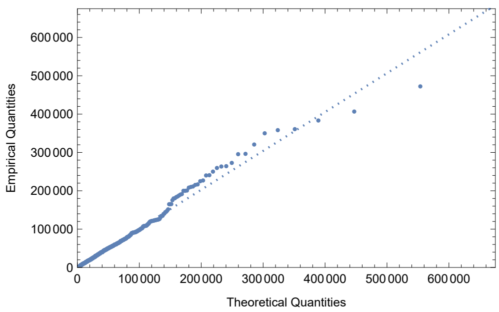
(25) Q-Q plot for Karnataka dataset w.r.t. Dagum.



(26) Q-Q plot for Goa dataset w.r.t. PP.



(27) Q-Q plot for Kerala dataset w.r.t. SM.



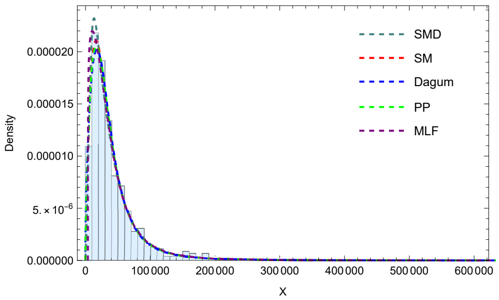
(28) Q-Q plot for Tamil Nadu dataset w.r.t. SMD.

Figure 5.1: (continued)

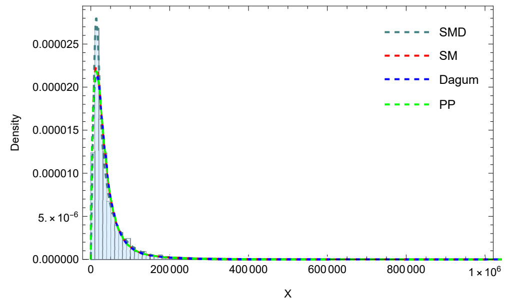
To identify the best model in each state, the data histogram is plotted along with the density functions of the accepted models, as illustrated in Figure 5.2.

In many states, the histogram does not explicitly reveal the most suitable model for modeling the income data, as several models exhibit similar levels of fit. Hence, we base our selection on the goodness of fit assessment using correlation, choosing the model with the highest correlation as the most suitable among the available options (Nair et al., 2024).

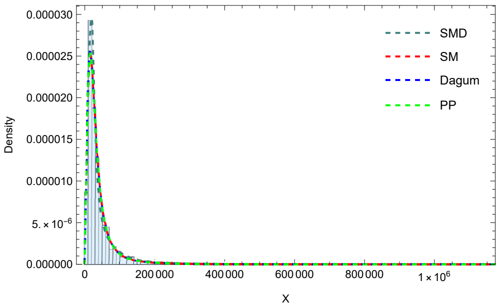
The correlation method used in this study is outlined as follows. For a given random sample and an assumed model, the pairs used to compute the correlation coefficient are expressed as $(x_{i:n}, Q\left(\frac{2i-1}{2n}\right))$. Here, $x_{i:n}$ represents the i^{th} ordered



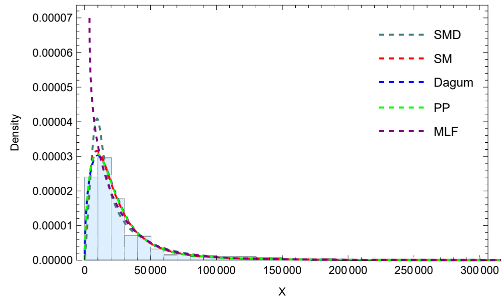
(1) Jammu & Kashmir data.



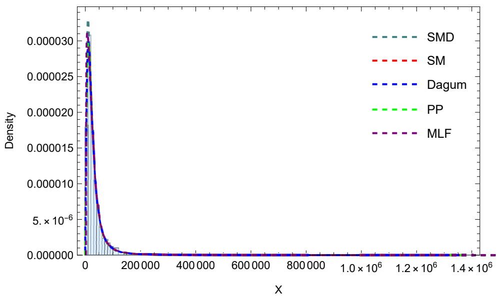
(2) Himachal Pradesh data.



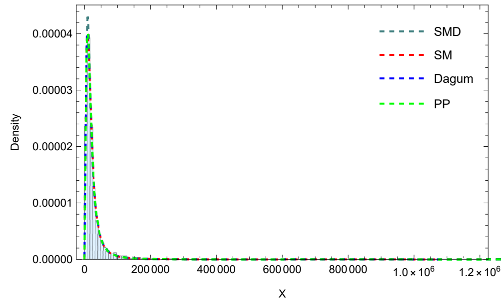
(3) Punjab data.



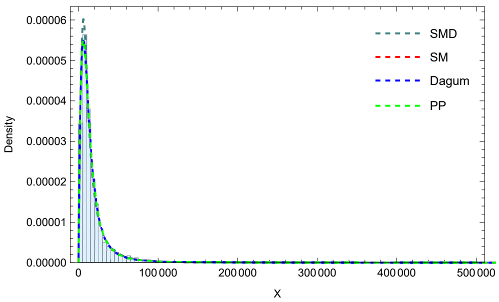
(4) Uttarakhand data.



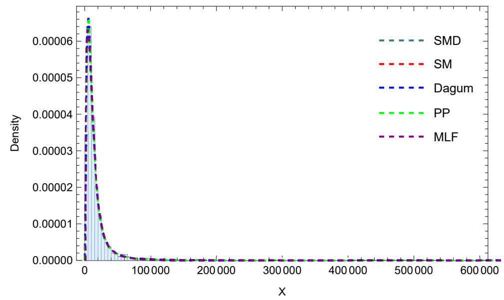
(5) Haryana data.



(6) Rajasthan data.

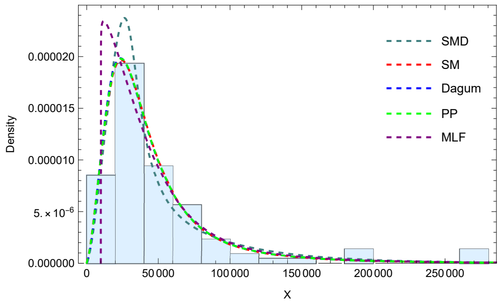


(7) Uttar Pradesh data.

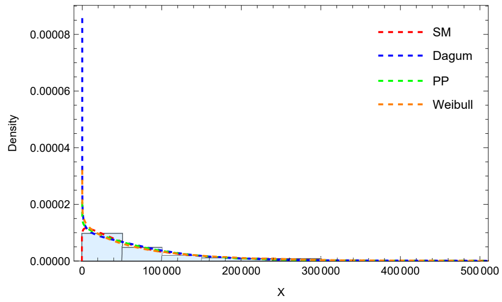


(8) Bihar data.

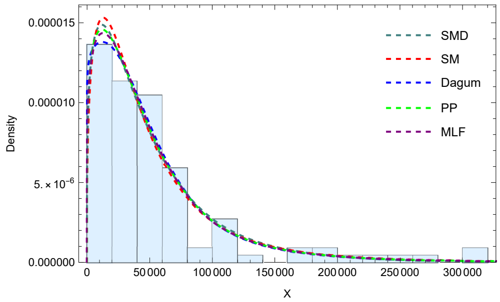
Figure 5.2: The densities of the accepted models for the per-capita household income data of Indian states



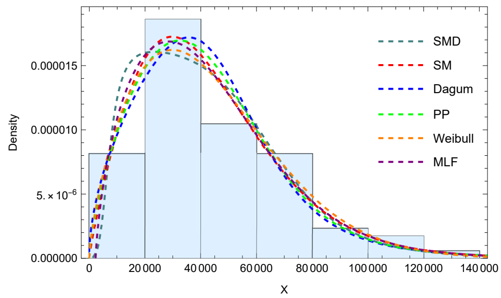
(9) Sikkim data.



