ECONOMIC STUDIES IN INEQUALITY, SOCIAL EXCLUSION AND WELL-BEING

Modeling Income Distributions and Lorenz Curves

Duangkamon Chotikapanich
Editor

Springer

Modeling Income Distributions and Lorenz Curves

ECONOMIC STUDIES IN EQUALITY, SOCIAL EXCLUSION AND WELL-BEING

This series will publish volumes that go beyond the traditional concepts of consumption, income or wealth and will offer a broad, inclusive view of inequality and well-being. Specific areas of interest will include Capabilities and Inequalities, Discrimination and Segregation in the Labor Market, Equality of Opportunities, Globalization and Inequality, Human Development and the Quality of Life, Income and Social Mobility, Inequality and Development, Inequality and Happiness, Inequality and Malnutrition, Inequality in Consumption and Time Use, Inequalities in Health and Education, Multidimensional Inequality and Poverty Measurement, Polarization among Children and Elderly People, Social Policy and the Welfare State, and Wealth Distribution.

Volume 1
De Janvry, Alain and Kanbur, Ravi
Poverty, Inequality and Development: Essays in Honor of Erik Thorbecke

Volume 2
Duclos, Jean-Yves and Araar, Abdelkrim
Poverty and Equity: Measurement, Policy and Estimation with DAD

Volume 3
Lemmi, Achille and Betti, Gianni
Fuzzy Set Approach to Multidimensional Poverty Measurement

Volume 4
Wagle, Udaya
Multidimensional Poverty Measurement: Concepts and Applications

Volume 5
Chotikapanich, Duangkamon
Modeling Income Distributions and Lorenz Curves

Modeling Income Distributions and Lorenz Curves

Edited by

Duangkamon Chotikapanich
Department of Econometrics and Business Statistics
Monash University
26 Sir John Monash Drive
Caulfield East VIC 3145
Australia

Springer

Editor

Duangkamon Chotikapanich
Department of Econometrics and Business Statistics
Monash University
26 Sir John Monash Drive
Caulfield East VIC 3145
Australia
Email: Duangkamon.Chotikapanich@buseco.monash.edu.au

Series Editor

Jacques Silber
Bar-Ilan University
Ramat Gan
Israel

ISBN: 978-0-387-72756-1 e-ISBN: 978-0-387-72796-7

Library of Congress Control Number: 2008928717

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

In Memory of
Camilo Dagum

Foreword

Jean-Jacques Rousseau wrote in the Preface to his famous Discourse on Inequality that "I consider the subject of the following discourse as one of the most interesting questions philosophy can propose, and unhappily for us, one of the most thorny that philosophers can have to solve. For how shall we know the source of inequality between men, if we do not begin by knowing mankind?" (Rousseau, 1754). This citation of Rousseau appears in an article in Spanish where Dagum (2001), in the memory of whom this book is published, also cites Socrates who said that the only useful knowledge is that which makes us better and Seneca who wrote that knowing what a straight line is, is not important if we do not know what rectitude is.

These references are indeed a good illustration of Dagum's vast knowledge, which was clearly not limited to the field of Economics. For Camilo the first part of Rousseau's citation certainly justified his interest in the field of inequality which was at the centre of his scientific preoccupations. It should however be stressed that for Camilo the second part of the citation represented a "solid argument in favor of giving macroeconomic foundations to microeconomic behavior" (Dagum, 2001). More precisely, "individualism and methodological holism complete each other in contributing to the explanation of individual and social behavior" (Dagum, 2001).

These excerpts from Camilo's fascinating article are the best proof that, no matter how important his contributions to the measurement of income inequality were, his concerns went much beyond this topic. Nevertheless we have to acknowledge that, even at the technical level of measurement, Dagum's contributions to the study of inequality are considerable. In the Foreword to a special issue of the Journal of Economic Inequality published in memory of Camilo, the editors Achille Lemmi and Gianni Betti stressed that Dagum's contribution to the economic-econometric modeling of personal income and wealth distributions was of great importance. The three-parameter model which bears his name "still represents one of the most complete formalizations with regard to economic theory, stochastic derivation and possibilities for use in empirical analysis. Every parameter has a precise economic interpretation and the model fits a very complete series of formal-logic properties." (Lemmi and Betti, 2007).

The present volume entitled Modeling Income Distributions and Lorenz Curves contains a very nice collection of classical papers in the field of income distribution modeling (among which is Camilo's famous article), important surveys on the main themes of this book and original contributions to a field in which research remains very much alive. Duangkamon Chotikapanich, the editor of this volume, has done a superb job in convincing so many well-known researchers in this field to contribute a chapter. This is indeed a clear testimony of the respect so many economists and statisticians have for Camilo Dagum.

Jacques Silber
Editor of the Springer Book Series
on Economic Studies in Inequality,
Social Exclusion and Well-Being

References

Dagum, C. (2001) Desigualdad del Rdito y Bienestar Social, Descomposicin, Distancia Direccional y Distancia Mtrica Entre Distribuciones, *Estudios de Economia Aplicada*, **17**(1), 5–52.

Lemmi, A. and G. Betti (2007) Guest Editors' Introduction, *Journal of Economic Inequality*, **5**(3), forthcoming.

Rousseau, J.-J. (1754) Discourse on Inequality, *Public domain*, translated by G. D. H. Cole.

Introduction

This volume is a collection of papers, drawn together to honour the memory of Camilo Dagum and his outstanding contributions to the study of personal income distribution and inequality measures. It is part of the book series “Economic Studies in Inequality, Social Exclusion and Well-Being” edited by Jacques Silber. One of Professor Dagum’s significant contributions to this area is his 1977 paper, reprinted here, which introduces a new model for income distribution. This new model is widely used in empirical work and is also known as the Dagum model. To honour this contribution the focus of the book is on modeling income distributions and Lorenz curves.

The volume is organized in three parts. Part One is a collection of five influential papers that have had a significant impact on this area. Part Two contains four survey papers on Lorenz functions, and generalizations and extensions of some income distributions, while in Part Three there are eight papers on current research and development written by well-known scholars who have worked extensively in the area.

Part One begins with the 1977 paper by Dagum (1977) on a new model for the size distribution of incomes that satisfies a set of important assumptions. In this paper Dagum established empirical foundations in the form of properties for a probability function to describe the size distribution of income. Given the established properties he proposed a model to represent the distribution of income. This model later became known as the Dagum distribution and is now widely used in empirical studies as one of the models that well represents income distributions.

The second chapter in Part One is a reprint of the paper on the model for income distribution proposed by Singh and Maddala (1976). The model is a three parameter income distribution derived from a generalization of the Pareto and the Weibull distributions and is based on the concept of failure rate. This model is also used widely in empirical studies as an income distribution model that fits the data from various countries very well.

Chapter 3 is a reprint of the paper by McDonald (1984) on using two generalized beta distributions as a model for the size distribution of incomes. These two generalized beta distributions are four parameter distributions and they were shown to include the beta of the first kind, the beta of the second kind, the Singh-Maddala, the lognormal, gamma, Weibull, Fisk and exponential distributions as special cases.

The fourth chapter is on the Lorenz curve. It is a reprint of the paper by Kakwani and Podder (1976) on a new coordinate system and the efficient estimation of the Lorenz curve using grouped data. The new coordinate system is an innovative representation of the Lorenz curve which proves to fit the data very well.

The final paper in Part One is a reprint of Paap and van Dijk (1998). They study the distributions of real GDP per capita for a combined 120 countries over the period 1960 to 1989. These distributions appear to be bimodal. In this paper a mixture of Weibull and truncated normal densities is used to model the bimodal distributions.

Part Two of the volume starts with Chapter 6, a survey paper written by Christain Kleiber entitled "A Guide to the Dagum Distributions". This paper introduces the Dagum distributions and their interrelations with other statistical distributions. It provides the basic statistical properties and inferential aspects of the Dagum distributions and a survey of their applications in economics.

Chapter 7 written by Barry Arnold provides a survey paper on the classical Pareto model and a hierarchy of generalized Pareto models. The properties of these models are introduced and the related distributions and inferential issues are discussed. The paper concludes by introducing the multivariate Pareto distribution.

Chapter 8 is another survey paper, this time written by James B. McDonald on the use of the generalized beta distribution for income distributions. It derives some inequality measures, the Gini, Pietra and Theil indices, as functions of the distributional parameters. It explores the use of numerical methods to calculate inequality measures for the case of the generalized beta distribution.

Jose M. Sarabia contributes Chapter 9 on "Parametric Lorenz Curves: Models and Applications". This chapter includes the basic properties for a function to represent a Lorenz curve and the Lorenz specifications corresponding to different classical income distributions. A general method for obtaining a hierarchical family of Lorenz curves is introduced. Sarabia also derives the Lorenz ordering conditions for a large number of well-known income distributions and also introduces the concept of the multivariate Lorenz curve which is an extension of the Lorenz curve to higher dimensions.

Part Three starts with Chapter 10 written by Hang K. Ryu on "Maximum Entropy Estimation of Income Distribution from Bonferroni Indices". This paper proposes using the Bonferroni Index (BI) to measure inequality in the distribution of income. The BI is defined using the ratio of the area between the Lorenz curve and the horizontal axis to the area between the 45 degree line and the horizontal axis. Based on this definition, more weight is given to the lower income groups and less weight to the upper income groups. The paper proceeds to compare the performance of the BI and the Gini coefficient by comparing the underlying distributions derived from them using the maximum entropy method with the empirical distributions from the income deciles of 113 countries.

Chapter 11 is written by William J. Reed on "A New Four- and Five-Parameter Models for Income Distributions". This paper introduces two new models to represent income distributions. They are the normal-Laplace distribution (NL) with four parameters and the generalized normal-Laplace distribution (GLN) with five parameters. The properties and the maximum likelihood estimation method for these two models are discussed. These two functional forms are fitted to nine empirical income distributions and the performances are compared to the four and five generalized beta distributions. It was found that both NL and GLN outperform the generalized beta.

Gianni Betti, Antonella D'Agustino, and Achille Lemmi provide Chapter 12 on "Fuzzy Monetary Poverty Measures under a Dagum Income Distributive Hypothesis". This paper derives the *Integrated Fuzzy Relative* poverty measure under the assumption that income follows the Dagum distribution. The authors apply their approach to Italian data obtained from the EU-SILC survey conducted in 2004.

Chapter 13 by Frank A. Cowell and Maria-Pia Victoria-Feser is on "Modelling Lorenz Curves: Robust and Semi-Parametric Issues". This paper considers the semi-parametric Lorenz curve and the estimation problem associated with contaminated data that normally occurs in the upper tail of the distribution. The semi-parametric Lorenz curve considers fitting a parametric distribution to the data on incomes above a certain level and the incomes below that level are treated non-parametrically using the empirical distribution function. The paper uses a Pareto distribution for the parametric distribution fitted to the upper tail. This approach is demonstrated and applied to UK household disposable incomes for 1981 with 7470 observations.

Chapter 14 by J. M. Henle, N. J. Horton and S. J. Jakus is on "Modelling Inequality with a Single Parameter". In this paper a new single parameter model is proposed for the Lorenz curve. This new functional form is tested using decile share data on income for 89 countries from the Luxembourg Income Study. This new specification for the Lorenz curve can also be used to represent a dynamic model for income growth.

Chapter 15 on "Lorenz Curves and Generalised Entropy Inequality Measures" is written by Nicholas Rohde. The paper establishes the general relationship between the Theil T inequality measure and the Lorenz curve. Analytical expressions for the Theil index are also derived from three parametric Lorenz curves. The empirical validity of the relationship between the Theil index and the Lorenz curve is examined using a simulation experiment.

Chapter 16 is the penultimate chapter by Duangkamon Chotikapanich and Bill Griffiths on "Estimating Income Distributions Using a Mixture of Gamma Densities". A Bayesian inference procedure to estimate a gamma mixture with two and three components is introduced. The predictive density and distribution function of income are described. The flexibility of the mixture is illustrated using a sample of Canadian income data. The paper obtains the posterior density for the Lorenz curve ordinates and the Gini coefficient.

The last chapter, Chapter 17, is written by Quentin Wodon and Shlomo Yitzhaki on "Inequality in Multidimensional Indicators of Well-Being: Methodology and Application to the Human Development Index". This paper introduces the Human Development Index which, in general, is a weighted average of three well-being indices involving life expectancy, educational attainment, and per capita GDP. The weighting schemes used are arbitrary and normally depend on the purpose of the analysis. The paper investigates the extent to which the Human Development Index is sensitive to a change in the weights.

I wish to thank Jacques Silber for his kind invitation to edit this volume. I would like to thank all the authors for preparing their contributions and the referees for helping to improve the final result. Lastly, I wish to thank Andrey Kostenko, Kulan Ranasinghe and Ainura Tursunalieva for their excellent editorial assistance in getting this book into publishable form.

Duangkamon Chotikapanich

References

Dagum, C. (1977) A New Model for Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, **30**, 413–437.

Kakwani, T. and N. Podder (1976) Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations, *Econometrica*, **44-1**, 137–149.

McDonald, J. B. (1984) Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52**, 647–663.

Paap, R. and H. K. van Dijk (1998) Distribution and Mobility of Wealth of Nations, *European Economic Review*, **42**, 1269–1293.

Singh, S. K. and G. S. Maddala (1976) A Function for the Size Distribution of Incomes, *Econometrica*, **44**, 963–970.

Contents

Part I
Collection of influential papers

CHAPTER 1

A New Model of Personal Income Distribution: Specification and Estimation[†]

Camilo Dagum [‡]

Abstract

The research deduces a new model of income distribution by size from a set of elementary assumptions. Its main properties are analyzed and five methods of parameter estimation are proposed. The mathematical form of the Lorenz curve and the Gini concentration ratio associated with the specified model are also deduced. The model is fitted to the observed income distributions of four very dissimilar countries: Argentina, Canada, Sri Lanka and the USA. The fits obtained for the USA in 1960 and in 1969 are compared with those obtained using the lognormal, the gamma and the Singh-Maddala models, working with the sum of squares of deviations and the bounds for the Gini concentration ratio proposed by Gastwirth. In conclusion, the specified model fared better than the others for almost all of the fourteen properties introduced in this paper.

1 Introduction

Since Pareto (1895, 1896, 1897) started the exploration of the field of income distribution and proposed his celebrated models (the first, second and third Pareto laws) a

[†] Reprint of Dagum, C. (1977), A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliquee*, 30, 413-437. With permission from the editor (Rolande Borrelly).

[‡] A Canada Council research grant is gratefully acknowledged. For the parameter estimation of the model specified in this research a nonlinear program written by Professor L. G. Birta for the University of Ottawa was used. I wish to express my thanks to Professor Birta and to my research assistants Leslie Gunapatne and Philip Bonardelli.

variety of probability functions have been suggested as suitable in describing the distribution of income by size (personal income distribution). These functional forms can, in a first approximation, be grouped in the following three main categories:

i) Functional forms proposed to describe the generation of an income distribution, by means of a stochastic process. In this category are included the contributions made by Champernowne (1953), Fisk (1961), Gibrat (1931), Mandelbrot (1960, 1963) and Rutherford (1955).

ii) Functional forms proposed solely by their practical bearing upon the encountered empirical distributions, that is, on the grounds of presenting a satisfactory goodness of fit. In this category are included the Gamma model proposed by Ammon (1895), March (1898) and Salem and Mount (1974); the Beta model proposed by Thurow (1970) and by Kakwani and Podder (1976), after performing an ingenious 45° rotation of the coordinates; the Pearson Type V distribution proposed by Vinci (1921); the generalized Gamma distribution deduced by Amoroso (1924-1925), which contains, as particular cases, the Gamma and Pearson Type V distributions; the hyperbolic distribution proposed by Champernowne (1952); the Weibull distribution proposed by Bartels and van Metelen (1975); and the log t (where t is the Student distribution) studied by Kloek and van Dijk (1976). The models proposed by Amoroso and Vinci partly overlap with those belonging to the first category for they are deduced from considerations of probability. An outstanding synthesis is achieved by D'Addario (1949) who specified a differential equation which contains as particular cases among others, the Pareto, the lognormal, the Amoroso (hence the Gamma and the Pearson Type V) models.

iii) Specification of differential equations that purport to capture the characteristics of regularity and permanence observed in the empirical distributions of income. The functional form is the solution of the corresponding differential equation. In this category are included the models proposed by Pareto (1896), Singh and Maddala (1976) and Dagum (1975).

The models most frequently applied are the Pareto, the lognormal and the Gamma. According to its goodness of fit, functional simplicity and the economic interpretation of its parameters, the Pareto model continues to be the best one to describe high income groups. It has the strong limitation, however, of only being useful in describing the upper tail of the distribution.

The lognormal and the Gamma fit the whole range of income distributions but are quite poor in describing both the upper and lower tails of the actual distributions, which happen to be the most relevant pieces of information in any measure of income inequality and in the elaboration of an income distribution policy. Empirical evidence favors the Gamma over the lognormal distribution, judging by goodness of fit criterion, as shown by Salem and Mount (1974) for the USA and Bartels and van Metelen (1975) for the Netherlands. The model deduced by Singh and Maddala (1976) outperforms both the lognormal and the Gamma distributions, as far as the goodness of fit is concerned.

The purpose of this research is to provide empirical foundations for and the derivation of a probability function as a general model to describe the personal distribution of income.

The plan of the paper is as follows: section 2 analyses a set of "ideal" properties to guide the identification of the best possible model among the large set of competing models which describe the size distribution of income. The third section derives a new model from a set of assumptions characterizing the observed regularities in the income distributions from both developed and developing countries. This section also includes the mathematical formulae deduced for the mean, the mode, the median, the moments of order r, the Lorenz curve and the Gini concentration ratio. Section 4 analyses the identification of intersecting Lorenz curves from their corresponding Gini ratios. The next section proves the convergence of the model specified in section 3 to the Pareto distribution. Section 6 presents five methods of parameter estimation. The following section fits the specified model to the income distributions of Canada, Argentina, Sri Lanka and the USA. Finally, section 8 concludes this study.

2 Set of Properties to Motivate the Identification of a Mathematical Model of Income Distribution

Although Ammon (1895), March (1898) and Amoroso (1924-1925) postulated and applied the Gamma function to describe the distribution of income and although McAlister (1879) is recognized as the first one to specify and provide the foundations for a theory of the lognormal distribution, it was Gibrat's pioneer contribution in 1931 that brought to the fore the lognormal as a model of unimodal income distribution. This preference for the lognormal remained virtually undisputed for nearly four decades. In 1970 Thurow postulated the Beta distribution; in 1974, Salem and Mount revived the Gamma distribution; in 1975 Bartels and van Metelen applied the lognormal, Gamma and the Weibull distributions, and C. Dagum postulated the three-parameter log-logistic model; in 1976 Singh and Maddala identified a new model which can be considered as a member of the logistic family, and finally, in this paper a new model is identifies which is a further generalization of the one proposed by Dagum (1973, 1975). Hence, in the decade of the seventies we have witnessed in this field the specification of several competing models.

The final choice of a particular model may be governed by its capacity to account fairly well to a set of economic, econometric, stochastic and mathematical properties. Aitchison and Brown (1957, p. 108) and Metcalf (1972, pp. 16-17) state four properties as a guide to identify the most representative model of the unknown stochastic process that generates an income distribution in a given time and space and for a given class of income and for any given unit of income receivers. Their four properties are included in a larger set proposed to guide the search and the identification of a functional form as a model of personal income distribution. This set takes into account: *i*) the undisputed acceptability of the Pareto distribution as the model of high income groups; *ii)* the properties and shortcomings of the different distributions already specified; *iii)* the existence of an explicit and tractable mathematical solution for the Lorenz curve and the Gini concentration ratio; and *iv)* whether the underlying stochastic process that generated the income distribution was specified, or whether the model follows a strictly pragmatic approach, that is, the fact that it fits the data is the main reason for its choice. The tentative set of properties are:

2.1 The extent to which the mathematical form of the distribution function can be derived from an elementary set of logico-empirical postulates or assumptions

In agreement with Morgenstern (1963, p. 93), the word "elementary" must in this connection be used in both its meanings: as *simple* from a technical point of view and as *fundamental* from the point of view of model-building. The requirement of a logico-empirical set of postulates underlying the model-building is an *elementary* tenet in the philosophy and methodology of sciences. It is also an *elementary* tenet in Pareto's contribution to income distribution. His model is a far reaching *elementary* outcome of a set of postulates that represent the theoretical counterpart of the empirical observations for several countries, regions and towns, in different periods of time.

2.2 Parsimony

The mathematical form of all proposed models contains unknown parameters whose values must be estimated from the data. This important practical and theoretical property requires that we make use of the smallest possible number of parameters for adequate and meaningful representation. However, all biparametric models of income distribution have failed to be empirically corroborated. The Pareto distribution is the only biparametric model that systematically stands the test of reality *provided that* it is restricted to the high income groups (right tail of the distribution). Of course, three- and four-parameter models imply a loss in simplicity, but an accurate description of empirical distributions with an associated measure of income inequality requires the specification of a model which must be a function of more than two parameters. As Metcalf (1972, p. 17) observed, "since it is argued above that a two-parameter function is *too* simple to reflect the impact of economic fluctuations on the size distribution, a convenient three- (or possible four-) parameter function will be sought".