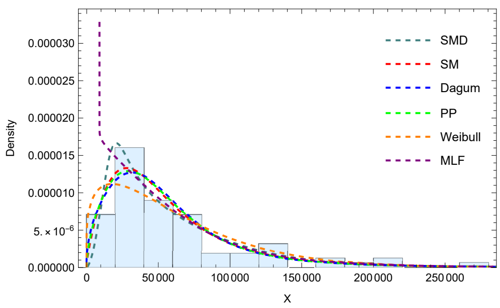
(10) Arunachal Pradesh data.



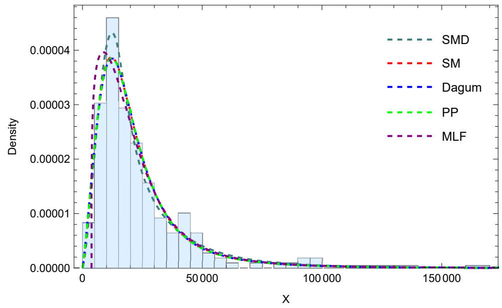
(11) Nagaland data.



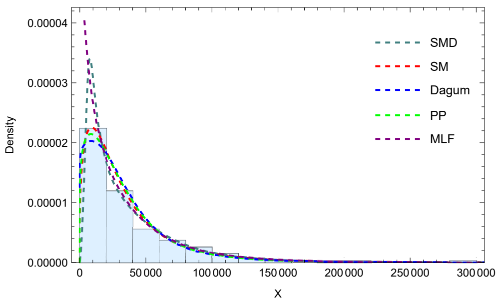
(12) Manipur data.



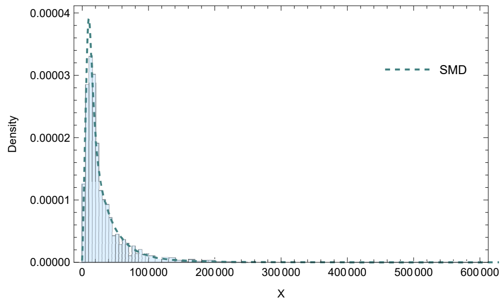
(13) Mizoram data.



(14) Tripura data.

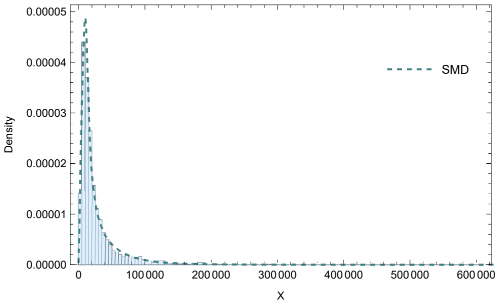


(15) Meghalaya data.

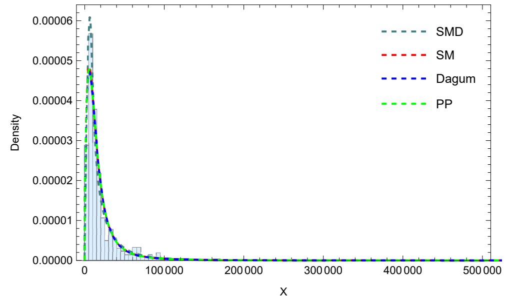


(16) Assam data.

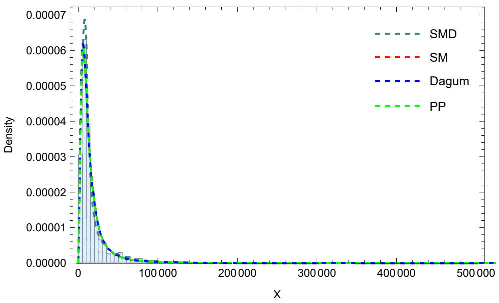
Figure 5.2: (continued).



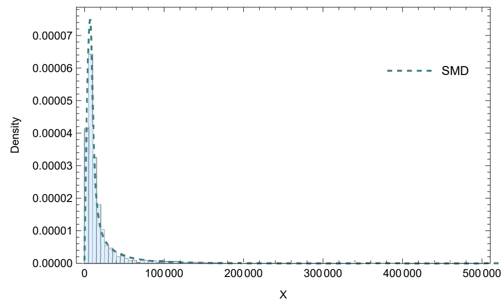
(17) West Bengal data.



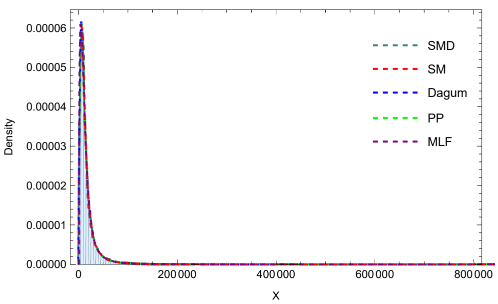
(18) Jharkhand data.



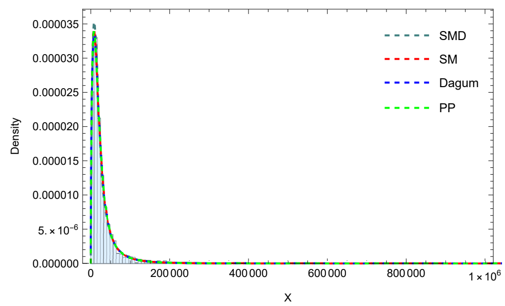
(19) Orissa data.



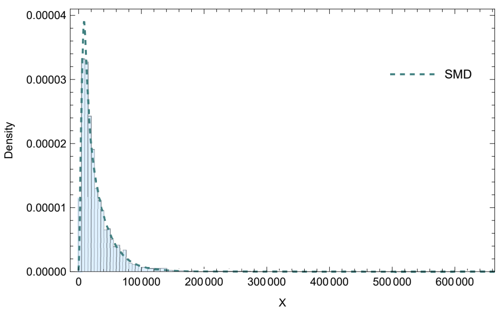
(20) Chhattisgarh data.



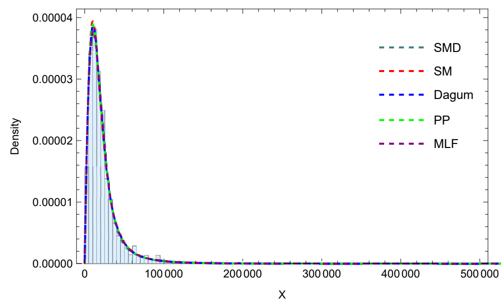
(21) Madhya Pradesh data.



(22) Gujarat data.

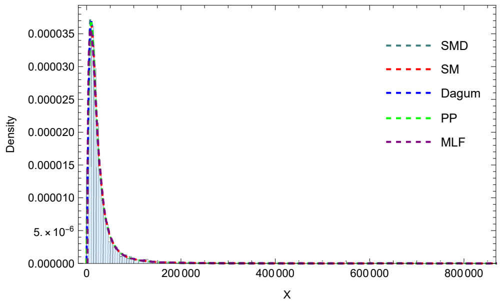


(23) Maharashtra data.

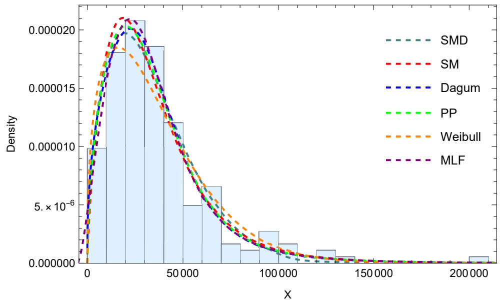


(24) Andhra Pradesh data.

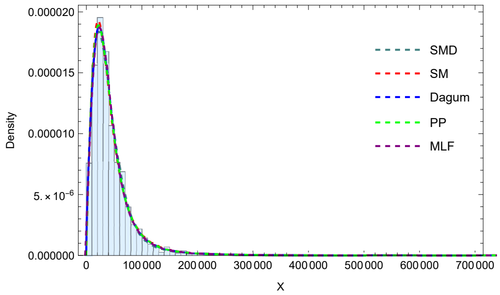
Figure 5.2: (continued)



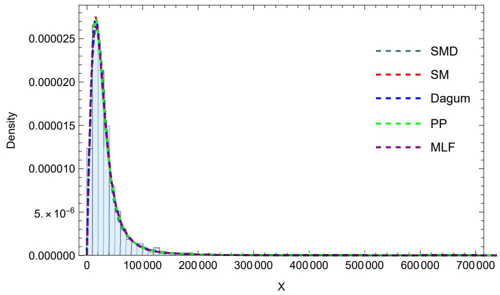
(25) Karnataka data.



(26) Goa data.



(27) Kerala data.



(28) Tamil Nadu data.

Figure 5.2: (continued)

observation in the sample, and $Q(.)$ denotes the quantile function of the assumed model. Let, $y_i = Q\left(\frac{2i-1}{2n}\right)$, then the correlation coefficient r, between $x_{i:n}$ and y_i is

$$r = \frac{\sum_{i=1}^{n}(x_{i:n} - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_{i:n} - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{5.4.1}$$

where $\bar{x}$ is the mean of the sample and $\bar{y}$ is the mean of the y_i values. The correlation coefficient between x_i and y_i is expected to be close to one when a sample is generated from the assumed model.

5.4.3 Results

The results of the primary goodness of fit criteria, including the Q-Q plot and chi-square test, are compiled in Table 5.3. Additionally, it illustrates the distributions that give comparable fits in the histogram and pinpoints the best model for each state based on the correlation coefficient. The table clearly shows that out of the six distributions, the SMD distribution offers the best fit for the per-capita household income data of sixteen states. In contrast, the Weibull and PP distributions each offer the best fit in only one state, while the MLF, Dagum, and SM distributions provide the best fit in two, three, and five states, respectively.

For each state, we calculate empirical income inequality measures, including the Gini, Pietra, Bonferroni, and Atkinson indices, as well as generalized entropy and Frigyes measures. We also compute the corresponding theoretical measures derived from the best-fitting model for that state. Table 5.4 displays the empirical and theoretical income inequality measures for various Indian states, derived from per-capita household income data in IHDS-II. The empirical and theoretical measures of the Atkinson index and generalized entropy measures have been estimated for $\epsilon = 0.5$ and $\theta = 0.5$, respectively. As mentioned, numerical integration is employed to calculate the theoretical inequality measures that lack closed-form expressions.

From Table 5.4, we can infer that the empirical and theoretical (estimated) income inequality measures such as Gini, Pietra, Atkinson, Bonferroni, generalized entropy, and Frigyes measures have their lowest values in Manipur and highest values in

Table 5.3: Results of data analysis on per-capita household income data across Indian states

State	Distributions that satisfy goodness of fit criteria - chi-square test and Q-Q plot	Distributions that gives good fit from histogram	Correlation coefficient	Distribution that gives best fit
Jammu & Kashmir	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.99234	MLF
Himachal Pradesh	PP, SM, SMD, Dagum	SMD	0.99093	SMD
Punjab	PP, SM, SMD, Dagum	SMD	0.99297	SMD
Uttarakhand	PP, SM, SMD, MLF, Dagum	PP, SM, Dagum	0.94221	SM
Haryana	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.99490	MLF
Rajasthan	PP, SM, SMD, Dagum	PP, SM, SMD, Dagum	0.99523	SMD
Uttar Pradesh	PP, SM, SMD, Dagum	PP, SM, SMD, Dagum	0.96279	SMD
Bihar	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.97466	Dagum
Sikkim	PP, SM, SMD, MLF, Dagum	PP, SM, Dagum	0.94690	Dagum
Arunachal Pradesh	PP, SM, Dagum, Weibull	SM	0.97579	SM
Nagaland	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.96664	SM
Manipur	PP, SM, SMD, MLF, Dagum, Weibull	PP, SM, MLF, Dagum, Weibull	0.99701	Weibull
Mizoram	PP, SM, SMD, MLF, Dagum, Weibull	PP, SM, SMD, Dagum	0.99144	SMD
Tripura	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.98867	SMD
Meghalaya	PP, SM, SMD, MLF, Dagum	PP, SM, Dagum	0.99559	SM
Assam	SMD	SMD	0.98683	SMD
West Bengal	SMD	SMD	0.97709	SMD
Jharkhand	PP, SM, SMD, Dagum	SMD	0.99168	SMD
Orissa	PP, SM, SMD, Dagum	PP, SM, SMD, Dagum	0.97416	SMD
Chhattisgarh	SMD	SMD	0.96188	SMD
Madhya Pradesh	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.98057	SMD
Gujarat	PP, SM, SMD, Dagum	PP, SM, SMD, Dagum	0.97577	SMD
Maharashtra	SMD	SMD	0.99752	SMD
Andhra Pradesh	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.99212	SMD
Karnataka	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.93827	Dagum
Goa	PP, SM, SMD, MLF, Dagum, Weibull	PP, SM, SMD, MLF, Dagum	0.99630	PP
Kerala	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.99296	SM
Tamil Nadu	PP, SM, SMD, MLF, Dagum	PP, SM, SMD, MLF, Dagum	0.98826	SMD

Chhattisgarh. However, an exception to this trend is the empirical Frigyes measure ρ, which attains its highest value in Arunachal Pradesh. Also, Table 5.5 summarizes the ranking of Indian states with respect to estimated income inequality measures.

5.5 Summary

In this chapter, the per-capita household income data of Indian states, obtained from the IHDS-II survey, has been effectively modeled using the Weibull, PP, SM, Dagum, SMD, and MLF distributions. The parameters of these distributions have been estimated using the L-moments method, and the model adequacy has been initially evaluated through Q-Q plots and the chi-square tests. Additionally, histograms with the density functions of the adequate models have been analyzed, and the correlation coefficient has been employed to determine the best model for each state.

The study demonstrates that the SMD distribution provided the best fit in sixteen states, whereas the Weibull and PP distributions each attained the best fit in only one state. Moreover, both empirical and theoretical income inequality measures such as the Gini, Pietra, Bonferroni, Atkinson, generalized entropy, and Frigyes measures, are computed and analyzed. Notably, all these income inequality measures attained their lowest values in Manipur and their highest values in Chhattisgarh, except for the empirical Frigyes measure, ρ, which peaked in Arunachal Pradesh. Future research could explore the factors contributing to the low-income inequality in Manipur and the elevated levels in Chhattisgarh. These insights provide valuable guidance for policy decisions to reduce income inequality in Chhattisgarh and foster more equitable economic development.

Table 5.4: Empirical and estimated income inequality measures across Indian states using per-capita household income data from the IHDS-II survey

State	Income inequality	Gini index	Pietra index	Atkinson index	Bonferroni index	Generalized entropy	Frigyes measures		
							ρ	ν	η
Jammu & Kashmir	empirical	0.47197	0.34250	0.18422	0.59410	0.38718	2.02279	4.17022	2.06161
	estimated-MLF	0.47263	0.34118	0.18434	0.59216	0.38744	2.02779	4.14444	2.04382
Himachal Pradesh	empirical	0.51711	0.38265	0.21896	0.63198	0.46494	2.19652	5.02117	2.28597
	estimated-SMD	0.51746	0.38285	0.22108	0.63244	0.46974	2.21073	5.02330	2.27224
Punjab	empirical	0.50978	0.37671	0.21538	0.61520	0.45684	2.07195	4.94292	2.38563
	estimated-SMD	0.51008	0.38108	0.21656	0.61667	0.45950	2.07736	5.06251	2.43699
Uttarakhand	empirical	0.51686	0.38525	0.21749	0.63319	0.46161	2.16546	5.10102	2.35563
	estimated- SM	0.51796	0.37493	0.22758	0.63737	0.48451	2.14787	4.84625	2.25630
Haryana	empirical	0.52339	0.38327	0.23273	0.63578	0.49624	2.14672	5.05920	2.35671
	estimated-MLF	0.52368	0.38067	0.23280	0.63252	0.49639	2.13422	4.99780	2.34175
Rajasthan	empirical	0.51944	0.38191	0.22653	0.62949	0.48212	2.06866	5.09884	2.46480
	estimated-SMD	0.51963	0.38134	0.22879	0.62852	0.48725	2.10468	5.03956	2.39446
Uttar Pradesh	empirical	0.52918	0.38957	0.23277	0.64131	0.49634	2.14186	5.24086	2.44688
	estimated-SMD	0.52932	0.38891	0.23599	0.64055	0.50370	2.16884	5.19911	2.39719
Bihar	empirical	0.53260	0.38973	0.24232	0.64034	0.51820	2.09168	5.31031	2.53877
	estimated-Dagum	0.53295	0.38888	0.24677	0.63864	0.52844	2.11620	5.24850	2.48016