2.3 Economic interpretation of the parameters

This property requires that all the parameters of the specified model should have a well-defined economic meaning. Two parameters which are encountered often are the scale and the inequality parameters. The first one is related to the unit of measurement of income and the second is of dimension zero and is related to the inequality of the income distribution.

2.4 Goodness of fit

This is the problem of testing the agreement between the model being identified and the actual observations. An ideal model should provide a good fit of the whole range of the distribution because all observations are relevant for: *i*) an accurate measurement of income inequality; *ii)* the supporting of a given social and income policy and *iii)* determining the taxation structure.

2.5 Simple and efficient method of parameter estimation

Given the modern computer facilities and the available methods of numerical analysis, it is always feasible to obtain the maximum likelihood estimators. However, a simple and efficient method of parameter estimation is always an advantage from the point of view of computer cost and the acceptance of the model in applied economics.

2.6 Model flexibility I

The model should be able to account for changes in the shape of the distribution through changes in parameter values. The fulfilment of this property contributes to support the claim of the universality of a given model's ability to describe income distributions of different countries, regions, socioeconomic groups and in different periods of time, as well as from different sources of income (wages and salaries, proprietor's income and property income).

2.7 Model flexibility II

The model should be able to account for negative and nil income through changes in the values of some of its parameters. The existence of negative and nil income strongly restricts the descriptive power of almost all specified models. Among the exceptions are the three- and four-parameter lognormal and the log-logistic models. Several models, such as the Gamma function, are unable to deal with negative incomes without a *tour de force* that sacrifices the descriptive accuracy of the actual distribution. A case in point is Salem and Mount's (1974) approach in fitting the Gamma function to the income distribution of the USA in 1970. In Table I, noted (p. 1122), they state: "The first two groups are combined in 1970 to overcome the problem of having a negative mean income in the first group". Afterward, when they try to explain the poor estimate of the concentration ratio for 1970, while using the best method of estimation, they say (p. 1122): "One reason for the low estimate using Method B is that the arithmetic mean of the first income group is - $201. The geometric mean, however, is estimated using $500 for the first group (...). In fact, the geometric mean cannot be defined with negative incomes in the sample, and, consequently, the two gamma parameters should be estimated from a sample truncated at zero, and families with negative incomes could be considered separately". The mathematical contradiction of this "pragmatic" approach becomes evident when we observe that the arithmetic mean is greater than the geometric mean, when both are defined and the random variable is not constant. Hence, the problem calls for a respecification of the Gamma model, as, for example, the generalized Gamma model specified by L.A. Amoroso (1924-1925), or the specification of a new model that can deal with negative incomes, while retaining its descriptive power for all income groups.

2.8 Model flexibility III

The model should be able to deal with a positive, and not predetermined, minimum income without truncating the distribution, i.e. to describe the actual distribution with income range equal to the closed-open infinite interval $[x_0, \infty]$ where $x_0 > 0$ and generally unknown. This is the case of regions with a minimum guaranteed income. This also could refer to the case when the population considered is the employed labour force and not the household population. This approach is found when the actual income distribution is integrated with the system of national accounts and is obtained as the income generated by the productive process of the economy (see Conade-Cepal, 1965, t. I, p. 7).

2.9 Model flexibility to deal with both unimodal and strictly decreasing (non-modal) income distributions

This is an important practical property that allows us to retain the model to describe non-modal sub-populations of income units.

2.10 Minimization of biased assumptions when dealing with the method of parameter estimation

One important source of bias is the assumption of equi-distribution within each interval of income required in the fitting of density functions. This source of bias is overcome by the cumulative distribution function approach corresponding to the model specified by Pareto (1895), Dagum (1975) and Singh and Maddala (1976).

2.11 Derivation of the explicit mathematical form of the Lorenz curve from the model of income distribution, and conversely

Several models do not have the Lorenz curve explicit solution. This is possible because either the solution does not exist or it is not yet known. Hence, when it comes to the use of the Lorenz curve associated with a given income distribution function, the work is done numerically and not analytically. This dichotomy is due to the mathematical limitations or shortcomings of the specified model.

2.12 Explicit mathematical solution for the Gini concentration ratio

"Gini concentration ratio is perhaps the most useful – and certainly the most widely used – measure of changes in inequality" (Budd, 1970, p. 247). The existence of a mathematical solution for the Gini ratio allows its direct computation and the verification of whether or not it lies within the interval derived by Gastwirth (1972) and proposed as a goodness of fit test.

2.13 The Gini concentration ratio associated with the model of income distribution should be able to account for intersecting Lorenz curves

Surprisingly enough, the literature on income distribution had agreed that the Gini ratio was unable to account for intersecting Lorenz curves. In fact, soon after Budd (1970, p. 247) states that the Gini concentration ratio is perhaps the most useful measure of changes in inequality, he adds "it does, as we know, produce an ambiguous measure of changes in inequality". Actually, the ambiguous measure of changes in inequality is a consequence of the mathematical limitations of the income distribution model being identified rather than a limitation of the Gini measure (see Dagum, 1977b). Whenever the specified model contains a single inequality parameter, as the lognormal and the Gamma, the Gini ratio will be a monotonic function of it, hence it will not be able to detect intersecting Lorenz curves. Amoroso (1924-1925, pp. 137-138), in his classical contribution, missed this important characteristic when he deduced the same value of the Gini ratio for different numerical combinations of his generalized Gamma function. Instead of making the right interpretation stating the existence of the intersecting Lorenz curves, he dismissed the Gini ratio as useless![1]

2.14 *Stochastic convergence of the model to the Pareto distribution for high levels of income*

That is, the model should fulfil the weak form of the Pareto law (Mandelbrot, 1960, p. 81). This important stochastic property is strengthened by the undisputed acceptability of the Pareto distribution as the model of high income groups. If the specified model possesses this important stochastic property, then it will demand a further investigation to determine if it does, or does not belong to the class of stable Pareto-Lévy distributions (Mandelbrot, 1960; Lévy, 1925, ch. VI).

3 Model Specification

The characteristics of regularity and permanence of the actual or empirical income distributions in both developed and developing countries can be represented by the following set of assumptions:

A.1: Empirical income distributions are, in general, unimodal and positively skewed.

A.2: There exists a finite percentage of economic units with nil or negative income. If the economic units are composed of unattached individuals, this percentage corresponds to those unemployed unattached individuals without any source of income (social insurance, etc.) and proprietors with net losses. If the economic units are families with two or more members, they receive the same interpretation as unattached individuals, however this is the income of all members of the family units.

A.3: The income range is the closed-open interval $[x_0, \infty]$, where $x_0 > 0$, when the population of economic units is integrated with the employed members of the labour force. That is, the income distribution starts from the right of the origin, since it is composed of economic units with positive income. This assumption corresponds to property 2.8 introduced in the preceding section and is related to the generation of the income distribution statistics of the productive process of the economy from an integrated system of national accounts. It is also the case of the population of

[1] "E ciò chiarisce ancora una volta come il rappoto di concentrazione non serva a misurare la concentrazione dei redditieri" (Amoroso, 1924-1925, p. 138).

economic units obtained after elimination of its members with zero or negative income. That is the approach followed by Figueroa (1974, p. 83).

A.4: The income elasticity of the cumulative distribution function $F(x)$, with respect to the origin α of $F(x)$, $\alpha < 1$, is a monotonic decreasing function of $F(x)$. This elasticity converges to a finite and positive value $\beta\delta$ when income x tends to zero; and it converges to zero when income x tends to infinity. That is, for a given constant proportional rate of growth of income, there corresponds a decreasing proportional rate of growth of the cumulative distribution function $F(x)$, which depends of the size of $F(x)$ itself.

The mathematical representation of the set of assumptions A.1-A.4 is as follows:[2]

$$\frac{d\log[F(x)-\alpha]}{d\log x} = \beta\delta\left[1-\left(\frac{F-\alpha}{1-\alpha}\right)^{1/\beta}\right], \tag{1.1}$$

$x > 0$ if $0 \leq \alpha < 1$ and $x > x_0 > 0$, where $F(x_0) = 0$, if $\alpha < 0$, subject to:

$$\beta > 0, \alpha < 1, \text{ and } \beta\delta > 1. \tag{1.2}$$

The solution of (1.1) is:

$$F(x) = \alpha + \frac{1-\alpha}{(1+\lambda x^{-\delta})^{\beta}}, \quad \lambda > 0, \tag{1.3}$$

where λ is strictly positive because it is the antilog of the constant of integration.

It can be verified that:

i) A.1 implies $\beta\delta > 1$; that is, the distribution is unimodal. If $0 < \beta\delta \leq 1$, then the distribution is non-modal. This situation can occur when the model describes that part of the actual distribution to the right of the model (the Pareto case) or the case in which the actual distribution is non-modal, as could be the case of a poor and overpopulated country.

ii) A.2 implies $0 \leq \alpha < 1$, where α can receive the interpretation of a pure rate of unemployment for the economic units considered, in the sense that its dominant part is constituted by those unemployed economic units without social insurance. Hence, $x_0 = 0$.

iii) A.3 implies $\alpha < 0$. Hence, x_0 is the solution of $F(x_0) = 0$. Therefore, the model is defined for all $x \geq x_0 > 0$.

iv) A.4 implies the differential equation (1.1).

Whenever $0 < \alpha < 1$ and given that the cumulative distribution function (c.d.f) (1.3) is monotonically increasing, the classical decomposition holds, i.e.:

$$F(x) = \alpha F_1(x) + (1-\alpha)F_2(x) \tag{1.4}$$

[2] For mathematical convenience, as can be seen in (1.3), the right hand coefficient in (1.1) is written as the product of two parameters.

where $F_1(0) = 1$, $F_1(x) = 0$, for all $x \neq 0$; $F_2(x) = 0$ for all $x \leq 0$ and $F_2(x)$ is a continuous and differentiable function for all $x > 0$. Its corresponding density function is:

$$f(x) = \begin{cases} \alpha, & \text{when } x = 0 \\ (1-\alpha)\beta\lambda\delta x^{-\delta-1}(1+\lambda x^{-\delta})^{-\beta-1}, & \text{for all } x > 0, \text{ and} \\ 0, & \text{for all } x < 0 \end{cases} \tag{1.5}$$

Unlike the lognormal and Gamma functions, the present model has an explicit mathematical solution for its c.d.f, which is given in (1.3). It also yields an explicit mathematical solution for both the mode of the density function (the point of inflexion of the c.d.f) when $\beta\delta > 1$, and the p-th percentile for all $p > \alpha$, hence the median, i.e. $p = 0.50$. It can be shown that the mode x_M is:

$$x_M = \lambda^{1/\delta}\left(\frac{\beta\delta-1}{\delta+1}\right)^{1/\delta}, \ \beta\delta > 1, \tag{1.6}$$

the median x_m:

$$x_m = \begin{cases} \lambda^{1/\delta}\left[\left(\frac{1-\alpha}{0.5-\alpha}\right)^{1/\beta} - 1\right]^{-1/\delta}, & \alpha < 0.5 \\ 0, & 0.5 \leq \alpha < 1 \end{cases} \tag{1.7}$$

and, in general, the p-th percentile x_p:

$$x_p = \lambda^{1/\delta}\left[\left(\frac{1-\alpha}{p-\alpha}\right)^{1/\beta} - 1\right]^{-1/\delta}, \quad p > \alpha. \tag{1.7'}$$

Finally, when $\alpha < 0$, the root of the equation $F(x) = 0$ is:

$$x_0 = \lambda^{1/\delta}\left[\left(\frac{\alpha-1}{\alpha}\right)^{1/\beta} - 1\right]^{-1/\delta} \tag{1.8}$$

It can be shown (Dagum, 1977a) that the r-th moment about the origin is, for all positive integers $r < \delta$.

$$E(X^r) = (1-\alpha)\beta\lambda^{r/\delta}B\left(1-\frac{r}{\delta}, \ \beta+\frac{r}{\delta}\right), \quad \text{when } 0 \leq \alpha < 1, \tag{1.9}$$

and

$$E(X^r) = (1-\alpha)\beta\lambda^{r/\delta}B\left(\frac{\lambda}{\lambda+x_0^\delta}; 1-\frac{r}{\delta}, \beta+\frac{r}{\delta}\right), \quad \text{when } \alpha < 0, \tag{1.10}$$

where $B\left(t_0;\ 1\text{-}\frac{r}{\delta}, \beta+\frac{r}{\delta}\right)$ is the incomplete Beta function defined as:

$$B\left(t_0;1-\frac{r}{\delta},\beta+\frac{r}{\delta}\right)=\int_0^{t_0} t^{-r/\delta}(1-t)^{\beta-1+r/\delta}dt, \quad 0\leq t_0<1. \tag{1.11}$$

The ratio between the incomplete and the complete Beta function is tabulated in Pearson (1934) for several combinations of its parameters.

From (1.9), when $0\leq\alpha<1$, and from (1.10), when $\alpha<0$, we can deduce the mean income and its variance, and it can also be shown that the coefficients of skewness and excess are functions of α,β and δ but not of λ. This conclusion is coherent with the economic interpretation of the parameters of the present model; that is with the set of assumptions A.1-A.4 introduced above and with its corresponding mathematical solution (1.3). In other words:

i) λ is a scale parameter. Moreover, $\lambda^{-1/\delta}$ has the same dimension as income x. Changes in the monetary unit will, *ceteris paribus*, change the parameter λ and will leave invariant the remaining three parameters;

ii) α,β and δ are dimensionless parameters. α is an inequality parameter and β and δ can be called equality parameters because the Gini ratio is an increasing function of the former and a decreasing function of the latter.

It can be shown (Dagum, 1977a) that the Lorenz curve associated with the model (1.3) is:

$$L(y)=\frac{\mathrm{B}\left(y^{1/\beta};\beta+\frac{1}{\delta},1-\frac{1}{\delta}\right)}{\mathrm{B}\left(\beta+\frac{1}{\delta},1-\frac{1}{\delta}\right)}, \quad \beta\delta>1, \quad 0\leq\alpha<1, \tag{1.12}$$

where

$$y=\frac{F(x)-\alpha}{1-\alpha}, \quad y\in[0,1]. \tag{1.13}$$

and its corresponding Gini concentration ratio is:

$$G=(2\alpha-1)+(1-\alpha)\frac{\Gamma(\beta)\Gamma\left(2\beta+\frac{1}{\delta}\right)}{\Gamma(2\beta)\Gamma\left(\beta+\frac{1}{\delta}\right)} \tag{1.14}$$

where $\Gamma(\cdot)$ is the complete Gamma function.

The economic interpretation of the parameters α,β and δ, that enter into the measure of income inequality, can also be derived from the Gini ratio (1.14), given that:

$$\frac{\partial G}{\partial\alpha}>0,\ \frac{\partial G}{\partial\beta}<0,\ \text{and}\ \frac{\partial G}{\partial\delta}<0,\ \beta>0\ \text{and}\ \beta\delta>1. \tag{1.15}$$

For $\beta=1$ it can be shown (Dagum, 1977b) that:

$$L(y)=\frac{\delta}{\Pi}\sin\frac{\Pi}{\delta}\mathrm{B}\left(y;1+\frac{1}{\delta},1-\frac{1}{\delta}\right), \quad \delta>1 \tag{1.16}$$

and

$$G=\alpha+\frac{1-\alpha}{\delta}, \quad 0\leq\alpha<1\ \text{and}\ \delta>1 \tag{1.17}$$

which can also be deduced from (1.12) and (1.14).

The mathematical forms of the Lorenz curve and the Gini ratio, when $\alpha < 0$, are deduced in Dagum (1977c). The economic interpretation of the parameters α, λ and δ, when $\beta = 1$, are the same as in the four-parameter model.

4 Intersecting Lorenz Curves

This topic was already introduced as property 2.13 to guide the identification of a descriptive model of income distribution. The Gini concentration ratio (1.14) deduced from the c.d.f. (1.3) can account for the existence of intersecting Lorenz curves, either with equal or different Gini ratios. This is the case where both tails (the upper and lower income groups) lose relative to the middle of the distribution.

Given two actual income distributions fitted with model (1.3), their corresponding Gini ratios are, applying (1.14):

$$G_1 = G(\alpha_1,\ \beta_1,\ \delta_1) \text{ and } G_2 = G(\alpha_2,\ \beta_2,\ \delta_2) \tag{1.18}$$

and the Lorenz curves are, applying (1.12) and (1.13):

$$L_1 = L(y;\ \alpha_1,\ \beta_1,\ \delta_1) \text{ and } L_2 = L(y;\ \alpha_2,\ \beta_2,\ \delta_2) \tag{1.19}$$

Since $\frac{d^2L}{dy^2} > 0$ for all $y \in (0,1)$ which implies strict convexity, the Lorenz curve deduced in (1.12) is well-behaved. The zeros of the equation $L_1 = L_2, y \in (0,1)$, will locate the point or points of intersection. In particular, if $G_1 = G_2, \beta_1 = \beta_2$, $0 < \alpha_1 < \alpha_2$ and $\delta_1 > \delta_2$, the Lorenz curve $L_1(y;.)$ does not intersect $L_2(y;.)$ and is completely to the left of L_2 (Fig. 1.1); but, if $\delta_1 = \delta_2$, then L_1 intersects L_2 from above (Fig. 1.2).

The three-parameter model specified in Dagum (1975) can also account for intersecting Lorenz curves, as can be deduced from (1.16) and (1.17).

5 Convergence to the Pareto Distribution

Property 2.14 motivates the introduction of this section. In a stimulating contribution, Mandelbrot (1960) developed the concepts of the strong and weak Pareto laws.

Let $P(x) = P(X > x | 0 < x^* \leq x)$ be the cumulative frequency of income earner units with income X greater than a positive income x (x is assumed to be a continuous variable). The strong Pareto law states that:

$$P(x) = \begin{cases} (x/x^*)^{-\theta} & x \geq x^*, \\ 1 & x < x^*. \end{cases} \tag{1.20}$$

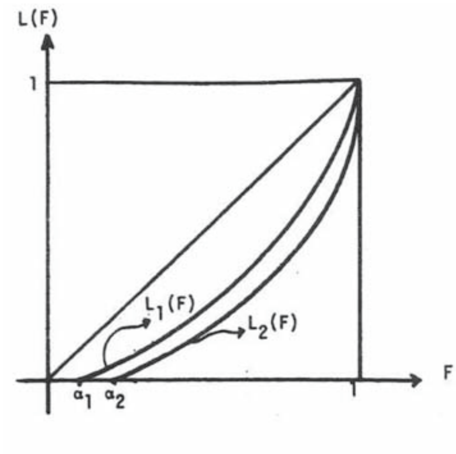


Fig. 1.1: $\alpha_1 < \alpha_2, \beta_1 = \beta_2$ and $\delta_1 > \delta_2$.

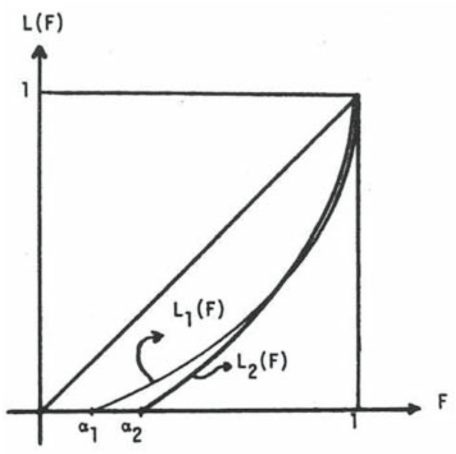


Fig. 1.2: $\alpha_1 < \alpha_2, \beta_1 = \beta_2$ and $\delta_1 < \delta_2$.

The weak Pareto law (Mandelbrot, 1960, p. 80) states that "$P(x)$ behaves like $(x/x^*)^{-\theta}$, as $x \to \infty$". Symbolically,

$$P(x) \sim (x/x^*)^{-\theta}. \tag{1.21}$$

Therefore,

$$\frac{P(x)}{(x/x^*)^{-\theta}} \to 1, \text{ as } x \to \infty \tag{1.22}$$

or

$$P(x) = (x/x^*)^{-\theta} + o(x/x^*)^{-\theta}, \text{ as } x \to \infty, \tag{1.23}$$

where the little o stands for *smaller order of magnitude*.

The model specified in (1.3) possesses this important property of the weak Pareto law, that is, (1.3) converges in distribution to the Pareto model (1.20) for x sufficiently large. In fact, from (1.3) and the definition of $P(x)$ we deduce:

$$P(x) = 1 - F(x) = (1-\alpha)\sum_{r=1}^{\infty}(-1)^{r-1}\binom{\beta+r-1}{r}\lambda^r x^{-r\delta}, \tag{1.24}$$

$$\lambda(\beta+r) < (r+1)x^{\delta},$$

where $(-1)^r\binom{-\beta}{r} = \binom{\beta+r-1}{r}$ is the negative binomial coefficient, and $\lambda(\beta+r) < (r+1)x^{\delta}$ is the condition of convergence of the series expansion in (1.24).

For sufficiently large values of x and performing the substitution

$$(1-\alpha)\beta\lambda = x^{*\delta} \tag{1.25}$$

we obtain (1.21). That is, model (1.3) converges in the weak sense to the Pareto distribution. This convergence for Canada and the USA is sufficiently close (with an absolute error less than $\varepsilon = 0.01$) for incomes as low as twice the mean income.

This important convergence property of model (1.3) is enhanced by its power to describe with similar accuracy the remaining part of actual income distributions, that is the lower and the middle range of income. This statement is substantiated in the applications (section 7), where model (1.3) is fitted to actual income distributions of developed as well as developing countries.

6 Methods of Parameter Estimation

All models of income distribution besides being non-linear belong to the class of transcendent functions. The Pareto model is the only one that can be linearized after performing a log transformation. Hence, the least squares (LS) and the maximum likelihood (ML) methods of parameter estimation require the solution of a system of non-linear (transcendent) equations. Although this is not a handicap today for most specified models, simpler methods of estimation have been developed. Herewith we propose five methods to estimate the parameters of model (1.3). They are: *i*) iterative method I; *ii)* iterative method II; *iii)* iterative method III; *iv)* unconstrained function minimization; and *v*) the method of maximum likelihood.

i) Iterative method I: After appropriate transformation of (1.3), it can be shown that:

$$\log\left[\left(\frac{1-\alpha}{F(x)-\alpha}\right)^{1/\beta}-1\right]=\log\lambda-\delta\log x \tag{1.26}$$

and

$$\log F(x)=-\beta\log(1+\lambda x^{-\delta})+\sum_{k=1}^{\infty}\alpha^k\frac{1-F^k}{kF^k}. \tag{1.27}$$

(1.27) can be approximated by the following linear form in α and β:

$$\log F(x)\doteq\alpha\frac{1-F(x)}{F(x)}-\beta\log(1+\lambda x^{-\delta}). \tag{1.28}$$

To estimate the parameter vector $(\alpha,\beta,\lambda,\delta)$ in (1.3), start with an initial value of α and β and estimate λ and δ by LS using (1.26). These estimates of λ and δ enter in (1.28) and then estimate α and β. These new estimates of α and β enter in (1.26) and then obtain a second estimate of λ and δ. The iteration continues until convergence is attained.

ii) Iterative method II: This method avoids the approximation (1.28) using three steps to complete the estimate of the four parameters. The iteration starts with an initial value of α and β and estimates λ and δ using (1.26). Given these estimates of λ and δ, and an initial value of β estimate α using transformation:

$$F(x)(1+\lambda x^{-\delta})^{\beta}-1=\alpha[(1+\lambda x^{-\delta})^{\beta}-1] \tag{1.29}$$

Finally, with the estimates of α, λ and δ as initial values, estimate β using the transformation:

$$\log\frac{1-\alpha}{F(x)-\alpha}=\beta\log(1+\lambda x^{-\delta}) \tag{1.30}$$

The estimated values of α and β are now used to obtain a second estimate of λ and δ, and the iteration continues until convergence is reached.

iii) Iterative method III: This is a simple and efficient non-linear method of estimation. It converges to the least squares estimate of the model.

Let α, β, λ, δ be the least squares estimates of (1.3), and let:

$$\Psi(\alpha,\beta,\lambda,\delta)=\sum[F(x)-\alpha-(1-\alpha)(1+\lambda x^{-\delta})^{-\beta}]^2 \tag{1.31}$$

be the sum of the squares of the deviations of the actual from the fitted values. Let $(\alpha_0,\beta_0,\lambda_0,\delta_0)$ be approximations to $(\alpha,\beta,\lambda,\delta)$ taken as initial estimates. The Taylor series expansion of $\Psi(\alpha,\beta,\lambda,\delta)$ about the origin $(\alpha_0,\beta_0,\lambda_0,\delta_0)$, using vector notations and symbolic operators, is

$$\Psi(\alpha,\beta,\lambda,\delta)=\sum_{r=0}^{\infty}\left[(\mu-\mu_0)'\frac{d}{d\mu}\right]^r\Psi(\mu_0) \tag{1.32}$$

where

$$(\mu-\mu_0)'=(\alpha-\alpha_0,\beta-\beta_0,\lambda-\lambda_0,\delta-\delta_0) \tag{1.33}$$

is a row vector of deviations of the least squares estimates from their approximated values, and

$$\frac{d}{d\mu} = \left(\frac{\partial}{\partial\alpha}\ \frac{\partial}{\partial\beta}\ \frac{\partial}{\partial\lambda}\ \frac{\partial}{\partial\delta}\right)', \tag{1.34}$$

symbolically multiplied by $\Psi(\mu_0) = \Psi(\alpha_0, \beta_0, \lambda_0, \delta_0)$, is the transpose of the column vector of first order derivatives evaluated at μ_0. For brevity the vector of partial derivatives evaluated at μ_0 may be written

$$\frac{d\Psi(\mu_0)}{d\mu} = (\Psi_\alpha(\mu_0)\ \Psi_\beta(\mu_0)\ \Psi_\lambda(\mu_0)\ \Psi_\delta(\mu_0))'. \tag{1.35}$$

To minimize $\Psi(\mu)$, differentiate with respect to the vector μ and equate to zero, thus obtaining a system of four linear equations. These may be expressed in matrix form as follows, after neglecting the terms with third and higher order derivatives of $\Psi(\mu)$:

$$\begin{bmatrix} \Psi_{\alpha\alpha}(\mu_0) & \Psi_{\alpha\beta}(\mu_0) & \Psi_{\alpha\lambda}(\mu_0) & \Psi_{\alpha\delta}(\mu_0) \\ & \Psi_{\beta\beta}(\mu_0) & \Psi_{\beta\lambda}(\mu_0) & \Psi_{\beta\delta}(\mu_0) \\ & & \Psi_{\lambda\lambda}(\mu_0) & \Psi_{\lambda\delta}(\mu_0) \\ & & & \Psi_{\delta\delta}(\mu_0) \end{bmatrix} \begin{bmatrix} \alpha - \alpha_0 \\ \beta - \beta_0 \\ \lambda - \lambda_0 \\ \delta - \delta_0 \end{bmatrix} = - \begin{bmatrix} \Psi_\alpha(\mu_0) \\ \Psi_\beta(\mu_0) \\ \Psi_\lambda(\mu_0) \\ \Psi_\delta(\mu_0) \end{bmatrix} \tag{1.36}$$

or, more compactly, as

$$H(\mu - \mu_0) = -A. \tag{1.37}$$

where H is the symmetric matrix of second order derivatives of $\Psi(\mu)$ evaluated at μ_0.

Hence, the approximate solution of μ given the initial vector μ_0 is:

$$\mu = \mu_0 - H^{-1}A \tag{1.38}$$

The process is repeated with the solution μ in (1.38) taken as μ_0 until convergence is attained.

iv) Unconstrained function minimization: This method estimates the parameter vector $\mu = (\alpha, \beta, \lambda, \delta)$ by iterative algorithm (Birta, 1976) that searches the minimization of the sum of the squares of the deviations of the actual from the fitted values, i.e. the minimization of $\Psi(\alpha, \beta, \lambda, \delta)$.

v) The method of maximum likelihood (ML): Let $F(x)$ in (1.3) be an unbiased predictor (Wold, 1961, 1963) of the sample realization $\eta(x)$.

Hence,

$$\eta(x) = F(x) + \varepsilon \tag{1.39}$$

where ε is a purely random variable. Assuming that ε is normally distributed[3], i.e.:

$$\varepsilon \stackrel{d}{=} N(0, \sigma^2) \tag{1.40}$$

[3] The symbol $\stackrel{d}{=}$ means “equal in distribution”.

the log of the likelihood equation is

$$\log L = -\tfrac{n}{2}\log 2\pi - \tfrac{n}{2}\log\sigma^2 - \tfrac{1}{2\sigma^2} \sum\left(\eta(x) - \alpha - (1-\alpha)(1+\lambda x^{-\delta})^{-\beta}\right)^2 \quad (1.41)$$

The ML estimator $\hat{\mu}$ of μ is obtained by solving the likelihood equation $\frac{d\log L}{d\mu} = 0$, where $\mu = (\alpha, \beta, \lambda, \delta)$.

The asymptotic variance-covariance matrix is deduced from the Cramér-Rao inequality (Cramér, 1946; Rao, 1973), which gives the lower bound for the estimators' variances. Denoting by $I(\mu)$ the information matrix, we obtain:

$$I(\mu) = -E\left(\frac{d^2\log L}{d\mu^2}\right) \quad (1.42)$$

where the partial derivatives are evaluated at $\hat{\mu}$. The inverse of (1.42) is the asymptotic covariance matrix of the estimators.

Under general conditions of stochastic regularity, the likelihood equation $\frac{d\log L}{d\mu} = 0$ has a solution $\hat{\mu}$ which converges in probability to the population vector μ, as the sample size tends to infinity. This solution is an asymptotically normal and asymptotically efficient estimate of μ. That is:

$$\hat{\mu} \xrightarrow{d} N(\mu, I^{-1}(\mu)), \quad (1.43)$$

hence,

$$(\hat{\mu} - \mu)' I(\mu)(\hat{\mu} - \mu) \to \chi^2(4) \quad (1.44)$$

and from the unbiased predictor assumption, (1.40), (1.44) and the Cochran theorem, the following is deduced:

$$\frac{(\hat{\mu} - \mu)' I(\mu)(\hat{\mu} - \mu)}{4\hat{\sigma}^2/\sigma^2} = F(4, n-4). \quad (1.45)$$

The F statistic in (1.45) is only a function of the unknown vector μ, for the population variance σ^2 vanishes, given that $I(\mu)$ contains the same factor in the denominator. Given a level of significance, (1.45) may be employed to build confidence regions for the vector μ or a subset of its components. The most important subset in the one formed with the parameters of dimension zero (α, β, δ) which determine the Gini concentration ratio.

7 Empirical Results. Four Country Study: Argentina, Canada, Sri Lanka and the USA

The parameters are estimated in all cases by minimizing the sum of squared residuals (1.31) using the search procedure of method *iv*.

Table 1.1 shows the estimated parameters of model (1.3) corresponding to the distribution of family income in Canada, in 1973 (Statistics Canada, 1975), Argentina, in 1959 (Conade-Cepal, 1965), Sri Lanka in 1973 (Central Bank of Ceylon, 1974) and the USA in 1960, 1965, 1969 and 1970 (US Bureau of the Census, 1961-1971). The sum of squared deviations of the observed frequencies of income earners by interval of income from their corresponding predicted probabilities are denoted by

$$\sum u^2 = \sum (f(x) - \hat{f}(x))^2 \tag{1.46}$$

and, for the cumulative distribution function,

$$\sum \varepsilon^2 = \sum (F(x) - \hat{F}(x))^2. \tag{1.47}$$

The sums of squared deviations reported in Table 1.1 show an exceptional goodness of fit for all cases studied. The last column in Table 1.1 reports the Gini concentration ratio applying the formula deduced from model (1.3) and presented in (1.14) for all but Argentina (1959) and Sri Lanka (1973), for which the estimated α's are negative. In these latter cases the negative values of α were taken into consideration to estimate the Gini ratio.

Table 1.1:

Country and Year[1]	*Estimated parameters* α	β	λ	δ	$\sum u^2$	No. of inter vals[2]	Gini ratio
Argentina 1959	-0.0033	20.307	2.24	2.1622	0.00097	20	0.380
Canada 1973	0.0028	0.427	84643.0	4.0739	0.00022	29	0.327
USA 1960	0.0209	0.381	6513.9	4.1095	0.00066	9	0.352
USA 1965	0.0006	0.343	20789.6	4.1923	0.00027	16	0.348
USA 1969	0.0000	0.348	117785.4	4.3734	0.00013	9	0.335
USA 1970	0.0000	0.374	55847.8	4.0435	0.00007	10	0.346
USA 1970	0.0000	0.369	65147.2	4.0969	0.00011	16	0.347
Sri Lanka 1973	-0.0330	0.409	191.68	3.0372	0.00133	26	0.410

[1] The scale parameter corresponds to income measured in 10^{-4} of the Argentinian pesos, in 10^{-2} of Sri Lanka rupees, in 10^{-3} of the Canadian dollars and in 10^{-3} of the USA dollars.
[2] Excluding the open-ended interval.

Table 1.2 presents the bounds for the Gini concentration ratio estimated by Salem and Mount (1974, p. 1122) for the USA in 1969 and 1970. These bounds were derived by Gastwirth (1972) and proposed as a goodness of fit test. It is indeed a very demanding test. Table 1.2 also reports the Gini ratio estimated by: *i*) Salem and Mount (1974) from the Gamma function using their best method of estimation (method B); *ii)* Singh and Maddala (1976) using the model specified by these authors; and *iii)* the model (1.3) in this paper applying formula (1.14). The estimates of the Gini ratio in 1970 (with 17 intervals) fall within the bounds for the Singh-Maddala (S-M) model and the present model (1.3).

Table 1.2: Estimated Values of and Bounds for the Gini Concentration Ratio USA Family Income

Year	No. of intervals	*Bounds*		*Gini Concentration Ratio*		
		Lower	Upper	Gamma	S-M	Present
1969	10	0.326	0.356	0.355		0.335
1970	10	0.337	0.372	0.352		0.346
1970	17	0.347	0.356	0.325	0.348	0.347

Table 1.3: Sum of Squared Deviations

Model	*Year*	
	1960	1969
Lognormal	0.01187	0.00752
Gamma	0.00391	0.00238
Singh-Maddala	0.00261	0.00156
Present model	0.00066	0.00013

Table 1.4: Observed and Estimated Values of the Median and Mean (in current values)

	Median Income			*Mean Income*		
Country and Year	Observed	Estimated from (3.7)	% of dif.	Observed	Estimated from (3.9) and (3.10)	% of dif.
Argentina 1959	69,250	68,830	0.61	112,821		
Canada 1973	11,533	11,455	0.68	12,716	12,757	0.32
Sri Lanka 1973	360	356	1.11	455	449	1.27
USA 1960	5,620	5,592	0.50	6,227	6,217	0.16
USA 1965	6,957	6,840	1.68	7,704	7,642	0.81
USA 1969	9,433	9,466	0.16	10,577	10,433	1.36
USA 1970 (10 intervals)	9,867	9,846	0.18	11,106	11,068	0.35
USA 1970 (16 intervals)	9,867	9,853	0.03	11,106	11,038	0.61

Table 1.3 reports the sum of squared deviations of the observed frequencies from their corresponding predicted probabilities, for the lognormal, Gamma, Singh-Maddala and the present model in years 1960 and 1969.

Finally, Table 1.4 shows the median and mean incomes as reported by the official statistics of each country (observed values) and estimated from the present model using formula (1.7) for the median and (1.9) or (1.10) for the mean income – depending on whether α is nonnegative or negative – after setting r=1. The estimated values of the parameters entering (1.7), (1.9) and (1.10) are reported in Table 1.1.

8 Conclusion

The presently large quantity of competing models specified to describe actual income distributions calls for the statement of a set of properties to guide the identification of the model to be adopted. This is done in section 2, by the proposition of a set of fourteen properties.

Section 3 derives a four-parameter model of income distribution from an elementary set of logico-empirical assumptions, hence it satisfies property 2.1. Although the number of unknown parameters is reasonable it is not as parsimonious as the two-parameter models. Other four-parameter models are Champernowne (1952) and Fisk (1961). The model specified by Amoroso (1924-1925) is five-parameter and those of Singh-Maddala, the displaced lognormal and the log-logistic (Dagum, 1975) are three-parameter models. Property 2.3 is fully accomplished by the present model since there is clear-cut economic interpretation of its four parameters.

The exceptional goodness of fit reported in section 7 for four very dissimilar countries (Argentina, Canada, Sri Lanka and the USA) stresses the fulfillment of property 2.4. Earlier, Singh and Maddala found that their model gives a better fit than the lognormal and the gamma. Table 1.3 shows that the present model gives a better fit than the Singh-Maddala. Moreover, it fits remarkably well over the whole income range, a possibility that was dismissed repeatedly in the literature by several researchers. Among them, Mandelbrot (1960, p. 82) stated, "The above reasons make it unlikely that (...) a single empirical formula could ever represent all the data"; Fisk (1961, p. 178) also stated that the "search for a single simple distribution function to describe the total distribution of incomes may prove fruitless"; and Budd (1970, p. 250) reiterated the past literature when he stated, "we know that it is virtually impossible to describe empirical distributions accurately by just one function".

Section 6 deals with five methods of parameter estimation. One of them provides the maximum likelihood estimators. The first three are simple iterative methods as required by property 2.5, even though they are not as simple as those methods of estimation available for the two-parameter models.

Properties 2.6 to 2.9 require flexibility of the specified model to account for: *i)* changes in the distribution; *ii)* nil and negative income; *iii)* income range starting from a positive, and not pre-determined, minimum income; and *iv)* unimodal and strictly decreasing density functions. The proposed model can account very well for these four properties as can be verified by the content of its set of logico-empirical assumptions in section 3 and its empirical results in section 7.

The possibility of working with the cumulative distribution function eliminates the source of bias which is present when using the density function, since one has to face the problem of selecting the representative point of each interval. For the c.d.f. the point is well determined, viz. it is the right end of each interval, and this also overcomes the problem created by the possible existence of negative mean income in the first interval. Therefore, property 2.10 is also fulfilled.

Properties 2.11 to 2.13 are also fulfilled, since section 3 presented the deduced Lorenz curve and Gini concentration ratio. Moreover, the parameters that enter in the latter can account for intersecting Lorenz curves.

Finally, section 5 proves the convergence of the present model to the Pareto distribution for high levels of income, hence it fulfils the weak form of the Pareto law.

References

Aitchison, J. and J. A. C. Brown (1957) *The Lognormal Distribution with Special Reference to its Uses in Enonomics*, Cambridge University Press, Cambridge.

Ammon, O. (1895) *Die Gesellschaftsordnung und ihre Natrlichen Grundlagen*, Jena.

Amoroso, L. (1924-1925) Ricerche Intorno alla Curva dei Redditi, *Annali di Mathematica Pura ed Applicata, Serie 4-21*, **II**, 123–157.

Bartels, C. P. A. and H. van Metelen (1975) Alternative Probability Density Functions of Income, Vrije University Amsterdam: Research Memorandum No. 29, 30 pages.

Birta, L. G. (1976) OPTPAK a Program Package for Unconstrained Function Minimization, Technical Report TR 76-02, University Ottawa, 22 pages.

Budd, E. C. (1970) Distribution Issues: Trends and Policies. Postwar Changes in the Size Distribution of Income in the U.S., *A. E. R.*, **60**(2), 247–260.

Central Bank of Ceylon (1974) Survey of Sri Lanka's Consumer Finances 1973, Colombo, Department of Economic Research, Central Bank of Ceylon.

Champernowne, D. G. (1952) The Graduation of Income Distributions, *Econometrica*, **20**(4), 591–615.

Champernowne, D. G. (1953) A Model of Income Distribution, *Economic Journal*, **63**(2), 318–351.

Conade-Cepal (1965) *Distribucion del Ingreso y Cuentas Nacionales en la Argentina*, tomos I-V, Buenos Aires, Presidencia de la Nacion Argentina.

Cramér, H. (1946) *Mathematical Methods of Statistics*, Princeton University Press, Princeton.

D'Addario, R. (1949) Ricerche sulla Curva dei Redditi, *Giornale degli Economisti e Annali di Economia*, **8**, 91–114.

Dagum, C. (1973) Un Modèle non Linéaire de Répartition Fonctionnelle du Revenu, *Economie appliquée*, **XXVI**(2-3-4), 843–876.

Dagum, C. (1975) A Model of Income Distribution and the Conditions of Existence of Moments of Finite Order, Proceedings of the 40th session of the International Statistical Institute, vol. XLVI, Book 3, Warsaw, pp. 196–202.

Dagum, C. (1977a) A New Model of Income Distribution, the Lorenz Curve and the Gini Concentration Ratio, forthcomming paper.

Dagum, C. (1977b) Analyse Régionale de la répartition du Revenu au Canada, *Economie appliquée*, **XXX**(2).

Dagum, C. (1977c) El Modelo Log-Logistico Y La Distribucion Del Ingreso En Argentina, *El Trimestre Economico*, **XLIV (4)**(176), 837–864.

Figueroa, A. (ed.) (1974) *Estructura del Consumo y Distribution de Ingresos en Lima Metropolitana, 1968-1969*, Serie Estudios Economicos No. 1, Lima, Pontificia Universidad Catolica del Peru.

Fisk, P. R. (1961) The Graduation of Income Distributions, *Econometrica*, **29**(2), 171–185.

Gastwirth, J. L. (1972) The Estimation of the Lorenz Curve and Gini Index, *Review of Economics and Statistics*, **54**, 306–316.

Gibrat, R. (1931) *Les Inégalits économiques*, Sirey, Paris.

Kakwani, N. C. and N. Podder (1976) Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations, *Econometrica*, **44**(1), 137–149.

Kloek, T. and H. K. van Dijk (1976) Efficient Estimation of Income Distribution Parameters, Report 7616/E, Econometric Institute, Erasmus University, Rotterdam, 15 pages.

Lévy, P. (1925) *Calculs des Probabilits*, Gauthier-Villars, Paris.

Mandelbrot, B. (1960) The Pareto-Lévy Law and the Distribution of Income, *International Economic Review*, **1**, 79–106.

Mandelbrot, B. (1963) The Pareto-Lévy Law and the Distribution of Income, *International Economic Review*, **IV**, 111–115.

March, L. (1898) Quelques Exemples de Distribution de Salaries, *Journal de la Société statistique de Paris*, pp. 193–206.