Table 5.4: (continued)

State	Income inequality	Gini index	Pietra index	Atkinson index	Bonferroni index	Generalized entropy	Frigyes measures		
							ρ	ν	η
Sikkim	empirical	0.44097	0.32409	0.15793	0.54773	0.32942	1.88890	3.85525	2.04100
	estimated-Dagum	0.44517	0.31915	0.16906	0.55753	0.35376	1.85029	3.78325	2.04469
Arunachal Pradesh	empirical	0.52866	0.39937	0.22523	0.65221	0.47917	2.59551	5.55445	2.14002
	estimated-SM	0.53205	0.39032	0.23949	0.66759	0.51170	2.43156	5.24753	2.15809
Nagaland	empirical	0.50375	0.36218	0.21249	0.63152	0.45034	2.04401	4.58879	2.24499
	estimated-SM	0.50837	0.36903	0.21677	0.63903	0.45998	2.22557	4.71547	2.11877
Manipur	empirical	0.32490	0.23528	0.08786	0.45461	0.17975	1.63697	2.61116	1.59512
	estimated- Weibull	0.32873	0.23656	0.09197	0.46983	0.18837	1.73429	2.66398	1.53607
Mizoram	empirical	0.42193	0.30957	0.14267	0.54263	0.29632	1.86691	3.60074	1.92871
	estimated- SMD	0.42741	0.31198	0.14766	0.55187	0.30710	1.95856	3.64384	1.86047
Tripura	empirical	0.42333	0.30785	0.14652	0.53883	0.30464	1.81953	3.58924	1.97262
	estimated-SMD	0.42528	0.30901	0.14823	0.54134	0.30834	1.83424	3.60223	1.96388
Meghalaya	empirical	0.50059	0.37007	0.20593	0.62581	0.43558	2.19751	4.73179	2.15325
	estimated-SM	0.50436	0.36597	0.21326	0.63567	0.45207	2.21394	4.65033	2.10048
Assam	empirical	0.49957	0.37160	0.20257	0.61759	0.42805	2.15134	4.76695	2.21581
	estimated-SMD	0.50008	0.37030	0.20428	0.61659	0.43188	2.17036	4.73556	2.18192

Table 5.4: (continued)

State	Income inequality	Gini index	Pietra index	Atkinson index	Bonferroni index	Generalized entropy	Frigyes measures		
							ρ	ν	η
West Bengal	empirical	0.53465	0.40113	0.23268	0.64257	0.49612	2.22345	5.51464	2.48021
	estimated-SMD	0.53488	0.40232	0.23531	0.64260	0.50213	2.25614	5.52822	2.45030
Jharkhand	empirical	0.55000	0.41488	0.24883	0.66151	0.53320	2.24347	5.94469	2.64977
	estimated-SMD	0.55065	0.41086	0.25218	0.66070	0.54094	2.33014	5.74622	2.46605
Orissa	empirical	0.54833	0.40998	0.25053	0.64859	0.53711	2.18044	5.86370	2.68923
	estimated-SMD	0.54860	0.41459	0.25331	0.64988	0.54354	2.20346	5.99560	2.72099
Chattisgarh	empirical	0.59610	0.45402	0.29500	0.69164	0.64143	2.38856	7.34012	3.07303
	estimated-SMD	0.59655	0.45712	0.30034	0.69286	0.65417	2.45317	7.36463	3.00208
Madhya Pradesh	empirical	0.53963	0.39475	0.24792	0.64725	0.53111	2.12692	5.42001	2.54829
	estimated-SMD	0.53981	0.39796	0.25022	0.64773	0.53640	2.13491	5.51465	2.58308
Gujarat	empirical	0.56442	0.41612	0.26721	0.67537	0.57588	2.31132	5.91944	2.56107
	estimated-SMD	0.56472	0.41643	0.27105	0.67537	0.58485	2.30983	5.93090	2.56768
Maharashtra	empirical	0.47308	0.34928	0.18167	0.59678	0.38154	2.10157	4.30167	2.04689
	estimated-SMD	0.47322	0.34791	0.18193	0.59540	0.38211	2.11303	4.27827	2.02471
Andhra Pradesh	empirical	0.45351	0.32309	0.17426	0.57721	0.36519	1.88833	3.83617	2.03152
	estimated-SMD	0.45373	0.32214	0.17539	0.57602	0.36767	1.90832	3.81040	1.99673

Table 5.4: (continued)

State	Income inequality	Gini index	Pietra index	Atkinson index	Bonferroni index	Generalized entropy	Frigyes measures		
							ρ	ν	η
Karnataka	empirical	0.52182	0.38104	0.23078	0.63325	0.49180	2.07818	5.06011	2.43487
	estimated-Dagum	0.52196	0.37891	0.23524	0.63207	0.50199	2.09317	4.98280	2.38051
Goa	empirical	0.38889	0.27668	0.13541	0.52831	0.28066	1.78663	3.11697	1.74460
	estimated-PP	0.39103	0.27780	0.12923	0.52790	0.26740	1.83948	3.14550	1.70999
Kerala	empirical	0.42557	0.30256	0.15437	0.55339	0.32166	1.85341	3.48871	1.88232
	estimated-SM	0.42585	0.30399	0.15138	0.55110	0.31517	1.86535	3.51033	1.88186
Tamil Nadu	empirical	0.45391	0.32394	0.17442	0.57346	0.36553	1.88827	3.85274	2.04035
	estimated-SMD	0.45414	0.32367	0.17529	0.57317	0.36746	1.87255	3.85716	2.05984

Table 5.5: Ranking of Indian states by estimated income inequality measures using per-capita household income data from the IHDS-II survey

Gini index	Pietra index	Bonferroni index	Atkinson index	Generalized entropy	Frigyes measures			Rank
					ρ	ν	η	
Manipur	Manipur	Manipur	Manipur	Manipur	Manipur	Manipur	Manipur	1
Goa	Goa	Goa	Goa	Goa	Tripura	Goa	Goa	2
Tripura	Kerala	Tripura	Mizoram	Mizoram	Goa	Kerala	Mizoram	3
Kerala	Tripura	Kerala	Tripura	Tripura	Sikkim	Tripura	Kerala	4
Mizoram	Mizoram	Mizoram	Kerala	Kerala	Kerala	Mizoram	Tripura	5
Sikkim	Sikkim	Sikkim	Sikkim	Sikkim	Tamil Nadu	Sikkim	Andhra Pradesh	6
Andhra Pradesh	Andhra Pradesh	Tamil Nadu	Tamil Nadu	Tamil Nadu	Andhra Pradesh	Andhra Pradesh	Maharashtra	7
Tamil Nadu	Tamil Nadu	Andhra Pradesh	Andhra Pradesh	Andhra Pradesh	Mizoram	Tamil Nadu	Jammu & Kashmir	8
Jammu & Kashmir	Jammu & Kashmir	Jammu & Kashmir	Maharashtra	Maharashtra	Jammu & Kashmir	Jammu & Kashmir	Sikkim	9
Maharashtra	Maharashtra	Maharashtra	Jammu & Kashmir	Jammu & Kashmir	Punjab	Maharashtra	Tamil Nadu	10
Assam	Meghalaya	Assam	Assam	Assam	Karnataka	Meghalaya	Meghalaya	11
Meghalaya	Nagaland	Punjab	Meghalaya	Meghalaya	Rajasthan	Nagaland	Nagaland	12
Nagaland	Assam	Rajasthan	Punjab	Punjab	Maharashtra	Assam	Arunachal Pradesh	13
Punjab	Uttarakhand	Karnataka	Nagaland	Nagaland	Bihar	Uttarakhand	Assam	14