McAlister, D. (1879) The Law of the Geometric Mean, Proceedings of the Royal Society, vol. 29, pp. 367.

Metcalf, E. C. (1972) *An Econometric Model of the Income Distribution*, Markham Publishing Company, Chicago.

Morgenstern, O. (1963) *On the Accuracy of Economic Observations*, 2nd ed., Princeton University Press, Princeton.

Pareto, V. (1895) La Legge della Domanda, *Giornale degli Economisti*, pp. 59–68, Gennaio.

Pareto, V. (1896) *Ecrits sur la Courbe de la Répartition de la Richesse*, Euvres complétes de Vilfredo Pareto publiées sous la direction de Giovanni Busino, Genéve, Librairie Droz, 1965.

Pareto, V. (1897) *Cours d'Economie Politique*, Lausanne, Rouge, New edition by Bousquet G. H. et G. Busino (1964). Genéve, Librairie Droz.

Pearson, K. (1934) *Tables of the Incomplete Beta-Function*, Cambridge University Press, Cambridge.

Rao, C. R. (1973) *Linear Statistical Inference and its Applications*, 2nd ed., John Wiley, New York.

Rutherford, R. S. G. (1955) Income Distibutions: A New Model, *Econometrica*, **23**(3), 277–294.

Salem, A. B. Z. and T. D. Mount (1974) A Convenient Descriptive Model of Income Distribution: the Gamma Density, *Econometrica*, **42**(6), 1115–1127.

Singh, S. K. and G. S. Maddala (1976) A Function for Size Distribution of Incomes, *Econometrica*, **44**(5), 963–970.

Statistics Canada (1975) Income Distributions by Size in Canada, 1973, Catalogue 13-207, Ottawa.

Thurow, L. C. (1970) Analyzing the American Income Distribution, *American Economic Review*, **48**, 261–269.

US Bureau of the Census (1961-1971) Income of Families and Persons in the United States, Current Population Reports, Series P-60, US Government Printing Office, Washington.

Vinci, F. (1921) Nuovi Contributi allo Studio della Distribuzione dei Redditi, *Giornale degli Economisti e Rivista di Statistica*, **61**, 365–369.

Wold, H. (1961) Unbiased Predictors, Proceedings of the Fourth Berkeley Symposium of Mathematical Statistics and Probability, vol. 1, pp. 719-761.

Wold, H. (1963) On the Consistency of the Least Squares Regression, *Sankya, A*, **25**(2), 211–215.

CHAPTER 2

A Function for Size Distribution of Incomes†

S. K. Singh and G. S. Maddala ‡

Abstract

The paper derives a function that describes the size distribution of incomes. The two functions most often used are the Pareto and the lognormal. The Pareto function fits the data fairly well towards the higher levels but the fit is poor towards the low income levels. The lognormal fits the lower income levels better but its fit towards the upper end is far from satisfactory. There have been other distributions suggested by Champernowne, Rutherford, and others, but even these do not result in any considerable improvement. The present paper derives a distribution that is a generalization of the Pareto distribution and the Weibull distribution used in analyses of equipment failures. The distribution fits actual data remarkably well compared with the Pareto and the lognormal.

1 Introduction

The derivation of a function that describes the size distribution of incomes and various other distributions that show similar shapes is the purpose of this paper. The two functions most often used are the Pareto function and the lognormal. The Pareto function fits the data fairly well toward the higher levels but the fit is poor toward the lower income levels. If one considers the entire range of income, perhaps the fit may be better for the lognormal but the fit toward the upper end is far from satisfactory (Cramer, 1971).

† Reprint of Singh, S.K. and Maddala, G. S. (1976) A Function for the Size Distribution of Incomes, *Econometrica*, 44, 963-970. © The Econometric Society, 1976.

‡ This research is part of an ongoing study on income distributions at the Development Research Center of the World Bank. Any opinions expressed are those of the authors and not of the Bank. The authors would like to thank the referees for helpful comments on an earlier draft.

Earlier, some efforts have been made by Champernowne (1953), Rutherford (1955), Mandelbrot (1960), Fisk (1961) to derive functional forms to describe the size distribution of incomes, based on reasoning about processes of income generation. The present paper derives a function based on the concept of hazard rate or failure rate which has been widely used for deriving distributions in reliability theory and for the analysis of the distribution of life times (see Barlow and Proschan (1965) and Lotka (1956)). The function derived here was also suggested by Burr (1942) though with a different purpose and reasoning. Also, there is a discussion of hazard rates in Gastwirth (1972), though again with a different purpose.

The plan of the paper is as follows: In Section 2 we present a derivation of the function through a discussion of failure rates. Section 3 presents an alternative derivation of the same function. Section 4 presents an empirical illustration, and the final section gives the conclusions.

2 The Genesis of the Function and Characterization through Failure Rate

If the life time of a person is distributed over the random variable x with probability density function $f(x)$, the probability of surviving at least up to time x is $R(x) = \int_x^\infty f(x)dx = 1 - F(x)$. The probability of death in a small interval of time dx is $f(x)$. After one has survived up to age x, the instantaneous death rate at age x, or the force of mortality, is $r(x) = f(x)/(1-F(x))$. This ratio is variously known as the failure rate or the hazard rate and considerable work has been done to study the characterization of distribution functions from this point of view. Distributions are characterized as IFR (increasing failure rate) or DFR (decreasing failure rate distribution) depending upon whether $f(x)/(1-F)$ rises or decreases with x. Generally speaking, one would not expect decreasing failure rate since time is most often the random variable and one does not expect a priori in most of the situations any particular kind of benefit to accrue with time to reduce the failure rate. Most of the distributions used - exponential, gamma function, normal - give IFR. Lognormal gives an increasing section of failure rate, followed by a decreasing section. This property, which appears questionable (see Barlow and Proschan (1965) and Jorgenson *et al.* (1967)) for other situations, is perhaps precisely the reason why it fits, to some extent, the income distribution.

When we change the random variable from time to income, a priori plausibility on theoretical reasoning for DFR after a point is obvious. While aging, as such, may not confer any advantage for living longer or the reduction of the hazard rate, income may help in earning more. The ability to make more money might increase with one's income. The various reasons are just a bit too obvious to be enumerated here. Therefore, it is appealing to consider distributions which are DFR at least after a point for income distribution. While Pareto is a DFR throughout the range, lognormal becomes a DFR only beyond a point.

For certain situations, it is perhaps more instructive to consider the hazard rate in terms of a transform of x rather than x itself. Consider the transform $z = \log x$. We may then try to find out the hazard rate with respect to this transform of $x : r*(z) = (dF/dz)/(1-F)$.

The Pareto diagram in $(\log(1-F), \log x)$ plane can be interpreted from this point of view. The first derivative of the Pareto transform is the hazard rate with respect to z.

A probability density function is defined to be IFR (increasing failure rate) if $(dr(x))/dx \geq 0$. It is called DFR (decreasing failure rate) if $(dr(x))/dx \leq 0$.

Similarly, a probability density function is defined to be IPFR (increasing proportionate failure rate) if $(dr*(z))/dz \geq 0$ and DPFR (decreasing proportionate failure rate) if $(dr*(z))/dz \leq 0$.

It can be easily checked that the Pareto distribution is monotone DFR for $r(x)$ though $r*(z)$ is constant. Lognormal has an $r(x)$ which has an IFR section followed by a DFR section. However, what is interesting is that $r*(z)$ is monotone increasing. This is the reason why the lognormal does not fit well at the high income level. As an empirical regularity, $r*(z)$ approaching constancy for high incomes appears to be fairly well accepted.

The intuitive economic meaning of $r*(z)$ is clear. At any income, it measures the odds against advancing further to higher incomes in a proportionate sense. It is a variable that should be allowed considerable flexibility because one would be interested in finding out its precise shape at varying levels of income. The restrictions imposed both by the shape of the Pareto function and the lognormal are rather severe. In the interpretation given above, lognormal asserts that it is easiest for one to improve one's relative position at low income groups, and the odds go on increasing monotonically tending to infinity as one's income increases. The Pareto distribution implies a constant value of the odds in the $r*(z)$ sense throughout at all income ranges.

A good starting point for deriving the distribution function is then the following: We accept the behaviour of $r*(z)$ toward the upper end of the income, i.e., asymptotic constancy on the basis of accumulated findings and received opinion. However, one must provide for lower $r*(z)$ at the lower income levels. This would mean allowing $r*(z)$ to rise with z and let it reach an asymptote. This can be done again variously. Should $r*(z)$ rise throughout with decreasing rate? Or should it rise first with z at an increasing rate, then a decreasing rate, and then asymptotically reach constancy? We will make the latter assumption.

For purposes of exposition, it is easier to take the negative of the Pareto transform, which is henceforth called $y : y = -\log(1-F))$, $z = \log x$, and $y = f(z); y' > 0$, $y'' > 0$.

We advance the following assumption:

$$y'' = a \cdot y'(\alpha - y'), \tag{2.1}$$

a being constant. We solve this differential equation to get the distribution function. The composite assumption consists of three parts: (A-1) $r*(z)$ reaches asymptotically a constant value a. (A-2) It first increases with an increasing rate, and then with a decreasing rate. (A-3) The rate of increase of $r*(z)$ is zero when the value of $r*(z)$ is zero. Rearranging (2.1) we get

$$\frac{y''}{y'} + \frac{y''}{\alpha - y} = a\alpha. \tag{2.2}$$

Integrating, we get

$$\log y' - \log(\alpha - y') = a\alpha z + c_1 \tag{2.3}$$

where c_1, is a constant of integration. This can be written as

$$\frac{y'}{\alpha - y'} = e^{a\alpha z + c_1}$$

or

$$y' = \frac{\alpha e^{a\alpha z + c_1}}{1 + e^{a\alpha z + c_1}}. \tag{2.4}$$

We note that y',which is the proportional failure rate, is the three-parameter logistic. Integrating (2.4) again we get

$$\log y = \frac{1}{a}\log(1 + e^{a\alpha z + c_1}) + c_2 \tag{2.5}$$

where c_2, is another constant of integration. After we substitute $-\log(1-F)$ for y and $\log x$ for z in (2.5) we get, with some algebra,

$$\log(1-F) = c - \frac{1}{a}\log(b + x^{a\alpha}), \tag{2.6}$$

where $c = (-c_2 - c_1)/\alpha$ and $b = 1/e^{c_1}$. Equation (2.6) gives the distribution function

$$F = 1 - \frac{c}{(b + x^{a\alpha})^{1/a}}. \tag{2.7}$$

The function in (2.7) has four constants. But since $F = 0$ for $x = 0$ we get $c = b^{1/a}$. Thus the three-parameter function is

$$F = 1 - \frac{b^{1/a}}{(b + x^{a\alpha})^{1/a}} \tag{2.8}$$

or

$$F = 1 - \frac{1}{(1 + a_1 x^{a_2})^{a_3}}, \tag{2.9}$$

where $a_1 = 1/b$, $a_2 = a\alpha$, and $a_3 = 1/a$. Note that $F = 0$ for $x = 0$ and, as $x \to \infty$, $F \to 1$.

In summary, F as in (2.9) is characterized by a PFR which is a logistic with respect to "income power", or z. Also, given that characterization, F as derived in (2.9) is unique. In upper income tail, the PFR is the same as for Pareto; at lower incomes it differs.

3 An Alternative Approach

An alternative derivation of the function derived in the previous section can be given in terms of models of decay. Let $F(x)$ be a certain mass at point x $(0 \leq x \leq \infty)$ which decays to zero as $x \to \infty$. dF/dx is the rate of decay. We standardize the initial mass to be one. If dF/dx depends only on the left-out mass $(1-F)$ then the process is said to be "memoryless". For the Poisson process, $dF/dx = a(1-F)$. The Pareto process can also be interpreted as memoryless since it implies

$$dF/dx = a(1-F)^{(1+1/a)}. \tag{2.10}$$

A process that introduces memory would be the so-called Weibull process which leads to the Weibull distribution. This implies

$$dF/dx = ax^b(1-F). \tag{2.11}$$

A generalization that combines elements of both (2.10) and (2.11) would be to start with the equation

$$dF/dx = ax^b(1-F)^c. \tag{2.12}$$

It can be readily verified that the solution to (2.12) gives equation (2.9) where (now, in terms of the parameters in (2.12))

$$a_1 = (c-1)(a/(b+1)), \quad a_2 = b+1, \quad \text{and} \quad a_3 = 1/(c-1).$$

The above derivation suggests the relationship between the Pareto, Weibull, and the distribution suggested here. One might wonder what the relationship is between this distribution and that suggested by Champernowne and Fisk. The distribution considered by Fisk (1961) is given by

$$\frac{dF}{d\phi} = \frac{e^\phi}{(1+e^\phi)^2} \quad \text{where } e^\phi = \left(\frac{x}{x_0}\right)^\alpha.$$

It can be easily verified that

$$\frac{dF}{dx} = \frac{a_1 a_2 x^{a_2-1}}{(1+a_1 x^{a_2})^2} \quad \text{where } a_1 = \left(\frac{1}{x_0}\right)^\alpha \quad \text{and } a_2 = \alpha.$$

Thus, putting $a_3 = 1$ we get the function suggested by Fisk.

4 Empirical Results

Salem and Mount (1974) used the method of maximum likelihood because, for the gamma density they considered, the estimating equations involve only the arithmetic and geometric means. For the present distribution, it is not possible to get any such simple expressions. The estimation of the Pareto distribution is customarily done by regressing $\log(1-F)$ on $\log x$. Fisk (1961) estimates the sech^2 distribution by regressing $\log F/(1-F)$ on $\log x$. For the distribution suggested here we have $\log(1-F) = -a_3 \log(1+a_1 x^{a_2})$.

Hence, following the customary procedures we estimated the parameters by using a nonlinear least squares method and minimizing

$$\sum [\log(1-F) + a_3 \log(1+a_1 x^{a_2})]^2.$$

The data used were from US Bureau of the Census (1960-1972) and the program was the nonlinear regression program from the Harvard computing center that uses the Davidon-Fletcher-Powell algorithm. The estimated parameters are shown in Table 2.1. The fit, as judged by the R^2's, was very good (they were all uniformly high around .99). But since this may not be an adequate evidence, we used some other checks with the results.

Salem and Mount (1974) have given the details of the observed and predicted probabilities for two years, 1960 and 1968, for the lognormal and the gamma. For comparison we plot the predicted probabilities from the present function in the same diagram that Salem and Mount (1974, Fig. 3) used. This is shown in Fig. 2.1. Also, the sum of squared deviations between the predicted and observed probabilities were as follows:

Year	Lognormal	Gamma	Present Function
1960	.01187	.00391	.00261
1969	.00752	.00238	.00156

Another check on the fit is to use the procedure suggested by Gastwirth and Smith (1972) which consists of computing bounds on the Lorenz concentration ratio and computing the implied value of this ratio from the estimated values of the parameters. For the years 1967 through 1970, we estimated these values by numerical integration using the estimated values of the parameters. The results are reported in Table 2.2. As can be easily seen, the estimates of the Lorenz ratio fall within the bounds.

Table 2.1:

Year	a_1	a_2	a_3
1972	.3070	2.064	2.538
1971	.3125	2.139	2.544
1970	.3102	2.121	2.546
1969	.3101	2.131	2.611
1968	.3071	2.111	2.712
1967	.3120	2.012	2.552
1966	.3109	2.197	2.558
1965	.3082	2.127	2.624
1964	.3184	2.080	2.550
1963	.3084	2.051	2.597
1962	.3079	2.063	5.609
1961	.2735	1.972	3.009
1960	.2931	1.992	2.803

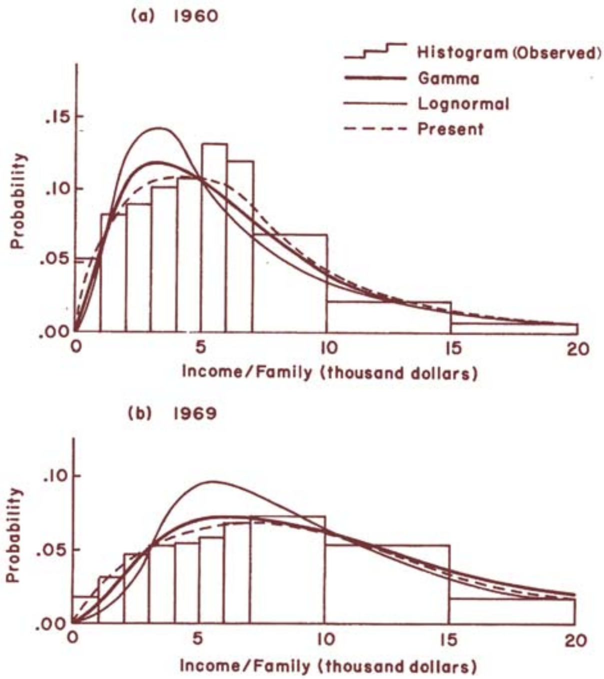


Fig. 2.1: Observed and predicted probabilities of United States families in ten income classes: 1960 and 1969.

Table 2.2:

Year	Bounds Reported by Salem and Mount		Estimates Obtained by the Fitted Function
	Lower	Higher	
1967	3,504	3,556	3,517
1968	3,391	3,457	3,402
1969	3,421	3,506	3,429
1970	3,466	3,565	3,484

5 Conclusions

The paper derives a function to describe the size distribution of incomes based on an analysis of hazard rates or failure rates. The distribution is a generalization of the Pareto and the Weibull distribution studied extensively in the analysis of equipment failures. The sech^2 distribution suggested by Fisk can also be considered as a special case of the distribution suggested here. The distribution has been fitted to United States income data and has been found to fit remarkably well. Earlier, Salem and Mount found that the gamma distribution gives a better fit than the lognormal. We find that the function suggested in the paper gives a better fit than the gamma.

References

Aitchison, J. and J. A. C. Brown (1957) *The Lognormal Distribution with Special Reference to its Uses in Enonomics*, Cambridge University Press, Cambridge.

Atkinson, A. B. (1970) On the Measurement of Inequality, *Journal of Economic Theory*, **2**, 244–263.

Barlow, R. E. and F. Proschan (1965) *Mathematical Theory of Reliability*, Wiley, New York.

Burr, I. W. (1942) Cumulative Frequency Functions, *Annals of Mathematical Statistics*, **13**, 215–235.

Champernowne, D. G. (1953) A Model of Income Distribution, *Economic Journal*, **63**(2), 318–351.

Cramer, J. S. (1971) *Empirical Econometrics*, North Holland, Amsterdam.

Fisk, P. R. (1961) The Graduation of Income Distributions, *Econometrica*, **29**(2), 171–185.

Gastwirth, J. L. (1972) The Estimation of the Lorenz Curve and Gini Index, *Review of Economics and Statistics*, **54**, 306–316.

Gastwirth, J. L. and J. T. Smith (1972) A New Goodness-of-Fit-Test, *ASA Proceedings of the Business and Economic Statistics Section*, pp. 320–322.

Jorgenson, D. J., J. J. McCall and R. Radner (1967) *Optimal Replacement Policy*, North Holland, Amsterdam.

Lotka, A. J. (1956) *Elements of Mathematical Biology*, Dover, New York.

Mandelbrot, B. (1960) The Pareto-Lévy Law and the Distribution of Income, *International Economic Review*, **1**, 79–106.

Rutherford, R. S. G. (1955) Income Distributions: A New Model, *Econometrica*, **23**(3), 277–294.

Salem, A. B. Z. and T. D. Mount (1974) A Convenient Descriptive Model of Income Distribution: the Gamma Density, *Econometrica*, **42**(6), 1115–1127.

US Bureau of the Census (1960-1972) *Current Population Reports*, Series P-60 and P-20, US Government Printing Office, Washington.

CHAPTER 3

Some Generalized Functions for the Size Distribution of Income†

James B. McDonald ‡

Abstract

Many distributions have been used as descriptive models for the size distribution of income. This paper considers two generalized beta distributions which include many of these models as special or limiting cases. These generalized distributions have not been used as models for the distribution of income and provide a unified method of comparing many models previously considered.

Expressions are reported which facilitate parameter estimation and the analysis of associated means, variances, and various measures of inequality.

The distributions considered are fit to US family income and their relative performance is compared.

1 Introduction

Many distributions have been considered as descriptive models for the distribution of income. These include, among others, the lognormal, gamma, beta, Singh-Maddala, Pareto, and Weibull distributions. In many applications, the Singh-Maddala distribution provides a better fit than the gamma which performs much better than the lognormal (McDonald and Ransom (1979); Salem and Mount (1974); Singh and Maddala (1976)). Thurow (1970) adopted the beta distribution as a model for the distribution of income, and this model includes the gamma as a limiting case;

† Reprint of McDonald, J.B. (1984), Some Generalized Functions for the Size Distribution of Income, *Econometrica*, 52, 647-663.

‡ The author appreciates comments and helpful suggestions from Alan Harrison. Dwight Israelsen, and the referees. The author appreciates the assistance of Jeff Green with the computer programming.

hence, the beta will provide at least as good a fit as the gamma. The Singh-Maddala distribution includes the Weibull and Fisk distributions as special cases.

Recently the generalized gamma has been used by Atoda *et al.* (1980); Esteban (1981); Kloek and van Dijk (1978); Taille (1981). Esteban (1981) demonstrates that the generalized gamma has similar tail behavior or includes the lognormal, Weibull, gamma, exponential, normal, and Pareto distributions as special or limiting cases. However, the beta, Singh-Maddala, and Fisk distributions are not included as members of this class of distributions.

In this paper two generalized beta distributions are considered. One of these includes the Singh-Maddala *and* the generalized gamma as special or limiting cases. This distribution provides a useful extension which facilitates a comparison of alternative models within the framework of a generalized model. The second includes the beta used by Thurow and the generalized gamma as special cases.

Section 2 includes a discussion of the generalized beta distributions and the relationships between these distributions and other widely used models for income distribution. Formulas describing associated population characteristics which are useful in the estimation and analysis of empirical data are also reported. Section 3 illustrates some applications of these results.

2 The Models

The generalized gamma (GG) and generalized beta of the first and second kind (GB 1, GB2) are defined by

$$f(y;a,\beta,p) = \frac{ay^{ap-1}e^{-(y/\beta)^a}}{\beta^{ap}\Gamma(p)}, \quad 0 \le y, \tag{3.1}$$

$$g(y;a,b,p,q) = \frac{ay^{ap-1}(1-(y/b)^a)^{q-1}}{b^{ap}B(p,q)}, \quad 0 \le y \le b, \tag{3.2}$$

$$\begin{aligned} h(y;a,b,p,q) &= \frac{ay^{ap-1}}{b^{ap}B(p,q)(1+(y/b)^a)^{p+q}}, \quad 0 \le y \\ &= 0 \quad otherwise. \end{aligned} \tag{3.3}$$

These distributions can be shown to include the beta of the first kind (B1) considered by Thurow, the beta of the second kind (B2), the Singh-Maddala (SM), the lognormal (LN), gamma (GA), Weibull (W), Fisk or Sech^2, and exponential (Exp) distributions as special or limiting cases. These relationships are depicted in Figure 3.1. The special cases of the generalized gamma distribution are carefully developed in the paper by Esteban (1981). Esteban characterizes density functions $f(y)$ in terms of an elasticity $-yf'(y)/f(y)$ and demonstrates that

$$\eta_f(y) = -yd(\ln f(y))/dy = 1 - ap + a(y/\beta)^a \tag{3.4}$$

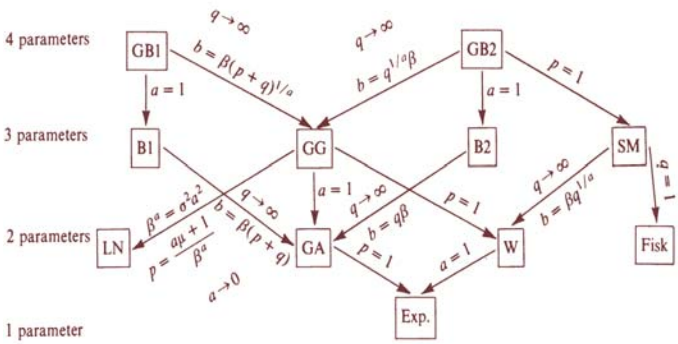


Fig. 3.1: Distribution trees.

uniquely characterizes the generalized gamma in equation (3.1). The corresponding elasticities for (3.2) and (3.3) are given by

$$\eta_g(y) = 1 - ap + \frac{a(q-1)(y/b)^a}{(1-(y/b)^a)} \tag{3.5}$$

and

$$\eta_h(y) = 1 - ap + a(p+q)\frac{(y/b)^a}{1+(y/b)^a}.^4 \tag{3.6}$$

From Figure 3.1, we observe that both of the generalized beta distributions include the generalized gamma as a limiting case:

$$\begin{aligned} f(y;a,\beta,p) &= \lim_{q\to\infty} g(y;\ a,\ \beta(p+q)^{1/a},\ p,\ q) \\ &= \lim_{q\to\infty} h(y;\ a,\ \beta(q)^{1/a},\ p,\ q) \end{aligned}\quad. \tag{3.7}$$

The details associated with these results are included in the Appendix. The *generalized beta of the second kind* is a particularly useful family of distributions and includes the generalized gamma, beta of the second kind, the Singh-Maddala, and all of the previously mentioned associated special cases as members. The distributions of the F statistic (variance ratio) are also a special case of the generalized beta of the second kind.[5]

[4] A referee pointed out that the monotonicity of (3.4), (3.5), and (3.6)) implies that associated graphs showing the relationship of the natural logarithm of the density function and $\ln y$ must be concave.

[5] The author has found that some of these distributions are known by different names in other disciplines. For example, Arnold (1980) refers to a GB2 with a non-zero threshold as a Feller-Pareto distribution. The Singh-Maddala function is a member of the Burr family (type 12) and has

Table 3.1: Distributions and Population Characteristics[a]

Model	Distribution Function	Moments	Gini Coefficient
GB1	$\frac{(y/b)^{ap}}{pB(p,q)}{}_2F_1\left[\begin{matrix} p,\ 1-q;\ (y/b)^a \\ p+1; \end{matrix}\right]$	$\frac{b^h B(p+q,h/a)}{B(p,h/a)}$	$\frac{B(2p+1/a,q)}{B(p,q)B(p+1/a,q)p(ap+1)}{}_4F_3\left[\begin{matrix} 2p+1/a, p, p+1/a, 1-q; 1 \\ 2p+q+1/a, p+1, p+1/a+1; \end{matrix}\right]$
GG	$\frac{e^{-(y/\beta)^a}(y/\beta)^{ap}}{\Gamma(p+1)}{}_1F_1\left[\begin{matrix} 1;\ (y/\beta)^a \\ p+1; \end{matrix}\right]$	$\frac{\beta^h \Gamma(p+h/a)}{\Gamma(p)}$	$\frac{1}{2^{2p+1/a}B(p,p+1/a)}\left\{\left(\frac{1}{p}\right){}_2F_1\left[\begin{matrix} 1, 2p+1/a;\ \frac{1}{2} \\ p+1; \end{matrix}\right] - \left(\frac{1}{p+1/a}\right){}_2F_1\left[\begin{matrix} 1, 2p+1/a; \frac{1}{2} \\ p+1/a+1; \end{matrix}\right]\right\}$
GB2	$\frac{\left(\frac{(y/b)^a}{1+(y/b)^a}\right)^p}{pB(p,q)}{}_2F_1\left[\begin{matrix} p, 1-q;\ \frac{(y/b)^a}{1+(y/b)^a} \\ p+1; \end{matrix}\right]$	$\frac{b^h B(p+h/a,q-h/a)}{B(p,q)}$	$\frac{B(2q-1/a,2p+1/a)}{B(p,q)B(p+1/a,q-1/a)}\left\{\left(\frac{1}{p}\right){}_3F_2\left[\begin{matrix} 1, p+q, 2p+1/a;\ 1 \\ p+1, 2(p+q); \end{matrix}\right] - \left(\frac{1}{p+1/a}\right){}_3F_2\left[\begin{matrix} 1, p+q, 2p+1/a;\ 1 \\ p+1/a+1, 2(p+q); \end{matrix}\right]\right\}$
B1	$\frac{(y/b)^p}{pB(p,q)}{}_2F_1\left[\begin{matrix} p, 1-q;\ (y/b) \\ p+1; \end{matrix}\right]$	$\frac{b^h B(p+q,h)}{B(p,h)}$	$\frac{B(p+q,\frac{1}{2})B(p+\frac{1}{2},\frac{1}{2})}{B(q,\frac{1}{2})\pi}$

[a] ${}_pF_q$ and B(,), respectively denote generalised hypergeometric series and beta functions which are defined in the Appendix. $\Lambda(y|\mu,\sigma^2)$ is standard notation for the distribution of a lognormal random variable; cf. Aitchison and Brown (1969). The results for the exponential can be obtained by letting $a=1$ in the row corresponding to the Weibull distribution

Table 3.1: Cont.

Model	Distribution Function	Moments	Gini Coefficient
B2	$\frac{\left(\frac{(y/b)}{1+y/b}\right)^p}{pB(p,q)}\, {}_2F_1\left[\begin{matrix} p, 1-q; & \frac{y/b}{1+y/b} \\ p+1; & \end{matrix}\right]$	$\frac{b^h B(q-h,p+h)}{B(p,q)}$	$\frac{2B(2p,2q-1)}{pB^2(p,q)}$
SM	$1-\frac{1}{(1+(x/b)^a)^q}$	$\frac{b^h B(1+h/a,q-h/a)}{B(1,q)}$	$1-\frac{\Gamma(q)\Gamma(2q-1/a)}{\Gamma(q-1/a)\Gamma(2q)}$
LN	$\Lambda(y\|\mu,\sigma^2)$	$e^{h\mu+h^2\sigma^2/2}$	$2N\left[\sigma/\sqrt{2};0,1\right]-1$
GA	$\frac{e^{-(y/\beta)}(y/\beta)^p}{\Gamma(p+1)}\, {}_1F_1\left[\begin{matrix} 1; y/\beta \\ p+1; \end{matrix}\right]$	$\frac{\beta^h\Gamma(p+h)}{\Gamma(p)}$	$\frac{\Gamma(p+\frac{1}{2})}{\Gamma(p+1)\sqrt{\pi}}$
W	$1-e^{-x^a/\beta^a}$	$\beta^h\Gamma(1+h/a)$	$1-(1/2^{1/a})$
F	$1-\frac{1}{(1+(y/b)^a)}$	$b^h\Gamma(1+h/a)\Gamma(1-h/a)$	$1/a$

Given an arbitrary estimation criterion, the higher a distribution is on a branch in Figure 3.1, the better it will perform as measured by the same criterion, e.g., least squares estimators of the generalized beta distribution of the first kind will have a sum of squared errors at least as small as the corresponding sum of squared errors for the lognormal distribution. However, no such conclusions can be drawn about the relative performance of distributions on different branches, e.g., the generalized gamma and Singh-Maddala distributions. If estimation is based upon one criterion such as maximum likelihood, distributions higher on a branch in Figure 3.1 will not necessarily perform better when compared according to another criterion such as sum of squared errors.

Expressions for the moments, distribution functions, and several measures of inequality corresponding to (3.1), (3.2), and (3.3) can be expressed in terms of

$$I(x,h) = \int_0^x y^h f(y)dy, \tag{3.8}$$

$$I^*(i,j) = \int_0^\infty x^i f(x) \int_0^x y^j f(y)dydx. \tag{3.9}$$

The associated moments, conditional on their existence, and the distribution function are defined by

$$E(y^h) = \lim_{x\to\infty} I(x,\ h), \tag{3.10}$$

$$F(y) = I(y,0), \tag{3.11}$$

respectively. The relative mean deviation of Pietra (P) and Gini (G) measures [6] of inequality can be expressed in terms of (3.8) and (3.9) by

$$P = E(|y-\mu|)/2\mu = I(\mu,0) - I(\mu,1)/\mu \tag{3.12}$$

and

$$G = E(|y-x|)/2\mu = (1/\mu)(I^*(1,0) - I^*(0,1)) \tag{3.13}$$

where $\mu = E(y)$. Expressions for $I(x,h)$ and $I^*(i,j)$ corresponding to (3.1), (3.2), and (3.3) are derived in the Appendix. In each case, these expressions are functions of the parameters defining the distribution function under consideration (a,b,β,p,q). Table 3.1 includes expressions for the distribution function, moments, and Gini coefficients for each of the distributions discussed. These expressions are functions of the parameters in the respective models and are useful in estimation and analysis of

also been referred to as a Beta-P distribution (Cronin (1979); Johnson and Kotz (1970)). Another special case of the generalized beta of the second kind encountered in other areas of application corresponds to q = 1 and is known as a three-parameter kappa distribution, Beta-*k* or Burr distribution of the third type (Tadikamalla, 1980). If *p* and *a* are both equal to one, then the corresponding distribution has been referred to as the Lomax distribution. The power and uniform distributions, among others, are special cases of the generalized beta of the first kind and B1.

[6] Gastwirth (1972) discusses some issues associated with nonparametric estimation of the Pietra index or relative mean deviation and Gini coefficient and their interpretation and historical background. McDonald and Ransom (1981) discuss some related inferential issues. Also see Kendall and Stuart (1961).

population characteristics. The expressions for the distribution function facilitate parameter estimation based upon data in a grouped format. Given parameter estimates, these results can then be used to estimate corresponding population characteristics of interest as well as providing indirect checks on the validity of the parameter estimates. This point will be illustrated by means of an example in the next section. The estimated population characteristics depend upon the estimation technique and the assumed distribution function. This point is covered in more detail in McDonald and Ransom (1979).

3 Applications

Estimation of the models discussed in Section 2 involves nonlinear techniques. Numerous estimation problems can arise in nonlinear estimation which will yield questionable results. An indirect check of the validity of the parameter estimates obtained from a nonlinear optimization routine is provided by comparing estimated population characteristics such as the mean with independently obtained results where available. The expressions in Table 3.1 facilitate such a comparison, and Section 3.1 provides an example of this. The relative performance of alternative models for the distribution of income is compared in Section 3.2.

3.1 Analysis of Parameter Estimates

Thurow's (1970) widely cited paper provides an example of estimation problems and an application of expressions in Table 3.1 in the detection of questionable results. The underlying distribution of income is assumed to be modeled by a beta density function.

$$g(y;\, a=1,\, b,\, p,\, q) = \frac{y^{p-1}(b-y)^{q-1}}{B(p,q)b^{p+q}}, \quad 0<y<b, \quad p,\, q>0, \tag{3.14}$$

which corresponds to (3.2) with $a = 1$. Thurow assumed that the maximum income (b) was equal to \$15,000 and obtained separate estimates of p and q for the distribution of income (1959 dollars) of families and unrelated individuals for whites and nonwhites for the period 1949-1966. Income characteristics associated with the estimated parameter values for (p,q,b) are inferred and their relationship with hypothesized explanatory variables considered. Thurow's results raise questions as to whether economic growth is associated with a more egalitarian distribution as well as suggesting that inflation may lead to a more equal distribution of income for whites. The accuracy of the estimated (p,q)'s is a critical element in the validity of the analysis of the estimated relationship between the hypothesized explanatory variables and the distribution of income. Thurow's estimates of $(p,\, q)$ were not reported in his paper, but were provided on request and are given in Table 3.2[7] The mean and Gini coefficient associated with the beta function (B1) in Table 3.1 are given by

$$E(y) = \frac{bp}{p+q}, \tag{3.15}$$

$$G = \frac{\Gamma(p+q)\Gamma(p+\frac{1}{2})\Gamma(q+\frac{1}{2})}{\Gamma(p+q+\frac{1}{2})\Gamma(p+1)\Gamma(q)\Gamma(\frac{1}{2})}. \tag{3.16}$$

The mean income level and Gini coefficients implied by Thurow's estimates can be readily obtained from equations (3.15) and (3.16) and are reported in Table 3.2.

The corresponding estimates reported in census publications are also given in this table to provide a useful comparison.

An analysis of the entries in Table 3.2 suggests that the distribution of income for whites is more egalitarian with a higher mean than for nonwhites. This qualitative result is consistent with Thurow's estimates as well as those reported in the current population reports; however, other implications of these two sets of estimates are not. For example, all of the associated estimated density functions are either "$\cup$" shaped $(p < 1, q < 1)$ or "$\imath$" shaped $(p < 1, q > 1)$ rather than "$\cap$" shaped. The agreement between the implied and census estimates of the mean is much closer than for the Gini coefficients. The magnitude and inter-temporal behavior (reductions in excess of 30 per cent) of the associated Gini coefficients implied by the estimated parameters (p, q) for the period under consideration are inconsistent with the census estimates and provide additional evidence of an estimation problem.

Thus there is relatively close agreement between the two estimates of mean income, but very poor agreement between the measures of inequality. The estimation procedure appears to have roughly preserved the mean characteristic, but implicitly modeled intra and/or inter-group variation incorrectly. The results could also have been partially due to the conjunction of the nature of the income groups and treatment of the maximum income.

The next section includes examples in which estimated distributions of the form considered by Thurow provide relatively accurate estimates of population characteristics. The parameters p, q, and b are estimated. The corresponding densities are "$\cap$" shaped and the associated estimates of the mean and Gini coefficient are very close to those reported by the census publications. The large values of b and q merely correspond to the estimated beta density (B1) being closely approximated by its limiting form, the gamma.

[7] The author appreciates Professor Thurow's assistance in providing these estimates and suggestions, but has been unable to replicate the results.

Table 3.2: Thurow's estimates of p, q*

	Whites						Nonwhites					
Year	p	q	mean**		Gini**		p	q	mean**		Gini**	
1949	.258	.666	4.188	(4.052)	.615	(.404)	.160	.930	2.202	(2.044)	.752	(.443)
1950	.279	.687	4.332	(4.294)	.600	(.407)	.172	.930	2.341	(2.245)	.738	(.438)
1951	.269	.625	4.513	(4.441)	.596	(.387)	.182	.908	2.505	(2.334)	.725	(.433)
1952	.298	.649	4.720	(4.579)	.576	(.398)	.205	.921	2.731	(2.518)	.702	(.407)
1953	.327	.667	4.935	(4.798)	.556	(.395)	.228	.920	2.979	(2.638)	.679	(.428)
1954	.334	.697	4.859	(4.741)	.557	(.401)	.217	.929	2.84	(2.579)	.691	(.456)
1955	.368	.718	5.083	(5.034)	.536	(.397)	.225	.913	2.966	(2.679)	.681	(.431)
1956	.411	.731	5.398	(5.311)	.511	(.391)	.249	.978	3.044	(2.875)	.666	(.427)
1957	.406	.728	5.370	(5.229)	.513	(.385)	.269	1.025	3.118	(2.860)	.652	(.435)
1958	.411	.750	5.310	(5.291)	.514	(.388)	.276	1.075	3.064	(2.893)	.651	(.448)
1959	.460	.765	5.633	(5.571)	.488	(.396)	.286	1.051	3.209	(2.977)	.641	(.452)
1960	.504	.815	5.732	(5.646)	.473	(.398)	.330	1.061	3.559	(3.276)	.608	(.459)
1961	.622	.979	5.828	(5.817)	.443	(.408)	.346	1.173	3.417	(3.268)	.607	(.462)
1962	.663	.971	6.086	(5.987)	.426	(.395)	.338	1.107	3.509	(3.278)	.607	(.443)
1963	.712	.985	6.294	(6.167)	.411	(.396)	.356	1.073	3.737	(3.513)	.591	(.440)
1964	.785	1.017	6.534	(6.333)	.391	(.400)	.406	1.095	4.057	(3.788)	.562	(.444)
1965	.842	1.029	6.750	(6.552)	.376	(.393)	.452	1.124	4.302	(3.859)	.538	(.427)
1966	.955	1.044	6.166	(6.912)	.348	(.390)	.514	1.104	4.756	(4.192)	.504	(.426)

* Thurow did not estimate the parameter b, but rather assumed it to be 15 ($ 15,000) and included any higher Incomes in the group with an upper bound of $15,000. The mean and Gini coefficients were evaluated using equations (15) and (16).

** The numbers in parentheses are the corresponding census estimates reported in current population reports (P60). The nominal figures for mean Income were adjusted by the CPI to obtain the figures in I959 dollars.

3.2 Estimation and Comparison of Alternative Distributions

The generalized gamma and generalized beta of the first and second kinds and special cases previously discussed were fit to US family nominal income for 1970-1980. The data were in a grouped format and the corresponding multinomial likelihood function is given by

$$L(\Theta) = N! \prod_{i=1}^{g} \frac{(P_i(\Theta))^{n_i}}{n_i!}$$

where $P_i(\Theta) = \int_{I_i} f(y;\ \Theta)dy$ denotes the predicted fraction of the population in the ith of g income groups defined by $I_i = [y_{i-1}, y_i)$. The $(n_i/N)$are the corresponding observed relative frequencies $(N = \sum n_i)$. The estimators obtained by maximizing the multinomial likelihood function will be asymptotically efficient relative to other estimators based on grouped data; however, they may be less efficient than maximum likelihood estimators based on individual observations (cf. Cox and Hinckley (1974)).[8] The results of this estimation for 1970, 1975, and 1980 are reported in Tables 3.3, 3.4, and 3.5. The reported values for the mean and Gini coefficients were obtained by substituting the estimated parameters into the relevant expressions given in Table 3.1.

The generalized beta of the second kind provides a better fit than the generalized gamma of the first kind based upon a sum of squared or absolute errors criterion (SSE, SAE), chi-square (χ^2), or log-likelihood criterion[9]. The values of the log-likelihood functions are consistent with the logical relationships between the

[8] The program GQOPT obtained from Richard Quandt was used to maximize the multinomial log-likelihood function. A convergence criterion of 10^{-8} was specified. The data used are given in the following table, and were taken from the Census Population Reports.

Endpoint (in thousands)	1970	1975	1980
2.5	6.6	3.5	2.1
5.0	12.5	8.5	4.1
7.5	15.2	10.6	6.2
10	16.6	10.6	6.5
12.5	15.8	11.4	7.3
15	11.0	10.9	6.9
20	13.1	18.8	14.0
25	4.6	11.6	13.7
35	3.0	9.5	19.8
50	1.1	3.2	12.8
∞	0.5	1.4	6.7

For cases in which the percentages do not add to 100, the percentages used in estimation were obtained by transforming the reported figures by multiplying them by (100/sum of percentages).