Table 5.5: (continued)

Gini index	Pietra index	Bonferroni index	Atkinson index	Generalized entropy	Frigyes measures			Rank
					ρ	ν	η	
Himachal Pradesh	Karnataka	Himachal Pradesh	Himachal Pradesh	Himachal Pradesh	Haryana	Karnataka	Uttarakhand	15
Uttarakhand	Haryana	Haryana	Uttarakhand	Uttarakhand	Madhya Pradesh	Haryana	Himachal Pradesh	16
Rajasthan	Punjab	Meghalaya	Rajasthan	Rajasthan	Uttarakhand	Himachal Pradesh	Haryana	17
Karnataka	Rajasthan	Uttarakhand	Haryana	Haryana	Uttar Pradesh	Rajasthan	Karnataka	18
Haryana	Himachal Pradesh	Bihar	Karnataka	Karnataka	Assam	Punjab	Rajasthan	19
Uttar Pradesh	Bihar	Nagaland	West Bengal	West Bengal	Orissa	Uttar Pradesh	Uttar Pradesh	20
Arunachal Pradesh	Uttar Pradesh	Uttar Pradesh	Uttar Pradesh	Uttar Pradesh	Himachal Pradesh	Arunachal Pradesh	Punjab	21
Bihar	Arunachal Pradesh	West Bengal	Arunachal Pradesh	Arunachal Pradesh	Meghalaya	Bihar	West Bengal	22
West Bengal	Madhya Pradesh	Madhya Pradesh	Bihar	Bihar	Nagaland	Madhya Pradesh	Jharkhand	23
Madhya Pradesh	West Bengal	Orissa	Madhya Pradesh	Madhya Pradesh	West Bengal	West Bengal	Bihar	24
Orissa	Jharkhand	Jharkhand	Jharkhand	Jharkhand	Gujarat	Jharkhand	Gujarat	25
Jharkhand	Orissa	Arunachal Pradesh	Orissa	Orissa	Jharkhand	Gujarat	Madhya Pradesh	26
Gujarat	Gujarat	Gujarat	Gujarat	Gujarat	Arunachal Pradesh	Orissa	Orissa	27
Chhattisgarh	Chhattisgarh	Chhattisgarh	Chhattisgarh	Chhattisgarh	Chhattisgarh	Chhattisgarh	Chhattisgarh	28

CHAPTER 6

CONLUSIONS AND RECOMMENDATIONS

Income distributions and inequality measures are widely discussed topics in modern econometrics and statistics. They provide insights into the distribution of wealth and resources among different populations. Analyzing these measures is essential for understanding economic disparities and informing policies for sustainable growth.

This work explores the significance of quantile functions in modeling and analyzing income data. Chapter 1 covers the fundamental properties of quantile functions, along with income inequality and poverty measures, discussed within the quantile framework.

A detailed literature review comprising four sections, including distribution function-based income distributions, quantile function-based income distributions, quantile functions in reliability analysis, and income inequality measures, is presented in

Chapter 2. The Lorenz curve, Gini index, Pietra index, Bonferroni index, Bonferroni curve, and Zenga curve for five distributions are also derived in this chapter. Additionally, the effectiveness of three quantile functions in income modeling is explored.

A thorough quantile-based income analysis of PP distribution has been carried out in Chapter 3. Major income inequality measures are derived, and the Lorenz ordering associated with the PP distribution is examined. A simulation study is conducted to evaluate the performance of three estimation methods, and the method of L-moments outperforms the others. We utilized the PP distribution to model two real income datasets, and both empirical and theoretical income inequality measures are computed.

Motivated by the property that the sum of two quantile functions results in another quantile function, we proposed the SMD distribution in Chapter 4. Several distributional properties, major income inequality, and poverty measures of the SMD distribution are derived. We transformed the poverty gap ratio and FGT measures into quantile form. The parameters are estimated using the method of L-moments, and the practical application of the SMD distribution is demonstrated using a real-life income dataset.

In Chapter 5, the six parametric models such as Weibull, PP, SM, Dagum, SMD, and MLF having closed-form quantile functions are utilized to model per capita household income data across Indian states. The L-moment estimation technique is employed to estimate the parameters of the models. We primarily used the Q-Q plot and chi-square test to identify the most suitable model for each state, followed by the histogram and correlation coefficient. Then for each state, we derived empirical and theoretical income inequality measures.

In this paragraph, we present additional research areas that could build upon this thesis. One possible direction is to interpret the parameters of each income distribution as income inequality parameters based on the study of their income inequality measures. As a future endeavor, one could explore the non-parametric estimation of income inequality measures using U-statistics, particularly for those measures that currently lack established non-parametric estimators. Our focus has primarily been on unimodal income data, however, in real-life situations, bimodal income patterns may also be observed. Therefore, it would be valuable to introduce quantile-based distributions capable of modeling bimodal income data. In this thesis, we consider only univariate income data. The analysis of high-dimensional income datasets using quantile functions remains an open area of research. The use of quantile-based approaches to study various income inequality and poverty measures in a multivariate setup has yet to be explored.

PUBLICATIONS AND PRESENTATIONS

List of published/ accepted papers

(1) Ashlin, V. and Haritha, N. H. (2023). A Review on Quantile Functions, Income Distributions, and Income Inequality Measures. Reliability: Theory & Applications, 18(2 (73)):50–62.

(2) Ashlin, V. and Haritha, N. H. (2024). Singh Maddala Dagum Distribution with Application to Income Data. Statistics and Applications, 22(1):321–341.

(3) Ashlin, V. and Haritha, N. H. (2025). Comparison of Income Inequality Among Indian States Using Quantile Functions. Computational Economics.

(4) Ashlin, V. and Haritha, N. H. (2025). A Quantile-Based Income Analysis of Power-Pareto Distribution. Statistica. (Accepted)

List of presentations

(1) Ashlin Varkey presented a paper on "A Quantile-Based Income Analysis of Power-Pareto distribution" in the Eighth International Conference on Statistics for Twenty-first Century-2022 (ICSTC-2022) organized by the International Statistical Fraternity (ISF), School of Physics and Mathematical Sciences and Department of Statistics, University of Kerala, Trivandrum on 16-19 December 2022.

(2) Ashlin Varkey presented a paper on "Singh Maddala Dagum Distribution" in the International Conference on Statistics, Probability, Data Sciences and Related areas organized jointly by the Department of Statistics, CUSAT and Indian Society for Probability and Statistics on 04-06 January 2023.

(3) Ashlin Varkey presented a paper on "Comparison of Income Inequality Among States of India Using Quantile Functions" in the three day International Conference on Statistics and Data Science (ICSDS-2024) organized by the Post-graduate Department of Statistics, Sree Sankara College, Kalady, in conjunction with the 45^{th} Annual Conference of the Kerala Statistical Association on 22-24 January 2024.

(4) Ashlin Varkey presented a contributed paper on "A Quantile-Based Study of Residual Income" in the Annual Conference of International Indian Statistical Association (IISA) jointly organized by the Department of Statistics, CUSAT & IISA on 27-31 December 2024.

BIBLIOGRAPHY

[1] Aaberge, R. (2000). Characterization of Lorenz curves and income distributions. Social Choice and Welfare, 17(4):639–653.

[2] Abd Raof, A. S., Haron, M. A., Safari, M. A. M., and Siri, Z. (2022). Modeling the incomes of the upper-class group in Malaysia using new Pareto-type distribution. Sains Malaysiana, 51(10):3437–3448.