[9] The SSE, SAE, and χ^2 values are obtained by evaluating $\sum_{i=1}^{11}\left(\frac{n_i}{N} - p_i\left(\widehat{\Theta}\right)\right)^2$, $\sum_{i=1}^{11}\left|\frac{n_i}{N} - P_i\left(\widehat{\Theta}\right)\right|$ and $N\sum_{i=1}^{11}\left(\frac{n_i}{N} - P_i\left(\widehat{\Theta}\right)\right)^2 / P_i\left(\widehat{\Theta}\right)$, respectively.

Table 3.3: Estimated Distribution Functions 1970 Family Income

	GB1	GB2	B1	GG	B2	SM	LN	G	W	F	E
a	.8954	5.0573	1.0000	0.8944	1.0000	1.9652	$\mu = 2.1924$	1.0000	1.5603	2.5123	1.0000
$b(\beta)$	649292.	13.5815	389112	(3.4106)	95.0194	18.7288	$\sigma = 0.6977$	(4.8274)	(12.3587)	9.3067	(11.1339)
p	2.8164	0.2961	2.3026	2.8228	2.5556	1.0000	–	2.3026	1.0000	1.0000	1.0000
q	53205.0	0.6708	80621.0	–	22.8234	2.9388	–	–	–	1.0000	–
Mean*	11.119	11.168	11.113	11.122	11.127	11.121	11.425	11.115	11.108	12.261	11.134
Gini**	.354	.337	.352	.354	.355	.350	.379	.352	.359	.398	.500
SSE	.0025	0.0001	0.0022	0.0025	0.0026	0.0014	0.0080	0.0022	0.0027	0.0051	0.0350
SAE	.1399	0.0307	0.1308	0.1399	0.1434	0.0998	0.2429	0.1307	0.1410	0.2043	0.4894
χ^2	1686.9	48.4	1997.6	1683.2	1326.0	644.1	3455.5	1994.9	7621.1	2619.9	11430.4
-ln L	710.6	79.9	722.6	710.6	654.7	363.5	1692.3	722.6	1180.6	1321.9	6051.5

* Census estimate : 11.106
** Census estimate: .354.

Table 3.4: Estimated Distribution Functions 1975 Family Income

	GB1	GB2	B1	GG	B2	SM	LN	G	W	F	E
a	1.0643	3.4977	1.0000	1.0643	1.0000	1.8648	μ=2.5157	1.0000	1.5923	2.4290	1.0000
$b(\beta)$	631397.	20.9572	393308	(8.0305)	984.238	31.5176	σ=0.7248	(6.8115)	(17.2351)	12.9153	(15.6136)
p	2.0376	0.4433	2.2726	2.0378	2.3038	1.0000	–	2.2729	1.0000	1.0000	1.0000
q	162350.	1.1372	57742.4	–	147.444	3.7657	–	–	–	–	1.0000
Mean*	15.472	15.593	15.479	15.475	15.484	15.506	16.094	15.482	15.460	17.368	15.614
Gini**	.353	.352	.354	.353	.355	.353	.392	.354	.353	.412	.500
SSE	.0012	0.0003	0.0013	0.0012	0.0014	0.0008	0.0063	0.0013	0.0014	0.0046	0.0281
SAE	.0912	0.0502	0.0963	0.0913	0.0982	0.0733	0.2226	0.0962	0.0964	0.1994	0.4322
χ^2	553.6	122.6	548.1	553.3	540.6	280.3	3108.7	547.8	1213.2	2200.9	9287.3
-ln L	319.9	117.6	323.3	319.9	322.3	196.2	1453.6	323.3	529.2	1169.4	5191.1

* Census estimate: 15.546

** Census estimate: .358.

Table 3.5: Estimated Distribution Functions 1980 Family Income

	GB1	GB2	B1	GG	B2	SM	LN	G	W	F	E
a	1.4008	2.5373	1.0000	1.4008	1.0000	1.6971	μ=2.9372	1.0000	1.6057	2.2768	1.0000
$b(\beta)$	273102.	40.7667	163.757	(21.8145)	3535660.	87.6981	σ=0.7797	(11.0473)	(26.3368)	19.7450	(24.5954)
p	1.2454	0.6117	1.9173	1.2454	2.1555	1.0000	–	2.1557	1.0000	1.0000	1.0000
q	549517.	2.1329	11.3828	–	320081.	8.3679	–	–	–	1.0000	–
Mean*	23.644	23.931	23.604	23.646	23.810	23.730	25.564	23.815	23.065	27.749	24.596
Gini**	.353	.359	.353	.353	.363	.355	.419	.363	.351	.439	.500
SSE	.0004	0.0002	0.0006	0.0004	0.0008	0.0003	0.0070	0.0008	0.0005	0.0053	0.0234
SAE	.0545	0.0385	0.0609	0.0545	0.0775	0.0495	0.2326	0.0775	0.0561	0.2038	0.4186
χ^2	143.0	84.3	188.4	143.0	353.5	114.2	4620.7	353.6	183.3	2438.9	9107.8
-ln$L$	129.1	99.1	151.6	129.1	228.3	114.5	1859.4	228.3	150.0	1219.5	5108.2

* Census estimate: 23.974

** Census estimate: .365.

distribution functions, i.e., the higher a distribution is on a branch in Figure 3.1, the larger is the associated likelihood function. The other measures of goodness of fit reported need not provide the same rankings.

It is interesting to note that the Singh-Maddala distribution function provides a better fit to the data than any of the distribution functions except for the generalized beta of the second kind regardless of the criterion used for comparison. It even performs better than the four-parameter generalized beta of the first kind for the data set considered. This result is of particular interest due to the relative simplicity of the distribution function. Atoda *et al.* (1980) found that the estimated Singh-Maddala distribution generally outperformed the corresponding B1 or GG for the distribution of Japanese income considered. They used a nonlinear least squares estimation technique and adopted chi-square and SSE criteria for purposes of comparing the distributions. For the US data considered here, the generalized gamma was generally the second best of the three-parameter models regardless of the criterion used for 1975 and 1980.

The lognormal is, with few exceptions, worse than any of the other two-parameter models. The lognormal overstates income in the upper tail (last group) and has a larger estimated mean income and suggests greater dispersion than for the other models with two or more parameters.

The likelihood ratio test provides the basis for comparing nested models. The asymptotic distribution of $2\,[\ln L(\Theta_{ML}) - \ln L(\Theta_R)]$ is chi-square with degrees of freedom equal to the number of independent restrictions imposed on the more general model in order to yield the nested model. Θ_{ML}, and Θ_R, respectively, denote maximum likelihood estimators of the general and restricted model. The literature dealing with nonnested hypotheses provides an approach for comparing distributions on different branches.

The differences between GB1 and GG are not statistically significant for any of the three years and have almost identical characteristics. Similarly, GB1, GG, B1, and GA have almost identical characteristics for 1975. Other differences between the nested models appeared to be statistically significant using either a likelihood ratio or Wald test.

The chi-square statistic provides a test of "goodness of fit" and has an asymptotic distribution which is chi-square with degrees of freedom equal to one less than the difference between the number of income groups and number of parameters. There is considerable variation in the value of chi-square across distributions, but all must be rejected at conventional levels of significance. This result is common in applications involving large sample sizes (Kloek and van Dijk, 1978; McDonald and Ransom, 1979), and suggests that it might be productive to consider the impact of sample size upon the power of such tests.

In summary, the generalized beta of the second kind provided the best relative fit and included many other distributions as special or limiting cases. The differences were statistically significant. The Singh-Maddala (or Burr) distribution provided a better fit than the generalized beta of the first kind (four parameters) and all of the two- and three-parameter models considered. The Singh-Maddala distribution function has a closed form which greatly facilitates estimation and analysis of results.

Appendix

Derivation of $I(x,h)$ *and* $I^*(i,j)$

The incomplete moments $I(x,h)$ for the generalized gamma can be obtained by substituting (3.1) into (3.8) to obtain

$$I(x,y) = \int_0^x \frac{ay^{ap+h-1}}{b^{ap}\Gamma(p)} e^{-(y/b)^a} dy. \tag{3.17}$$

Equation (3.17) can be evaluated by making the change of variable $s = (y/b)^a$; hence,

$$\begin{aligned} I(x,h) &= b^h \int_0^{(x/b)^a} \frac{s^{p+h/a-1}}{\Gamma(p)} e^{-s} ds \\ &= \frac{b^h e^{-(x/b)^a} (x/b)^{ap+h}}{\Gamma(p)(p+\frac{h}{a})} {}_1F_1 \left[\begin{matrix} 1;\ (x/b)^a \\ p+\frac{h}{a}+1; \end{matrix} \right] \end{aligned} \tag{3.18}$$

(cf. McDonald and Jensen (1979); Rainville (1960)).

The

$${}_1F_1 \left[\begin{matrix} a; \\ b; y \end{matrix} \right]$$

is the confluent hypergeometric series and is a special case of the generalized hypergeometric series defined by

$${}_pF_q \left[\begin{matrix} a_1, \ldots, a_p;\ x \\ b_1, \ldots, b_q; \end{matrix} \right] = \sum_{i=0}^{\infty} \frac{(a_1)_i \ldots (a_p)_i}{(b_1)_i \ldots (b_q)_i} \frac{x^i}{i!}$$

where $(a)_i = (a)(a+1)\ldots(a+i-1)$. (A.2) can be used to evaluate $I^*(i,j)$.

$$I^*(i,j) = \int_0^{\infty} x^i f(x) \int_0^x y^j f(y) dy dx = \int_0^{\infty} x^i f(x) I(x,j) dx$$

$$= \frac{b^j}{\Gamma(p)(p+j/a)} \int_0^{\infty} x^i \frac{ax^{ap-1} e^{-(x/b)^a}}{b^{ap}\Gamma(p)} \frac{e^{-(x/b)^a} x^{ap+j}}{b^{ap+j}} {}_1F_1 \left[\begin{matrix} 1;\ (x/b)^a \\ p+\frac{j}{a}+1; \end{matrix} \right] dx \tag{3.19}$$

$$= \frac{a}{\Gamma^2(p)(p+j/a)b^{2ap}} \int_0^{\infty} x^{2ap+i+j-1} e^{-2(x/b)^a} {}_1F_1 \left[\begin{matrix} 1;\ (x/b)^a \\ p+j/a+1; \end{matrix} \right] dx.$$

Making the change of variable $s = (x/b)^a$,

$$I^*(i,j) = \frac{b^{i+j}}{\Gamma^2(p)(p+j/a)} \int_0^\infty s^{2p+(i+j)/a-1} e^{-2s} {}_1F_1\left[\begin{matrix} 1;\ s \\ p+\frac{j}{a}+1; \end{matrix}\right] ds$$

$$= \frac{b^{i+j}\Gamma(2p+\frac{i+j}{a})}{\Gamma^2(p)(p+j/a)2^{2p+(i+j)/a}} {}_2F_1\left[\begin{matrix} 1, 2p+\frac{i+j}{a};\ \frac{1}{2} \\ p+\frac{j}{a}+1; \end{matrix}\right]. \tag{3.20}$$

See Gradshteyn and Rhyzik (1965, p. 851, 7.5229).

The derivation for the generalized beta for the first kind is similar to that for the generalized gamma except that the evaluation of $I(x,h)$ makes use of the incomplete beta function (Rainville, 1960) and the evaluation of $I^*(i,j)$ makes use of an integral reported in Gradshteyn and Rhyzik (1965, p. 850). The corresponding results can be shown to be

$$I(x,h) = \frac{b^h (x/b)^{ap+h}}{B(p,q)\left(p+\frac{h}{a}\right)} {}_2F_1\left[\begin{matrix} p+\frac{h}{a}, 1-q;\ \left(\frac{x}{b}\right)^a \\ p+\frac{h}{a}+1; \end{matrix}\right],$$

$$I^*(i,j) = \frac{b^{i+j} B\left(2p+\frac{i+j}{a}, q\right)}{B^2(p,q)\left(p+\frac{j}{a}\right)} {}_3F_2\left[\begin{matrix} 2p+\frac{i+j}{a}, p+\frac{j}{a}, 1-q;\ 1 \\ 2p+q+\frac{i+j}{a}, p+\frac{j}{a}+1; \end{matrix}\right], \tag{3.21}$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

The derivation of $I\ (x,h)$ for the generalized beta of the second kind also makes use of the incomplete beta function. The evaluation of the corresponding $I^*(i,j)$ involves the integral reported in Gradshteyn and Rhyzik (1965, p. 849, #5). $I(x,h)$ and $I^*(i,j)$ are given by

$$I(x,h) = \frac{b^h z^{p+h/a}}{B(p,q)\left(p+\frac{h}{a}\right)} {}_2F_1\left[\begin{matrix} p+\frac{h}{a}, 1+\frac{h}{a}-q;\ z \\ p+\frac{h}{a}+1; \end{matrix}\right] \tag{3.22}$$

where

$$z = \frac{(x/b)^a}{1+(x/b)^a}$$

and

$$I^*(i,j) = \frac{b^{i+j} B\left(2p+\frac{i+j}{a}, q-\frac{i}{a}\right)}{B^2(p,q)\left(p+\frac{j}{a}\right)} {}_3F_2\left[\begin{matrix} p+\frac{j}{a}, 1+\frac{j}{a}-q, 2p+\frac{i+j}{a};\ 1 \\ p+\frac{j}{a}+1, 2p+q+\frac{j}{a}; \end{matrix}\right].$$

Expressions for the moments and distributions can be easily obtained from equations (3.10) and (3.11) using the expressions for $I(x,h)$ in (3.18), (3.21), and (3.22).

The expressions for $I^*(i,j)$ can be substituted into (3.13) to yield formulas for the Gini measure of inequality. In some instances these equations have been transformed into simpler representations reported in Table 3.1.

Limiting behavior of the generalized beta of the first kind

The generalized beta density of the first kind is given by

$$g(y) = \frac{ay^{ap-1}(1-(y/b)^a)^{q-1}}{b^{ap}B(p,q)}. \tag{3.23}$$

This density function approaches the generalized gamma density as $q \to \infty$ where the scale factor changes with q as

$$b = \beta(p+q)^{1/a}. \tag{3.24}$$

Making this substitution into (3.23) for b yields

$$g(y) = \left(\frac{ay^{ap-1}}{\Gamma(p)\beta^{ap}}\right)\left(\frac{\Gamma(p+q)}{\Gamma(q)(p+q)^p}\right)\left(1-\frac{y^a}{\beta^a(p+q)}\right)^{q-1}. \tag{3.25}$$

For large values of q, the gamma function can be approximated by Stirling's formula,

$$\Gamma(x) \doteq e^{-x}x^{x-1/2}\sqrt{2\pi}. \tag{3.26}$$

See Kendall and Stuart (1961, v. 1, p. 811).

The second bracketed expression in (3.25) can be shown to approach 1 by making the substitution (3.26) for the gamma functions and taking the limit as $q \to \infty$. Similarly, the last bracketed expression in (3.25) approaches $e^{-(y/\beta)^a}$ as $q \to \infty$. Therefore the generalized beta in (3.23) approaches the generalized gamma (3.1) as $q \to \infty$.

Limiting behavior of the generalized beta of the second kind

Substituting $b = q^{1/a}\beta$ into (3.3) and grouping terms yields

$$h(y) = \left(\frac{ay^{ap-1}}{\beta^{ap}\Gamma(p)}\right)\left(\frac{\Gamma(p+q)}{\Gamma(q)q^p}\right)\left[\frac{1}{\left(1+\frac{y^a}{\beta^a q}\right)^{p+q}}\right]. \tag{3.27}$$

Using (3.26) in the second bracketed expression in (3.27) and taking the limit of (3.27) as $q \to \infty$ yields the expression for the generalized gamma density given in equation (3.1).

References

Aitchison, J. and J. A. C. Brown (1969) *The Lognormal Distribution with Special References to its Uses in Economics*, Cambridge University Press, Cambridge.

Arnold, B. (1980) Pareto Distributions: Pareto and Related Heavy-tailed Distributions, mimeographed manuscript, University of California at Riverside.

Atoda, N., T. Suruga and T. Tachibanaki (1980) Statistical Inference of Functional Forms for Income Distribution, manuscript. Kyoto University.

Cox, D. R. and D. V. Hinckley (1974) *Theoretical Statistics*, Chapman and Hall, London.

Cronin, D. C. (1979) A Function for the Size Distribution of Income: A Further Comment, *Econometrica*, **47**, 773–774.

Esteban, J. (1981) Income-Share Elasticity, Density Functions and the Size Distribution of Income, mimeographed manuscript, University of Barcelona.

Gastwirth, J. L. (1972) The Estimation of the Lorenz Curve and Gini Index, *Review of Economics and Statistics*, **54**, 306–316.

Gradshteyn, I. S. and I. M. Rhyzik (1965) *Tables of Integrals, Series, and Products*, Academic Press, New York.

Johnson, N. L. and S. Kotz (1970) *Continuous Univariate Distributions*, vol. 1, John Wiley and Sons, New York.

Kendall, M. G. and A. Stuart (1961) *The Advanced Theory of Statistics*, vol. 1, 2nd ed., Hafner, New York.

Kloek, T. and H. K. van Dijk (1978) Efficient Estimation of Income Distribution Parameters, *Journal or Econometrics*, **8**, 61–74.

McDonald, J. B. and B. C. Jensen (1979) An Analysis of Estimators of Alternative Measures of Income Inequality Associated with the Gamma Distribution Function, *Journal of the American Statistical Association*, **74**, 856–860.

McDonald, J. B. and M. J. Ransom (1979) Functional Forms, Estimation Techniques and the Distribution of Income, *Econometrica*, **47**, 1513–1526.

McDonald, J. B. and M. J. Ransom (1981) An Analysis of the Bounds for the Gini Coefficient, *Journal of Econometrics*, **17**, 177–188.

Rainville, E. D. (1960) *Special Functions*, MacMillan, New York.

Salem, A. B. Z. and T. D. Mount (1974) A Convenient Descriptive Model of Income Distribution: the Gamma Density, *Econometrica*, **42**(6), 1115–1127.

Singh, S. K. and G. S. Maddala (1976) A Function for Size Distribution of Incomes, *Econometrica*, **44**(5), 963–970.

Tadikamalla, P. R. (1980) A Look at the Burr and Related Distributions, *International Statistical Review*, **48**, 337–349.

Taille, C. (1981) Lorenz Ordering Within the Generalized Gamma Family of Income Distributions, in G. P. Taille, C. and B. Balderssari (eds.) *Statistical Distributions in Scientific Work*, vol. 6, pp. 181–192, Reidel, Boston.

Thurow, L. C. (1970) Analyzing the American Income Distribution, *American Economic Review*, **48**, 261–269.

US Bureau of the Census (1961-1971) Income of Families and Persons in the United States, Current Population Reports, Series P-60, US Government Printing Office, Washington.

CHAPTER 4

Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations†

N. C. Kakwani and N. Podder ‡

Abstract

This paper introduces a new coordinate system for the Lorenz curve . Particular attention is paid to a special case of wide empirical validity. Four alternative methods have been used to estimate the proposed Lorenz curve from the grouped data. The well known inequality measures are obtained as the function of the estimated parameters of the Lorenz curve. In addition the frequency distribution is derived from the equation of the Lorenz curve. An empirical illustration is presented using the data from the Australian Survey of Consumer Expenditure and Finances 1967-68.

1 Introduction

The Lorenz curve is widely used to represent and analyze the size distribution of income and wealth. The curve relates the cumulative proportion of income units to the cumulative proportion of income received when the units are arranged in ascending order of their income.

The equation of the Lorenz curve can be derived from the density function of the income distribution. In practice, the density function is not known, and one approach

† Reprint of Kakwani, N. C., and N. Podder (1976) Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations, *Econometrica*, 44, 137-148.

‡ This research was partially supported by a grant given by the Reserve Bank of Australia. The authors would like to thank two referees for their valuable comments.

has been to fit some well known density function, for example, the Pareto or the lognormal. The shortcoming of such an approach is that the usual density function hardly gives a reasonably good fit to actual data. An alternative approach is to find an equation of the Lorenz curve which would fit actual data reasonably well. The Lorenz curve has a number of properties which can be effectively utilized to specify such an equation.

The purpose of this paper is to introduce a new coordinate system for the Lorenz curve. Particular attention is paid to a special case of wide empirical validity. Four alternative methods have been used to estimate the proposed Lorenz curve from the grouped observations. The well known inequality measures are obtained as the function of the estimated parameters of the Lorenz curve. The procedure of estimating the asymptotic standard errors of the inequality measures is also provided. In addition the frequency distribution is derived from the equation of the Lorenz curve.

A new representation of the Lorenz curve is introduced in the next section. Section 3 provides the relationship between this representation of the Lorenz curve and a number of conventional measures of income inequality. Section 4 describes a number of estimation methods. The last section reports some empirical results based on the data from the Australian Survey of Consumer Expenditure and Finances (1967-1968).

2 A New Co-Ordinate System for the Lorenz Curve

Suppose that income X of a family is a random variable with probability distribution function $F(x)$. Further, if it is assumed that mean μ of the distribution exists and X is defined only for positive values,[2] the first moment distribution function of X is then given by

$$F_1(x) = \frac{1}{\mu}\int_0^x Xg(X)dX \tag{4.1}$$

where $g(X)$ is the density function.

The Lorenz curve is the relationship between $F(x)$ and $F_1(x)$. The curve is shown in Figure 4.1. The equation of the line $F_1 = F$ is called the egalitarian line, which in the diagram, is the diagonal through the origin of the unit square.

Let P be any point on the curve with co-ordinates (F, F_1), and

$$\pi = \frac{1}{\sqrt{2}}(F+F_1) \quad \text{and} \quad \eta = \frac{1}{\sqrt{2}}(F-F_1); \tag{4.2}$$

then η will be the length of the ordinate from P on the egalitarian line and π will be the distance of the ordinate from the origin along the egalitarian line. Since the Lorenz curves lie below the egalitarian line, $F_1 \leq F$ which implies $\eta \geq 0$. Further,

[2] The income X can be negative for some families but is assumed to be always positive here because of notational convenience.

if income is always positive the equation (4.2) will imply η to be less than or equal to π.

The equation of the Lorenz curve in terms of π and η can now be written as:

$$\eta = f(\pi) \tag{4.3}$$

where π varies from zero to $\sqrt{2}$.

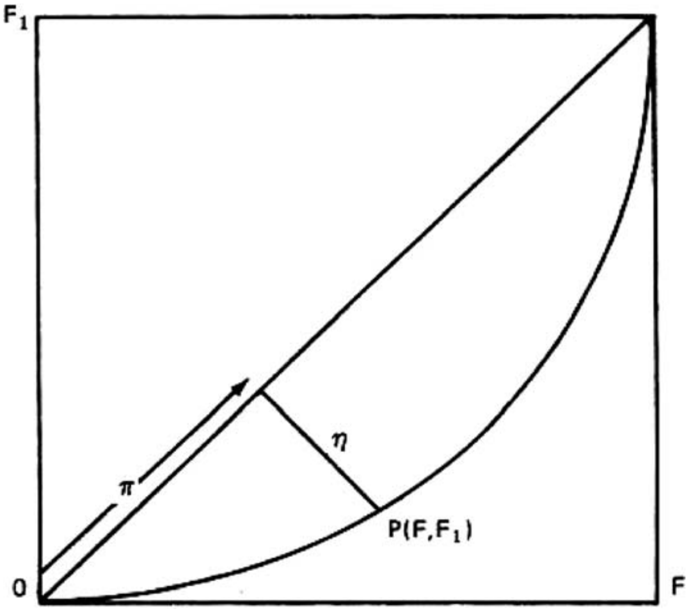


Fig. 4.1:

If $g(X)$ is continuous, the derivatives of $F(x)$ and $F_1(x)$ exist; $dF/dx = g(x)$ and $dF_1/dx = xg(x)/\mu$. Using these values in (4.2) gives the derivatives of η with respect to π as:

$$\frac{d\eta}{d\pi} = \frac{\mu - x}{\mu + x} \tag{4.4}$$

and

$$\frac{d^2\eta}{d\pi^2} = -\frac{2\sqrt{2}\mu^2}{g(x)(\mu + x)^3} < 0. \tag{4.5}$$

Thus η will be maximum at $x = \mu$.

If the Lorenz curve represented by the equation (4.3) is symmetric[3], the value of η at π and $(\sqrt{2} - \pi)$ should be equal for all values of π, which implies

[3] The symmetricity of the Lorenz curve is defined with respect to the diagonal drawn perpendicular to the egalitarian line .

$$f(\pi) = f(\sqrt{2} - \pi) \quad \text{for all } \pi \tag{4.6}$$

The curve will be skewed towards (1,1) if $f(\pi) > f(\sqrt{2} - \pi)$ for $\pi < 1/\sqrt{2}$ and it will be skewed towards (0,0) if $f(\pi) < (\sqrt{2} - \pi)$ for $\pi < 1/\sqrt{2}$. For instance, assume that the equation of the curve is

$$\eta = a\pi^{\alpha}(\sqrt{2} - \pi)^{\beta}, \quad a > 0,\ \alpha > 0, \text{ and } \beta > 0. \tag{4.7}$$

The restriction $a > 0$ implies that $\eta \geq 0$, i.e., the Lorenz curve lies below the egalitarian line. Further, $\alpha > 0$ and $\beta > 0$ mean that η assumes value zero when $\pi = 0$ or $\pi = \sqrt{2}$. Using (4.6) it is seen that the curve is symmetric if $\alpha = \beta$, skewed towards (1,1) if $\beta > \alpha$, and skewed towards (0,0) otherwise. Further restrictions on the coefficients of (4.7) can be imposed on the basis of equations (4.4) and (4.5). If $f'(\pi)$ stands for the first derivative of $f(\pi)$ with respect to π, the equation (4.4) implies that for $X \geq 0$, $[1 - f'(\pi)]$ and $[1 + f'(\pi)]$ should be of the same sign so that their ratio is always positive. The equation (4.5) means that for all values of X, the second derivative $f''(\pi)$ should be negative. For the equation (4.7), these three quantities are obtained as

$$1 - f'(\pi) = \frac{\sqrt{2}F_1 + (1-\alpha)\eta}{\pi} + \frac{\beta\eta}{\sqrt{2} - \pi}, \tag{4.8}$$

$$1 + f'(\pi) = \frac{\sqrt{2}(1-F) + (1-\beta)\eta}{(\sqrt{2} - \pi)} + \frac{\alpha\eta}{\pi}, \tag{4.9}$$

and

$$f''(\pi) = -\eta\left[\frac{\alpha(1-\alpha)}{\pi^2} + \frac{\beta(1-\beta)}{(\sqrt{2}-\pi)^2} + \frac{2\alpha\beta}{\pi(\sqrt{2}-\pi)}\right], \tag{4.10}$$

where use has been made of (4.2). It is thus obvious that the sufficient conditions for equations (4.4) and (4.5) to be satisfied are $0 < \alpha \leq 1$ and $0 < \beta \leq 1$. These sufficient conditions rule out the possibility of points of inflexion on the curve which are, of course, not permissible in the Lorenz curve.

An alternative class of equations of the Lorenz curve which look similar to the well known CES production function proposed by Arrow *et al.* (1961) is given by

$$\eta = a[\delta\pi^{-\rho} + (1-\delta)(\sqrt{2} - \pi)^{-\rho}]^{-\nu/\rho} \tag{4.11}$$

where the parameters a, δ, ρ, and ν, are all greater than zero. Rearranging the equation (4.11) we obtain

$$\eta = a\pi^{\nu}(\sqrt{2} - \pi)^{\nu}[\delta(\sqrt{2} - \pi)^{\rho} + (1-\delta)\pi^{\rho}]^{-\nu/\rho} \tag{4.12}$$

which clearly shows that η assumes value zero when $\pi = 0$ and $\pi = \sqrt{2}$. The curve is symmetric if $\delta = 1/2$, skewed towards (1,1) if $\delta > 1/2$, and skewed towards (0,0) if $\delta < 1/2$. Further, the limit of (4.12) as ρ approaches zero becomes

$$\eta = a\pi^{\delta\nu}(\sqrt{2}-\pi)^{\nu(1-\delta)} \tag{4.13}$$

which is the same class of equations as (4.7) with $\alpha = \delta\nu$ and $\beta = \nu(1-\delta)$. Finally, the sufficient conditions that the equations (4.4) and (4.5) are always satisfied for this class of equations, i.e. (4.11), are $0 < \delta < 1$ and $0 < \nu < 1$.

The income density function underlying the Lorenz curve (4.3) is obtained as

$$g(x) = \frac{1}{\sqrt{2}}\left[\frac{d\pi}{dx} + \frac{d\eta}{dx}\right] = \frac{1}{\sqrt{2}}\left[1 + \frac{d\eta}{d\pi}\right]\frac{d\pi}{dx} = \frac{\sqrt{2}\mu}{(\mu+x)}\frac{d\pi}{dx} \tag{4.14}$$

where use has been made of equations (4.2) and (4.4).

The equation (4.4), written as

$$f'(\pi) = \frac{\mu - x}{\mu + x}, \tag{4.15}$$

gives the relationship between π and x. Under the sufficient conditions discussed above, $f''(\pi) < 0$, which implies that $f'(\pi)$ is a monotonically decreasing function of π and, therefore, the equation (4.16) can always be solved for π in terms of x. Substituting the value of π for a given value of x in (4.2) gives the distribution functions $F(x)$ and $F_1(x)$. Differentiating (4.15) with respect to π gives

$$f''(\pi)\frac{d\pi}{dx} = \frac{-2\mu}{(\mu+x)^2} \tag{4.16}$$

which implies that $d\pi/dx > 0$, i.e., π increases as x increases. Using the value of π solved from (4.15) into (4.16), we obtain the value of $d\pi/dx$ in terms of x, which on substituting in (4.14) gives the density function $g(x)$. Thus, the condition that $f''(\pi) < 0$ is satisfied for the given equation for the Lorenz curve it is always possible to derive the income density function underlying the equation of the Lorenz curve.

3 Inequality Measures and Their Derivation

Among all the inequality measures, the most widely used is Gini's concentration ratio which is equal to twice the area between the Lorenz curve and the egalitarian line. Thus if the Lorenz curve is formulated in terms of π and η, the concentration ratio becomes

$$CR = 2\int_0^{\sqrt{2}} f(\pi)d\pi \tag{4.17}$$

which for the specific curve (4.7) is

$$CR = 2\int_0^{\sqrt{2}} a\pi^{\alpha}(\sqrt{2}-\pi)^{\beta}d\pi = 2a(\sqrt{2})^{1+\alpha+\beta}B(1+\alpha, 1+\beta) \tag{4.18}$$

where $B(1+\alpha, 1+\beta)$ is the Beta function which has been widely tabulated (see Pearson and Johnson (1968)).

The partial derivative of CR with respect to a, α, and β are evaluated as

$$\frac{\partial(CR)}{\partial a} = \frac{CR}{a}, \tag{4.19}$$

$$\frac{\partial(CR)}{\partial \alpha} = [\log\sqrt{2} + \Psi(1+\alpha) - \Psi(2+\alpha+\beta)](CR), \tag{4.20}$$

$$\frac{\partial(CR)}{\partial \beta} = [\log\sqrt{2} + \Psi(1+\beta) - \Psi(2+\alpha+\beta)](CR), \tag{4.21}$$

where $\Psi(1+\alpha)$ is the Euler's psi function which can be numerically computed by making use of the following relationship:[4]

$$\Psi(1+\alpha) - \Psi(2+\alpha+\beta) = \sum_{k=0}^{\infty}\left(\frac{1}{(2+\alpha+\beta+k)} - \frac{1}{(1+\alpha+k)}\right). \tag{4.22}$$

Using these partial derivatives, the asymptotic variance of CR can now be obtained from the estimated variances and covariances of the parameter estimates $\hat{a}$, $\hat{\alpha}$, and $\hat{\beta}$ (see Kakwani and Podder (1972)).

Another important measure of inequality which is well known in the literature is relative mean deviation. This measure is defined as

$$T = \frac{1}{2\mu}\frac{1}{N}\sum_{i=1}^{N}|x_i - \mu| \tag{4.23}$$

where x_i is the income of the ith family.

It can be shown (see Gastwirth (1972)) that T is equal to the maximum discrepancy between $F(x)$ and $F_1(x)$, which is also equal to $\sqrt{2}$ times the maximum value of η. In order to obtain the maximum value of η, equation (4.3) is to be differentiated with respect to π and equated to zero. Then solving for π, the maximum value of η can be obtained from the equation of the Lorenz curve. For instance, if the equation of the Lorenz curve is (4.7), equating its derivative to zero, we obtain

$$\frac{d\eta}{d\pi} = a\alpha\pi^{\alpha-1}(\sqrt{2}-\pi)^{\beta} - a\beta\pi^{\alpha}(\sqrt{2}-\pi)^{\beta-1} = 0 \tag{4.24}$$

which gives $\pi = \sqrt{2}\alpha/(\alpha+\beta)$ and, therefore, the relative mean deviation will be

$$T = (\sqrt{2})^{1+\alpha+\beta}\frac{a\alpha^{\alpha}\beta^{\beta}}{(\alpha+\beta)^{\alpha+\beta}}. \tag{4.25}$$

[4] In order to find the derivatives $\partial(CR)/\partial\alpha$ and $\partial(CR)/\partial\beta$, we require the partial derivatives of $B(1+\alpha, 1+\beta)$ with respect to α and β. Formula 4-2531 of Gradsheteyn and Ryshik (1965) is used to evaluate the integral obtained after differentiating partially the Beta function.

Again if the variances and covariances of $\hat{a}$, $\hat{\alpha}$, and $\hat{\beta}$ are known, it is possible to compute the asymptotic variance of T.

Elteto and Frigyes (1968) have recently proposed a set of three new inequality measures which can be easily computed from the equation of the Lorenz curve (4.3) by using the value of π at which η is maximum. It is thus obvious that the derivation of Elteto and Frigyes (1968) measures are similar to the relative mean deviation. Recently, Kondor (1971) has shown that these measures do not convey much more information than the relative mean deviation and it is, therefore, unnecessary to discuss their derivation here. However, the numerical values of these measures along with their asymptotic standard errors have been computed using Australian data in Section 5.

Further, the estimated Lorenz curve (4.3) can be used to obtain any percentile of the distribution. To illustrate this point the estimated shares of income going to the poorest and richest 5 and 10 per cent have been computed in Section 5.

4 The Estimation of the Lorenz Curve

The estimation of the Lorenz curve from grouped observations is considered here. Suppose there are N families which have been grouped into $(T+1)$ income classes, viz., $(0 \text{ to } x_1), (x_1 \text{ to } x_2), \ldots, (x_T \text{ to } x_{T+1})$. Let n_t, be the number of families earning income in the interval x_{t-1} and x_t; then $f_t = n_t/N$ is the relative frequency; $f_t = n_t/N$ is a consistent estimator of the probability ϕ_t of a family belonging to the tth income group.[5]

If x_t^* is the sample mean for the tth income group, then the consistent estimates of $F(x_t)$ and $F_1(x_t)$ are

$$p_t = \sum_{\gamma=1}^{t} f_\gamma \quad \text{and} \quad q_t = \frac{1}{Q}\sum_{\gamma=1}^{t} x_\gamma^* f_\gamma, \tag{4.26}$$

respectively, where $t = 1, 2, \ldots, T$ and $Q = \sum_{\gamma=1}^{T+1} x_t^* f_\gamma$ is the mean income of all the families. Now using the equation (4.2), the consistent estimators of π_t and η_t are obtained as

$$r_t = \frac{p_t + q_t}{\sqrt{2}} \quad \text{and} \quad y_t = \frac{p_t - q_t}{\sqrt{2}}, \tag{4.27}$$

respectively. (The terms r_t, and y_t, differ from π_t, and η_t, by some random disturbance terms.) Then the equation of the Lorenz curve (4.7) in terms of the observations on r_t, and y_t can be written as

$$\log\, y_t = a' + \alpha \log\, r_t + \beta \log(\sqrt{2} - r_t) + w_{1t}, \tag{4.28}$$

[5] $\phi_t = \int_{x_{t-1}}^{x_t} g(x)dx$.

where $a' = \log a$ and w_{1t} is the random disturbance which can be shown to be of order $N^{-1/2}$ in probability (see Kakwani and Podder (1972)).

In what follows, it will be useful to write the above equation in vector and matrix notations as

$$Y_1 = X_1\delta + w_1, \tag{4.29}$$

where Y_1 is a $T\times 1$ vector of T observations on log y_1,X_1, is a $T\times 3$ matrix of T observations on the right-hand side variables of (4.28), w_1 is the column vector of T observations on the disturbance term, and δ is the column vector consisting of the three elements a', α, and β. Then the least squares estimator of δ is

$$\hat{\delta} = (X_1'X_1)^{-1}X_1'Y_1 \tag{4.30}$$

which will be referred to as Method I in subsequent discussions.

Following Kakwani and Podder (1972) it can be shown that $\hat{\delta}$ is a consistent estimator of δ and its asymptotic variance-covariance matrix is given by

$$\mathrm{var}(\hat{\delta}) = (\mathrm{X}_1'\mathrm{X}_1)^{-1}\mathrm{X}_1'\Omega_{11}\mathrm{X}_1(\mathrm{X}_1'\mathrm{X}_1)^{-1} \tag{4.31}$$

where

$$E(w_1w_1') = \Omega_{11} \tag{4.32}$$

is the variance and covariance matrix of w_1. However, the asymptotically more efficient estimator of δ is

$$\hat{\hat{\delta}} = (X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1}Y_1, \tag{4.33}$$

which can also be shown to be consistent and its asymptotic variance-covariance matrix would be

$$var(\hat{\hat{\delta}}) = (X_1'\Omega_{11}^{-1}X_1)^{-1}. \tag{4.34}$$

This generalized least squares method will be referred to as Method II. The information on income ranges is available for most income distributions which can be effectively utilized to improve the precision of the estimates. To show this, we consider the equation (4.4) which for the Lorenz curve (4.7) can be written as

$$\frac{\mu - x_t}{\mu + x_t}\frac{\pi_t(\sqrt{2}-\pi_t)}{\eta_t} = (\sqrt{2}-\pi_t)\alpha - \pi_t\beta. \tag{4.35}$$

Substituting the estimates of π_t, η_t, and μ, the above equation becomes

$$\frac{Q - x_t}{Q + x_t}\frac{r_t(\sqrt{2}-r_t)}{y_t} = (\sqrt{2}-r_t)\alpha - r_t\beta + w_{2t} \tag{4.36}$$

where w_{2t}, is the random error which can again be shown to be of order $N^{-1/2}$ in probability.

Write (4.36) in vector and matrix notations as

$$Y_2 = X_2\delta + w_2, \tag{4.37}$$

where Y_2 is a column vector of T observations on the dependent variable in the equation (4.36), X_2 is a $T \times 3$ matrix, the first column of which consists of T observation on the explanatory variables $(\sqrt{2} - r_t)$ and $-r_t$, of the equation (4.36), and w_2 is the vector of stochastic disturbances.

The equations (4.29) and (4.37) can now be combined together as

$$Y = X\delta + w, \tag{4.38}$$

where

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \ X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \text{ and } w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{4.39}$$

w is now the vector of $2T$ disturbances with zero mean and covariance matrix

$$Eww' = \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}. \tag{4.40}$$

The coefficient vector δ can now be estimated from (4.38) by the direct least squares method which will be referred to as Method III. However, the asymptotic more efficient estimator of δ will be

$$\hat{\hat{\delta}}^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \tag{4.41}$$

with its asymptotic variance-covariance matrix

$$var(\hat{\hat{\delta}}^*) = (X'\Omega^{-1}X)^{-1}. \tag{4.42}$$

The estimator $\hat{\hat{\delta}}$ will be referred to as Method IV.[6]

The above procedure of estimating the parameters by combining two stochastic equations has been earlier used by Theil (1963) in connection with mixed estimation and by Zellner (1962) in connection with seemingly unrelated regressions. It can be demonstrated that the estimators of the coefficient vector δ obtained from the combined equation are more efficient than those obtained from the individual equation (4.29).

5 Some Empirical Results

Results of the estimation of the Lorenz curve and associated inequality measures are presented in this section. The source of data used for this purpose is the Australian Survey of Consumer Finances and Expenditures, 1967-68, carried out by the Macquarie University and the University of Queensland. The nature of the Survey has

[6] The derivation of the consistent estimator of the covariance matrix Ω follows a pattern similar to that given in the earlier work of the authors Kakwani and Podder (1972) and, therefore, has been omitted here.

been extensively discussed elsewhere (Podder (1972)). The data were supplied to us in the form of a set of individual observations. The income considered here is net of taxes but does not include imputed rent from owner occupied houses. Individual income figures made it possible to compute the actual values of the concentration ratio and other measures which were useful in judging the degree of accuracy of the methods discussed in this paper. The grouped data are presented in Table 4.1.

Table 4.1: Income Distribution

Income Range ($)	*Number of families*	*Mean Income ($)*	*p*	*q*	*r*	*y*
Below 1000	310	674.39	0.05698	0.009274	0.046849	0.033730
1000-1999	552	1426.10	0.15846	0.044193	0.143297	0.080799
2000-2999	1007	2545.79	0.34357	0.157912	0.354601	0.131280
3000-3999	1193	3469.35	0.56287	0.341510	0.639493	0.156525
4000-4999	884	4470.33	0.72536	0.516805	0.878343	0.147471
5000-5999	608	5446.60	0.83713	0.663701	1.061248	0.122633
6000-6999	314	6460.93	0.89485	0.753693	1.165696	0.099813
7000-7999	222	7459.14	0.93566	0.827147	1.246493	0.076730
8000-8999	128	8456.66	0.95919	0.875164	1.297084	0.059415
9000-10999	112	9788.38	0.97978	0.923794	1.346030	0.039588
11000 and over	110	15617.69	1.00000	1.000000	1.414213	0.000000

Table 4.2 presents the estimated parameters of the Lorenz curve (4.7) along with the different inequality measures. The equation has been estimated using four alternative methods of estimation discussed in Section 4.

In the table u', v', and w' represent the three new inequality measures proposed by Elteto and Frigyes (1968). The last row of the table presents the actual values of the inequality measures computed on the basis of the individual observations. It is observed that the inequality measures computed by all four methods are very close to the actual values and their standard errors are generally low. Method IV gives the best result in the sense that the standard errors are the lowest and the estimated inequality measures are closest to their actual values.