[3] Aitchison, J. and Brown, J. A. C. (1957). The Lognormal Distribution. Cambridge University Press, Cambridge.

[4] Amoroso, L. (1925). Ricerche intorno alla curva dei redditi. Annali di Matematica Pura ed Applicata, 2:123–159.

[5] Arnold, B. C. (1983). Pareto Distributions. International Co-operative Publishing House, Fairland, MD.

[6] Arnold, B. C. (1987). Some related orderings. In Majorization and the Lorenz Order: A Brief Introduction, pages 77–89. Springer, New York.

[7] Arnold, B. C. (2015). On Zenga and Bonferroni curves. METRON, 73(1):25–30.

[8] Arora, S. and Jain, K. (2006). Testing for generalized Lorenz dominance. Statistical Methods and Applications, 15:75–88.

[9] Aswin, I. C., Sankaran, P. G., and Sunoj, S. M. (2020). A class of distributions with quadratic hazard quantile function. Journal of the Indian Society for Probability and Statistics, 21:409–426.

[10] Atkinson, A. B. (1970). On the measurement of inequality. Journal of Economic Theory, 2(3):244–263.

[11] Bakar, S. A. A. and Dharini, P. (2020). Income modeling with the Weibull mixtures. Communications in Statistics - Theory and Methods, 51(11):3612–3628.

[12] Banerjee, S., Chakrabarti, B. K., Mitra, M., and Mutuswami, S. (2020a). Inequality measures: The Kolkata index in comparison with other measures. Frontiers in Physics, 8:562182.

[13] Banerjee, S., Chakrabarti, B. K., Mitra, M., and Mutuswami, S. (2020b). On the Kolkata index as a measure of income inequality. Physica A: Statistical

Mechanics and its Applications, 545:123178.

[14] Barroso, F. J. C., García-Pérez, C., and Prieto-Alaiz, M. (2020). Modelling income distribution using the log Student's t distribution:New evidence for European Union countries. Economic Modelling, 89:512–522.

[15] Bartels, C. P. A. (1977). Economic Aspects of Regional Welfare: Income Distribution and Unemployment. Martinus Nijhoff, Leiden.

[16] Behdani, Z. and Borzadaran, G. M. (2019). Measures of income inequality based on quantile function. In Proceeding of the 5^{th} *Seminar on Reliability Theory and its Applications*.

[17] Bijamma, T., Midhu, N. N., and Sankaran, P. G. (2014). A software reliability model using quantile function. Journal of Probability and Statistics, 2014:1–8.

[18] Bonferroni, C. E. (1930). Elements of General Statistics. Seeber, Florence.

[19] Bourguignon, M., Saulo, H., and Fernandez, R. N. (2016). A new Pareto-type distribution with applications in reliability and income data. Physica A: Statistical Mechanics and its Applications, 457:166–175.

[20] Bowley, A. (1920). Elements of Statistics, volume 2. P. S. King & Son, Limited.

[21] Brazauskas, V., Greselin, F., and Zitikis, R. (2024). Measuring income in-

equality via percentile relativities. Quality & Quantity, 58:4859–4896.

[22] Brzezinski, M. (2015). Power laws in citation distributions: Evidence from Scopus. Scientometrics, 103:213–228.

[23] Burrell, Q. L. (1991). The Bradford distribution and the Gini index. Scientometrics, 21(2):181–194.

[24] Burrell, Q. L. (2005). Symmetry and other transformation features of Lorenz / Leimkuhler representations of informetric data. Information Processing & Management, 41(6):1317–1329. Special Issue on Infometrics.

[25] Butler, R. J. and McDonald, J. B. (1989). Using incomplete moments to measure inequality. Journal of Econometrics, 42(1):109–119.

[26] Caeiro, F. and Norouzirad, M. (2024). Comparing estimation methods for the Power–Pareto distribution. Econometrics, 12(3).

[27] Calderín-Ojeda, E., Azpitarte, F., and Gómez-Déniz, E. (2016). Modelling income data using two extensions of the exponential distribution. Physica A: Statistical Mechanics and its Applications, 461:756–766.

[28] Census Bureau, U. S. (2007). County and City Data Book:2007. Washington,DC.

[29] Chatterjee, A., Ghosh, A., and Chakrabarti, B. K. (2017). Socio-economic inequality: Relationship between Gini and Kolkata indices. Physica A:

Statistical Mechanics and its Applications, 466:583–595.

[30] Chotikapanich, D. (1993). A comparison of alternative functional forms for the Lorenz curve. Economics Letters, 41(2):129–138.

[31] Chotikapanich, D., editor (2008). Modeling Income Distributions and Lorenz Curves, volume 5. Springer, New York.

[32] Chotikapanich, D., Griffiths, W., Karunarathne, W., and Prasada Rao, D. S. (2013). Calculating poverty measures from the generalized beta income distribution. Economic Record, 89(S1):48–66.

[33] Clementi, F. (2023). The Kaniadakis distribution for the analysis of income and wealth data. Entropy, 25(8):1141.

[34] Clementi, F., Di Matteo, T., Gallegati, M., and Kaniadakis, G. (2008). The κ-generalized distribution: A new descriptive model for the size distribution of incomes. Physica A: Statistical Mechanics and its Applications, 387(13):3201–3208.

[35] Clementi, F., Gallegati, M., and Kaniadakis, G. (2007). κ-generalized statistics in personal income distribution. The European Physical Journal B, 57:187–193.

[36] Clementi, F., Gallegati, M., and Kaniadakis, G. (2009). A κ-generalized statistical mechanics approach to income analysis. Journal of Statistical

Mechanics: Theory and Experiment, 2009(2):P02037.

[37] Clementi, F., Gallegati, M., Kaniadakis, G., and Landini, S. (2016). κ-generalized models of income and wealth distributions: A survey. The European Physical Journal Special Topics, 225:1959–1984.

[38] Cowell, F. A. and Kuga, K. (1981). Additivity and the entropy concept: An axiomatic approach to inequality measurement. Journal of Economic Theory, 25(1):131–143.

[39] D'Addario, R. (1974). Intorno ad una funzione di distribuzione. Giornale degli Economisti e Annali di Economia, 33:205–214.

[40] Dagum, C. (1977). A new model of personal income distribution: Specification and estimation. Économie Appliquée, 30(3):413–437.

[41] Dagum, C. (1980). Inequality measures between income distributions with applications. Econometrica, 48(7):1791–1803.

[42] Dalton, H. (1920). The measurement of the inequality of incomes. Economic Journal, 30(119):348–461.

[43] Deepthy, G. S., Nicy, S., and Chandra, N. (2023). Applications of Burr III-Weibull quantile function in reliability analysis. Statistical Theory and Related Fields, 7(4):296–308.

[44] Desai, S. and Vanneman, R. (2018). India Human Development Survey-II

(IHDS-II), 2011-12. Inter-university Consortium for Political and Social Research [distributor].

[45] Desai, S., Vanneman, R., and National Council of Applied Economic Research, N. D. (2018). India Human Development Survey (IHDS), 2005. Inter-university Consortium for Political and Social Research [distributor].

[46] Dileep Kumar, M. and Sankaran, P. G. (2019). Reliability modelling and analysis using quantile functions. Doctoral dissertation, Cochin University of Science and Technology, Cochin.

[47] Donaldson, D. and Weymark, J. A. (1980). A single-parameter generalization of the Gini indices of inequality. Journal of Economic Theory, 22(1):67–86.

[48] Donaldson, D. and Weymark, J. A. (1983). Ethically flexible Gini indices for income distributions in the continuum. Journal of Economic Theory, 29(2):353–358.

[49] Duclos, J. Y. and Araar, A. (2006). Poverty and Equity: Measurement, Policy and Estimation with DAD. Springer, New York.

[50] Eltetö, Ö. and Frigyes, E. (1968). New income inequality measures as efficient tools for causal analysis and planning. Econometrica, 36(2):383–396.

[51] Fellman, J. (2012). Properties of Lorenz curves for transformed income distributions. Theoretical Economics Letters, 2(5):487–493.

[52] Fellman, J. (2018). Income inequality measures. Theoretical Economics Letters, 8:557–574.

[53] Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. Econometrica, 52(3):761–766.

[54] 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(10):3547–3567.

[55] Frigyes, E. (1965). Analysis and planning of the income distribution of workers and employees (in Hungarian). candidate's dissertation, Budapest.

[56] 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(322):33–46.

[57] Gastwirth, J. L. (1971). A general definition of the Lorenz curve. Econometrica, 39(6):1037–1039.

[58] Ghosh, A., Chattopadhyay, N., and Chakrabarti, B. K. (2014). Inequality in societies, academic institutions and science journals: Gini and k-indices. Physica A: Statistical Mechanics and its Applications, 410:30–34.

[59] Gibrat, R. (1931). Les inégalités économiques. Sirey, Paris.

[60] Gilchrist, W. (2000). Statistical Modeling with Quantile Functions (1st ed.).

Chapman and Hall/CRC, New York.

[61] Gini, C. (1912). Variabilità e Mutabilità. Bologna.

[62] 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.

[63] Giorgi, G. and Crescenzi, M. (2001). A proposal of poverty measures based on the Bonferroni inequality index. Metron - International Journal of Statistics, 59:3–16.

[64] Giorgi, G. M. and Nadarajah, S. (2010). Bonferroni and Gini indices for various parametric families of distributions. Metron - International Journal of Statistics, 68(1):23–46.

[65] Gómez-Déniz, E. (2016). A family of arctan Lorenz curves. Empirical Economics, 51:1215–1233.

[66] Gómez-Déniz, E., Sarabia, J. M., and Jordá, V. (2022). Parametric Lorenz curves based on the beta system of distributions. Communications in Statistics - Theory and Methods, 51(23):8371–8390.

[67] 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(5):1049–1054.

[68] Guerra, R. R., Peña-Ramírez, F. A., and Cordeiro, G. M. (2023). The Logistic Burr XII distribution: Properties and applications to income data. Stats, 6(4):1260–1279.

[69] Hankin, R. K. S. and Lee, A. (2006). A new family of non-negative distributions. Australian & New Zealand Journal of Statistics, 48(1):67–78.

[70] Haritha, N. H. and Nair, K. R. M. (2005). Models of income distributions and measures of income inequality - A review. Journal of Indian Society for Probability and Statistics, 9:54–76.

[71] 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.

[72] Haritha, N. H., Nair, N. U., and Nair, K. R. M. (2008). Modelling income using generalized lambda distributions. Journal of income distribution, 17(2):37–51.

[73] 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(3):413–426.

[74] Hlasny, V. (2021). Parametric representation of the top of income distri-

butions: Options, historical evidence, and model selection. Journal of Economic Surveys, 35(4):1217–1256.

[75] Hogg, R. V. and Klugman, S. A. (1983). On the estimation of long tailed skewed distributions with actuarial applications. Journal of Econometrics, 23(1):91–102.

[76] Hosking, J. (1996). Some Theoretical Results Concerning L-moments. IBM Thomas J. Watson Research Division.

[77] 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(1):105–124.

[78] Huang, S. and Oluyede, B. O. (2014). Exponentiated Kumaraswamy-Dagum distribution with applications to income and lifetime data. Journal of Statistical Distributions and Applications, 1(1):8.

[79] Jeena, J. and Asisha, A. P. (2021). Power-exponential geometric quantile function. Reliability: Theory & Applications, 16(4):294–307.

[80] Jones, M. (2004). On some expressions for variance, covariance, skewness and L-moments. Journal of Statistical Planning and Inference, 126(1):97–106.

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

[82] Kakwani, N. (1980b). On a class of poverty measures. Econometrica, 48(2):437–446.

[83] Kakwani, N. (1999). Inequality, Welfare and Poverty: Three Interrelated Phenomena. In Silber, J., editor, Handbook of Income Inequality Measurement, pages 599–634. Springer Netherlands, Dordrecht.

[84] Kämpke, T. and Radermacher, F. (2015). Income Modeling and Balancing: A Rigorous Treatment of Distribution Patterns, volume 679. Springer International Publishing, Cham.

[85] Karian, Z. A. and Dudewicz, E. J. (1999). Fitting the generalized lambda distribution to data: A method based on percentiles. Communications in Statistics - Simulation and Computation, 28(3):793–819.

[86] Kendall, M. G. (1940). Note on the distribution of quantiles for large samples. Supplement to the Journal of the Royal Statistical Society, 7(1):83–85.

[87] Kleiber, C. and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences. Wiley, New Jersey.

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

[89] Kumar, D. (2017). The Singh–Maddala distribution: Properties and estimation. International Journal of System Assurance Engineering and Manage-

ment, 8 (Suppl. 2):1297–1311.

[90] Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. Publications of the American Statistical Association, 9(70):209–219.

[91] Majid, M. H. A., Ibrahim, K., and Masseran, N. (2023). Three-part composite Pareto modeling for income distribution in Malaysia. Mathematics, 11(13):2899.

[92] March, L. (1898). Quelques exemples de distribution des salaires. Journal de la Société de Statistique de Paris, 39:193–206 and 241–248.

[93] McDonald, J. B. (1984). Some generalized functions for the size distribution of income. Econometrica, 52(3):647–663.

[94] Mdingi, K. and Ho, S. (2021). Literature review on income inequality and economic growth. MethodsX, 8:101402.

[95] Mehran, F. (1976). Linear measures of income inequality. Econometrica, 44(4):805–809.

[96] Mirzaei, S., Borzadaran, G. R. M., Amini, M., and Jabbari, H. (2018). A new generalized Weibull distribution in income economic inequality curves. Communications in Statistics - Theory and Methods, 48(4):889–908.

[97] Moors, J. J. A. (1988). A quantile alternative for kurtosis. Journal of the Royal Statistical Society. Series D (The Statistician), 37(1):25–32.

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

[99] Muliere, P. and Scarsini, M. (1989). A note on stochastic dominance and inequality measures. Journal of Economic Theory, 49(2):314–323.

[100] Nair, N. U., Nair, K. R. M., and Haritha, N. H. (2008). Some properties of income gap ratio and truncated Gini coefficient. Calcutta Statistical Association Bulletin, 60:245–254.

[101] Nair, N. U., Sankaran, P. G., and Balakrishnan, N. (2013a). Quantile-Based Reliability Analysis. Springer, New York.

[102] Nair, N. U., Sankaran, P. G., and Sunoj, S. M. (2013b). Quantile based stop-loss transform and its applications. Statistical Methods Applications, 22:167–182.

[103] Nair, N. U., Sankaran, P. G., and Sunoj, S. M. (2018). Proportional hazards model with quantile functions. Communications in Statistics - Theory and Methods, 47(19):4710–4723.

[104] Nair, N. U., Sankaran, P. G., and Vineshkumar, B. (2012). Modelling lifetimes by quantile functions using Parzen's score function. Statistics, 46(6):799–811.

[105] Nair, N. U., Subhash, S., and Sunoj, S. M. (2024). A simple method of

estimation and testing based on Q–Q plots. Operations Research Forum, 5.

[106] Nair, N. U. and Vineshkumar, B. (2010). L-moments of residual life. Journal of Statistical Planning and Inference, 140(9):2618–2631.

[107] Nair, N. U. and Vineshkumar, B. (2021). Properties of bivariate distributions represented through quantile functions. American Journal of Mathematical and Management Sciences, 42(1):1–12.

[108] Nair, N. U. and Vineshkumar, B. (2022). Modelling informetric data using quantile functions. Journal of Informetrics, 16(2):101266.

[109] Ord, J. K., Patil, G. P., and Taillie, C. (1981). The choice of a distribution to describe personal income. Statistical Distributions in Scientific Work, 6:193–202.

[110] Ord, J. K., Patil, G. P., and Taillie, C. (1983). Truncated distributions and measures of income inequality. Sankhyā: The Indian Journal of Statistics, Series B, 45(3):413–430.

[111] Pareto, V. (1897). Cours d'Économie Politique, volume 2. F. Rouge, Lausanne.

[112] Parzen, E. (1979). Nonparametric statistical data modeling. Journal of the American Statistical Association, 74(365):105–121.

[113] 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.

[114] Pillai, P. P. C., Rajesh, G., and Abdul Sathar, E. I. (2014). Semi-parametric estimation of Lorenz curve and Gini-index in an exponential distribution. Sri Lankan Journal of Applied Statistics, 15(1):19–30.

[115] Porro, F. (2015). Zenga distribution and inequality ordering. Communications in Statistics - Theory and Methods, 44(18):3967–3977.

[116] Pundir, S., Arora, S., and Jain, K. (2005). Bonferroni curve and the related statistical inference. Statistics & Probability Letters, 75(2):140–150.

[117] Pérez, C. G. and Alaiz, M. P. (2011). Using the Dagum model to explain changes in personal income distribution. Applied Economics, 43(28):4377–4386.

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

[119] Rao, U. L. G. and Tam, A. Y. P. (1987). An empirical study of selection and estimation of alternative models of the Lorenz curve. Journal of Applied Statistics, 14(3):275–280.

[120] Rasche, R. H., Gaffney, J., Koo, A. Y. C., and Obst, N. (1980). Functional forms for estimating the Lorenz curve. Econometrica, 48(4):1061–1062.

[121] Rattan, P., Arora, S., and Mahajan, K. K. (2021). Testing for multisample generalized Lorenz curves dominance. Journal of Statistics and Management Systems, 24(5):895–915.

[122] Rousseau, R. (1987). The nuclear zone of a Leimkuhler curve. Journal of Documentation, 43(4):322–333.

[123] Safari, M. A. M., Masseran, N., and Hussain, S. I. (2021a). Modeling urban household income in Malaysia using parametric approach. Journal of Physics: Conference Series, 1988:012090.

[124] Safari, M. A. M., Masseran, N., Ibrahim, K., and Hussain, S. I. (2019). A robust and efficient estimator for the tail index of inverse Pareto distribution. Physica A: Statistical Mechanics and its Applications, 517:431–439.

[125] Safari, M. A. M., Masseran, N., Ibrahim, K., and Hussain, S. I. (2020). Modeling the income distribution of poor households in Malaysia. In AIP Conference Proceedings, volume 2266, page 090004.

[126] Safari, M. A. M., Masseran, N., Ibrahim, K., and Hussain, S. I. (2021b). Measuring income inequality: A robust semi-parametric approach. Physica A: Statistical Mechanics and its Applications, 562:125359.

[127] Sánchez, L., Leiva, V., Galea, M., and Saulo, H. (2021). Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. Applied Stochastic Models in Business and Industry, 37(1):53–73.

[128] 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(2):566–582.

[129] Sankaran, P. G. and Preetha Kumari, P. V. (2010). A parametric quantile regression model useful in reliability analysis. Calcutta Statistical Association Bulletin, 62(3-4):183–206.

[130] Sarabia, J. M. (2008). A general definition of the Leimkuhler curve. Journal of Informetrics, 2(2):156–163.

[131] Sarabia, J. M., Gómez-Déniz, E., Sarabia, M., and Prieto, F. (2010). A general method for generating parametric Lorenz and Leimkuhler curves. Journal of Informetrics, 4(4):524–539.

[132] Sarabia, J. M. and Jordá, V. (2014). Explicit expressions of the Pietra index for the generalized function for the size distribution of income. Physica A: Statistical Mechanics and its Applications, 416:582–595.

[133] Sarabia, J. M., Jordá, V., and Remuzgo, L. (2017). The Theil indices in parametric families of income distributions—A short review. Review of

Income and Wealth, 63(4):867–880.

[134] Sarabia, J. M. and Sarabia, M. (2008). Explicit expressions for the Leimkuhler curve in parametric families. Information Processing & Management, 44(5):1808–1818.

[135] 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(2):448.

[136] Sen, A. (1976). Poverty: An ordinal approach to measurement. Econometrica, 44(2):219–231.

[137] Sen, P. K. (1986). The Gini coefficient and poverty indexes: Some reconciliations. Journal of the American Statistical Association, 81(396):1050–1057.

[138] 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(1):79–93.

[139] Shahzad, M. and Asghar, Z. (2013b). Parameter estimation of Singh Maddala distribution by moments. International Journal of Advanced Statistics and Probability, 1(3):121–131.

[140] 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(3):641–650.

[141] 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.

[142] Singh, S. K. and Maddala, G. S. (1976). A function for size distribution of incomes. Econometrica, 44(5):963–970.

[143] Sodomova, E., Pacak, B., and Sipkova, L. (2005). Models of households in the Slovak Republic. In Statistics in Management of Social and Economic Development: 11th Ukrainian-Polish-Slovak Scientific Conference, pages 80–91.

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

[145] Sudheesh, K. K., Dewan, I., and Sreelakshmi, N. (2021). Non-parametric estimation of Gini index with censored observations. Statistics & Probability Letters, 175:109113.

[146] Sudheesh, K. K., Sreelakshmi, N., and Balakrishnan, N. (2022). Non-parametric inference for Gini covariance and its variants. Sankhya A,

84(2):790–807.

[147] Takayama, N. (1979). Poverty, income inequality, and their measures: Professor Sen's axiomatic approach reconsidered. Econometrica, 47(3):747–59.

[148] Tarsitano, A. (2004). Fitting the generalised lambda distribution to income data. In COMPSTAT 2004 Symposium, pages 1861–1867. Springer.

[149] Theil, H. (1967). Economics and Information Theory, volume 7. North-Holland Publishing Company, Amsterdam.

[150] Thompson, W. A. (1976). Fisherman's luck. Biometrics, 32(2):265–271.

[151] Trzcińska, K. (2022). Income and inequality measures in households in Czech Republic and Poland based on Zenga distribution. Statistika: Statistics and Economy Journal, 102(1):46–58.

[152] Tukey, J. W. (1977). Exploratory Data Analysis. Addsion Wesley, Massachusetts.

[153] Tyagi, A. (2023). Income inequality in Indian states. Economic & Political Weekly, 58(12):60–67.

[154] Vartia, P. L. I. and Vartia, Y. O. (1980). Description of the income distribution by the scaled F-distribution model. In Klevmarken, N. A. and Lybeck, J. A., editors, The Statics and Dynamics of Income, pages 23–36. Tieto, Clevedon, UK.

[155] Vineshkumar, B. and Nair, N. U. (2019). Bivariate quantile functions and their applications to reliability modelling. Statistica, 79(1):3–21.

[156] Watts, H. W. (1968). An economic definition of poverty. In Moynihan, D. P., editor, On Understanding Poverty, pages 316–329. Basic Books, New York.

[157] Weibull, W. (1939a). The phenomenon of rupture in solids. Technical Report 153, Ingeniörs Vetenskaps Akademiens Handlingar, Stockholm.

[158] Weibull, W. (1939b). A statistical theory of the strength of material. Technical Report 151, Ingeniörs Vetenskaps Akademiens Handlingar, Stockholm.

[159] Yitzhaki, S. (1983). On an extension of the Gini inequality index. International Economic Review, 24(3):617–628.

[160] Zenga, M. (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(5/6):301–326.

[161] Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. Statistica & Applicazioni, 5:3–27.

[162] Zenga, M. (2010). Mixture of Polisicchio's truncated Pareto distributions with beta weights. Statistica & Applicazioni, 8(1):3–25.

[163] Zenga, M., Domma, F., and Latorre, G. (2012). The Dagum distribution in

reliability analysis. Statistica e Applicazioni, X:97–113.

[164] Zheng, B. (1993). An axiomatic characterization of the Watts poverty index. Economics Letters, 42(1):81–86.