It should be pointed out that various approximations have been used in the past to estimate the concentration ratio from grouped data. The most common is the linear approximation which assumes that within each income range the inequality is zero (see Morgan (1962)). Therefore this approximation provides only the lower limit of the *CR* which in the present case is 0.3134. Gastwirth (1972) has recently suggested the method of obtaining the upper limit of the *CR* which has been computed as 0.3223 in our case. The *CR* computed in Table 4.3 lies within the lower and upper limit for all four methods of estimation. Thus the goodness-of-fit test criterion

Table 4.2: Results of the Different Methods of Estimation

Method of Estimation	*Coefficients Estimates*	Inequality Measures *CR*	*T*	*u'*	*v'*	*w*
I	α=0.7542 (0.0061) β=0.8042 (0.0072) *a*=0.2728 (0.0156)	0.3203 (0.0041)	0.2250 (0.0081)	0.3772 (0.0051)	0.6002 (0.0015)	0.3579 (0.0048)
II	α=0.7535 (0.0059) β=0.8029 (0.0067) *a*=0.2730 (0.0132)	0.3208 (0.0039)	0.2252 (0.0070)	0.3774 (0.0043)	0.6004 (0.0013)	0.3583 (0.0038)
III	α=0.7615 (0.0042) β=0.8061 (0.0058) a=0.2744 (0.0121)	0.3206 (0.0038)	0.2255 (0.0022)	0.3767 (0.0041)	0.6009 (0.0009)	0.3596 (0.0037)
VI	α=0.7611 (0.0009) β=0.8049 (0.0020) *a*=0.2732 (0.0091)	0.3195 (0.0008)	0.2246 (0.0018)	0.3753 (0.0009)	0.5993 (0.0008)	0.3583 (0.0029)
	Actual Values of Inequality Measures	0.3196	0.2219	0.3750	0.5951	0.3521

NOTE: Figures in brackets are the asymptotic standard errors

Table 4.3: Actual and Estimated *y*

	Methods of Estimation			
Actual y	I	II	III	IV
0.0337	0.0349	0.0349	0.0343	0.0342
0.0808	0.0764	0.0766	0.0758	0.0755
0.1313	0.1307	0.1310	0.1306	0.1300
0.1565	0.1586	0.1588	0.1589	0.1583
0.1475	0.1498	0.1500	0.1503	0.1498
0.1226	0.1235	0.1237	0.1240	0.1236
0.0998	0.0999	0.1002	0.1004	0.1001
0.0767	0.0766	0.0769	0.0769	0.0768
0.0594	0.0592	0.0593	0.0594	0.0593
0.0396	0.0394	0.0395	0.0395	0.0395

Table 4.4: Actual and Estimated Frequency Distribution of Family Income

	Relative Frequency		Mean Income	
Income Range	*Actual*	*Estimated*	*Actual*	*Estimated*
Below 1000	0.0569	0.0403	674.39	380.47
1000-1999	0.1015	0.1188	1426.10	1583.66
2000-2999	0.1851	0.2089	2545.79	2515.38
3000-3999	0.2193	0.2029	3469.35	3482.29
4000-4999	0.1625	0.1527	4470.33	4469.68
5000-5999	0.1118	0.1033	5446.60	5463.85
6000-6999	0.0577	0.0661	6460.93	6451.13
7000-7999	0.0408	0.0406	7459.14	7471.49
8000-8999	0.0235	0.0245	8456.66	8440.27
9000-10999	0.0206	0.0236	9788.38	9868.39
11000+	0.0203	0.0183	15617.69	15964.67

proposed by Gastwirth and Smith (1972) is satisfied for the density function fitted here.[7]

The equation of the curve (4.28) was also fitted to 20 income groups earlier considered by the authors (Kakwani and Podder (1972)) using Method I of estimation. The *CR* and *T* have been computed as 0.3201 and 0.3241 respectively. Thus increasing the number of income groups from 11 to 20 improves the accuracy of the technique.

Table 4.5: Shares of Incomes: The poorest and Richest 5 and 10 per Cent

Shares of Income	*Estimated from 11 Groups*	*Estimated from 20 Groups*	*Actual from Individual Observations*
Poorest 5%	0.5857	0.623	0.767
Poorest 10%	2.2600	2.314	2.132
Richest 5%	14.2000	14.380	14.424
Richest 10%	23.6000	23.780	23.757

In order to obtain the estimated frequency distribution of the income we need to solve for π in terms of x from the following non-linear equation:

$$\alpha a\pi^{\alpha-1}(\sqrt{2}-\pi)^{\beta} - \beta a\pi^{\alpha}(\sqrt{2}-\pi)^{\beta-1} = \frac{\mu - x}{\mu + x}, \tag{4.43}$$

where the estimates of a, α, and β are given in Table 4.2. The Newton-Raphson method was used to compute π for given values of x (see Henrici (1967)). The estimated frequency distribution for the family income so obtained is given in Table 4.4.

[7] This goodness-of-fit test is based on the idea that any fitted distribution whose theoretical Gini index falls outside the lower and upper bounds should be declared to fit the data inadequately. According to this test the lognormal distribution did not fit United States data.

It can be concluded from this table that the density function underlying the Lorenz curve (4.7) provides a reasonably good fit to the whole range of the observed income distribution.

The estimated shares of incomes going to the poorest and richest 5 and 10 per cent are presented in Table 4.5. It is clear from the table that the estimated income shares are quite close to the actual based on the individual observations. The twenty income groups again provide more accurate results than ten groups.

The above procedure can be used to obtain the relative frequency and the mean income of income ranges which could be made as small as one wishes. By dividing the whole income range into a large number of income groups it is thus possible to compute accurately a number of inequality measures which could not be otherwise obtained from group observations. In addition, the density function can be useful for other purposes which need not be mentioned here.

References

Arrow, K. J., H. B. Chenery, B. S. Minhas and R. M. Solow (1961) Capital-Labor Substitution and Economic Efficiency, *Review of Economics and Statistics*, **43**, 225–250.

Elteto, O. and E. Frigyes (1968) New Income Inequality Measures as Efficient Tools for Causal Analysis and Planning, *Econometrica*, **36**, 383–396.

Gastwirth, J. L. (1972) The Estimation of the Lorenz Curve and GINI Index, *The Review of Economics and Statistics*, **54**, 306–316.

Gastwirth, J. L. and J. T. Smith (1972) A New Goodness-Of-Fit Test, *Proceedings of the American Statistical Association*, pp. 320–322.

Gradsheteyn, I. S. and I. M. Ryshik (1965) *Tables of Integral Series and Products*, Academic Press, New York.

Henrici, P. (1967) *Elements of Numerical Analysis*, John Wiley and Sons Ltd, New York.

Kakwani, N. C. and N. Podder (1972) On the Estimation of Lorenz Curves from Grouped Observations, *International Economic Review*, **14**, 278–291.

Kondor, Y. (1971) An Old-New Measure of Income Inequality, *Econometrica*, **39**, 1041–1042.

Morgan, J. (1962) The Anatomy of Income Distribution, *The Review of Economics and Statistics*, **44**, 270–283.

Pearson, E. S. and N. L. Johnson (1968) *Tables of the Incomplete Beta-Function*, The Cambridge University Press, Cambridge.

Podder, N. (1972) Distribution of Household Income in Australia, *The Economic Record*, **48**, 181–201.

Theil, H. (1963) On the Use of Incomplete Prior Information in Regression Analysis, *Journal of the American Statistical Association*, **58**, 401–414.

Zellner, A. (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Testsfor Aggregation Bias, *Journal of the American Statistical Association*, **57**, 348–368.

CHAPTER 5

Distribution and Mobility of Wealth of Nations*

Richard Paap† and Herman K. van Dijk‡

Abstract

We estimate the empirical bimodal cross-section distribution of real Gross Domestic Product per capita of 120 countries over the period 1960–1989 by a mixture of a Weibull and a truncated normal density. The components of the mixture represent a group of poor and a group of rich countries, while the mixing proportion describes the distribution over poor and rich. This enables us to analyse the development of the mean and variance of both groups separately and the switches of countries between the two groups over time. Empirical evidence indicates that the means of the two groups are diverging in terms of levels, but that the growth rates of the means of the two groups over the period 1960–1989 are the same.

1 Introduction

Empirical evidence on convergence of national economies has usually been investigated by regressing growth rates of real Gross Domestic Product [GDP] on initial levels, sometimes after correcting for exogenous variables (conditional convergence), see among others, Baumol (1986), Barro (1991), Mankiw, Romer and Weil (1992) and Sala-i-Martin (1994). A negative regression coefficient, usually labelled the β-coefficient, is interpreted as an indication of so called β-convergence. It implies that countries with a relatively low level of GDP grow faster than countries with a high level of GDP, indicating catching-up, compare also Abramowitz (1986).

---

* Reprint of Richard Paap and Herman K. van Dijk (1998) Distribution and Mobility of Wealth of Nations, *European Economic Review*, 42, 1269-1293. Copyright (1998), with permission from Elsevier.

† Econometric and Tinbergen Institutes, Erasmus University Rotterdam.

‡ Econometric and Tinbergen Institutes, Erasmus University Rotterdam.
Email: hkvandijk@few.eur.nl

A different concept of convergence, called σ-convergence, refers to a reduction in cross-sectional variance or dispersion over time, see Barro and Sala-i-Martin (1992). Friedman (1992) and especially Quah (1993a) show, using Galton's fallacy, that a negative β-regression coefficient can be perfectly consistent with the absence of σ-convergence, even when conditioning on exogenous variables. Furthermore, Levine and Renelt (1991, 1992) discuss the robustness of the regression approach with respect to the conditioning variables and the consistency of the results, see also Durlauf and Johnson (1995). Another limitation of the regression approach is that the dynamics of the economic process is summarised in a growth rate and an initial level, neglecting the short run dynamics of the variable investigated.

This paper deals with the analysis of convergence in terms of several characteristics of the distribution of real GDP per capita and is related to the work by Quah (1993a,b), Desdoigts (1994), Jones (1997), Quah (1996a,b) and Bianchi (1997). In these studies nonparametric methods are usually applied to analyse convergence. In the present paper we take a parametric approach. More, generally, we analyse the *development of the distribution and mobility of wealth* of 120 countries from 1960 until 1989. As measure for wealth we take the real Gross Domestic Product per capita, which can be interpreted as a rough approximation of the basic idea about wealth, see Parente and Prescott (1993)[10]. We start with presenting some stylized facts on the observed real GDP per capita over the period. This leads to the conclusion that the data may be described by a bimodal distribution. Next, we divide the further analysis into two parts[11]. In the first part, the empirical cross sectional bimodal distribution of the real GDP per capita in each year is described by a finite mixture density. Efficient estimation of the parameters of several classes of finite mixtures results in a partitioning of the countries into two groups in each year, a group with a relatively high level of real GDP per capita and a group with a low level of real GDP per capita and two estimated conditional density functions for the two groups. The use of mixtures enables us to analyse the distribution of countries over poor and rich as well as the development of the distribution of each group.

In the second part, the results of the estimated mixture distributions are used to consider the intra-distribution dynamics. By examining the movements of countries between the poor group and the rich group, we obtain insight into the extent of catching-up of poor countries with rich countries.

The outline of the paper is as follows. In section 2 we describe the data and present some stylized facts. In section 3 we briefly discuss the interpretation, representation and estimation of finite mixture distributions. Section 4 considers the estimation results of the mixture distribution for the cross-section real GDP per capita distribution including the development of the mean, the variance and the mixing parameter through time. The mobility in wealth between and within both groups is investigated in section 5. The final section contains our conclusions.

[10] Of course, the real GDP per capita of a country is a measure which neglects information about the spread of wealth among people living in this country. There can be a small group of persons living in a country with a high level of income, while the majority has low income.

[11] Here we differ from Quah who considers the year by year distribution and intra-distribution dynamics simultaneously.

2 Stylized Facts

In order to analyse the distribution and mobility of wealth of nations empirically, one needs a suitable data set containing per capita data over a long period for a large number of countries. Usually, one has data over several periods (years) but only a limited number of (industrialized) countries or one has many countries over a small number of years. In this paper we analyse the distribution and mobility of wealth using a reasonably large collection of countries over 30 years. The obvious data set for our analysis is the Penn World Table version 5.6 of Summers and Heston (1991). This table contains a set of economic time series, based on national accounts covering 152 countries for the period 1950–1992. Because observations are not available for each country over the whole period, we focus on the period 1960–1989. By restricting ourselves to this period, there remain observations for 120 countries. The variable we analyse in this paper is the real Gross Domestic Product [GDP] per capita, which is constructed by dividing nominal GDP per capita by a special price index made up of the weighted averages across countries of relative prices of all goods in a particular basket of final goods and services. This is intended to make real GDP per capita comparable across time and countries. For a discussion of the construction of the special price index and the data in general, we refer to Summers and Heston (1991).

Figure 5.1 shows smoothed versions of histograms for real GDP per capita of 120 countries in each year[12]. Several features of the data are shown in this figure. First, the cross-section distribution of the real GDP per capita is bimodal. There is a group of countries with a relative small real GDP per capita (poor countries) and a smaller group of countries with a relative large real GDP per capita (rich countries). Second, the gap between these groups seems to become larger over time, as the peak of the real GDP per capita of the rich countries shifts more to the right than the peak of the poor countries, leaving very few countries in a middle group.

In order to obtain better insight into the stylized facts of our data set we divide our sample into six subperiods of five years and compute the average real GDP per capita for all 120 countries over these subperiods, *i.e.* for 1960–1964, 1965–1969, 1970–1974, 1975–1979, 1980–1984 and 1985–1989. Figure 5.2 displays the histograms for the mean real GDP per capita in each subperiod in a 3-dimensional space, similar to figure 5.1. This figure shows the data features mentioned before even more clearly. In addition, we notice that the variance of the poor group in the early sixties seems to be smaller than in the early eighties. For the rich countries this seems to be the opposite. The same features of the data can be detected from figure 5.3 which shows the histograms of the real GDP per capita in the six subperiods in a one-dimensional setting. The six histograms give good insight in the development of the cross-section distribution of the real GDP per capita. From the stylized facts

[12] This figure is constructed by making a histogram for real GDP per capita in each year and putting these histograms in a 3-dimensional space. For visual convenience we use small ribbons, which connect the midpoints of the bars, instead of 3-dimensional bars. Furthermore, the real GDP per capita data, are divided by 1000 for the convenience of representation, like in the remainder of this paper.

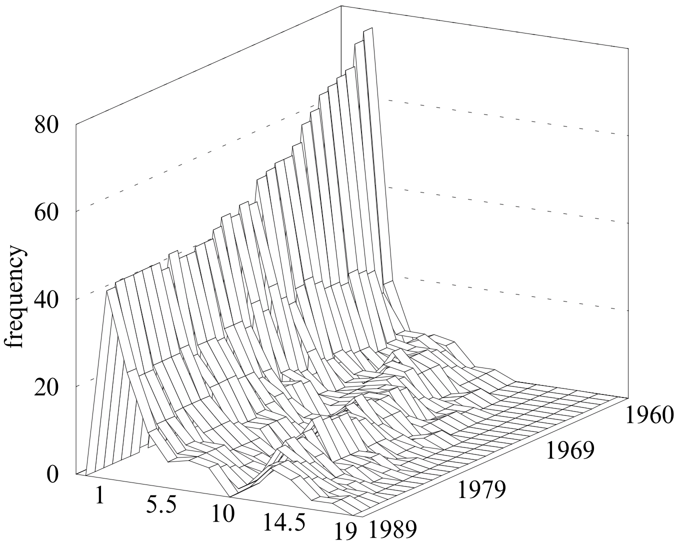


Fig. 5.1: Histograms of real GDP per capita divided by 1000 (1960–1984).

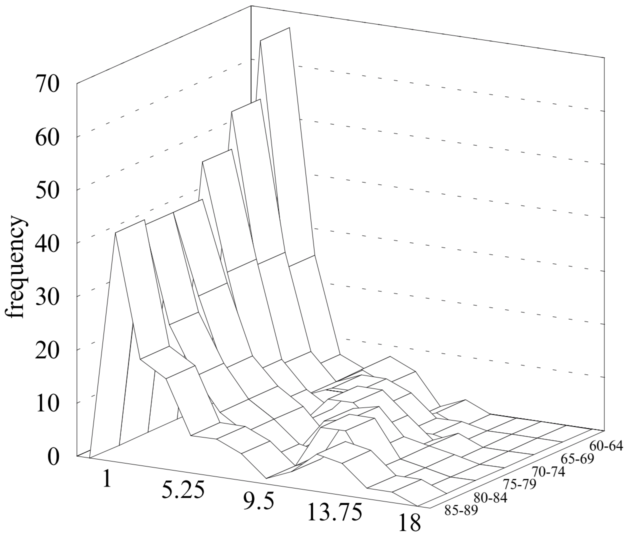


Fig. 5.2: Histograms of the average real GDP per capita divided by 1000 in six subperiods.

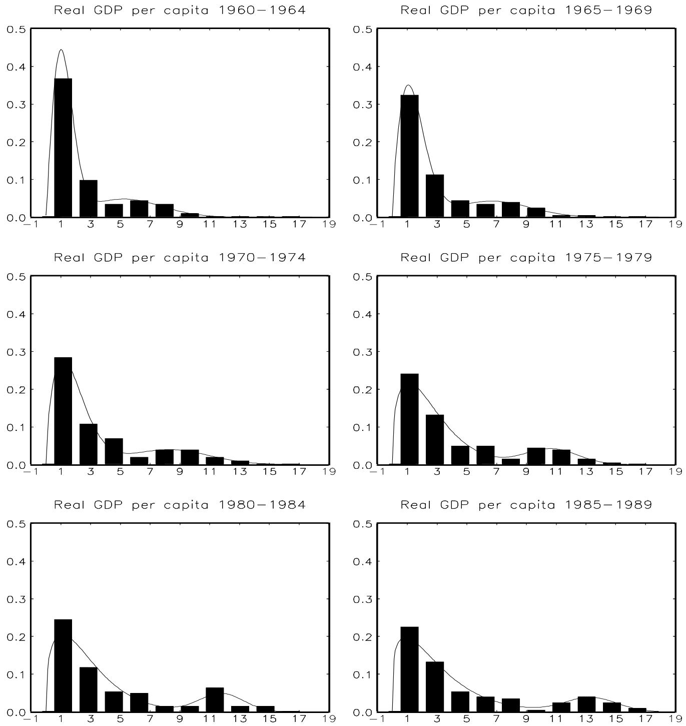


Fig. 5.3: Histograms of the average real GDP per capita divided by 1000 with the fitted density functions.

we conclude that: the distribution in each period is bimodal; a gap arises between the poor and rich group, which increases over time; the number of countries with an extremely low real GDP per capita decreases, but the spread of wealth within the poor group seems to rise. Similar findings are reported in *e.g.* Quah (1993a,b), Bianchi (1997) and Quah (1996a,b).

We end this section with three remarks. First on the loss of individual information through our histogram analysis, we note that a data summarization of 3600 individual observations into 30 yearly histograms - with only a relatively small number of cells - involves some loss of individual information. The optimal level of aggregation of information depends on the purpose of the empirical analysis. We are interested in describing and estimating efficiently such stylized facts as the behaviour of the poor and rich countries and their relative position through the post-war period. From the data summarization presented in this section we conclude as main stylized fact the bimodality of the empirical distribution of real GDP per capita.

Second on the relative merits of parametric and non-parametric analysis of income distributions, we note that we estimate the bimodal cross-section distribution of real GDP per capita per year by means of a mixture of two densities using individual observations per country. A mixture density belongs to a parametric class of densities which are defined as a convex combination of two or more densities. In our case these densities describe the distribution of the poor and the distribution of the rich countries, with a mixing distribution, representing the distribution over poor and rich. The separate analysis of the components of the mixture and of the relative importance of these components over time are the main advantages over a non-parametric approach as performed by, for instance, Desdoigt (1994). A clear choice between a parametric or a non-parametric approach depends on the availability of large data sets and on the purpose of the analysis. If there are many data over a long period then the asymptotically valid non-parametric approach is attractive in the sense that one can let the data ‘speak for themselves’. Often in economics there are not enough data to have a reliable non-parametric analysis. The parametric analysis is attractive in case there are no overly restrictive assumptions. In the next section we perform a sensitivity analysis with respect to the chosen functional form of the components of the mixture. One might also discuss the proper number of components in the mixture. Our choice of bandwidth and therefore the number of classes in the histograms are to some extent arbitrary. Using a different bandwidth in the histograms may result in the conjecture of more than two modes in the cross-section distribution. It is difficult to estimates a component of a mixture if the number of observations belonging to the components is very small, see also section 3 for a discussion about singularities in the likelihood function. Furthermore, the extra modes which occur using a smaller bandwidth, may also be due to noise. Bianchi (1997) rejects the hypothesis of more than two modes using a non-parametric approach based on the choice of the bandwidth. This supports our choice of two components in the mixture.

Third, on the choice between level, log of the level and relative level of real GDP per capita we note that in this paper we are interested in investigating convergence in the level of real GDP per capita. That is, that convergence implies that the differences in the level of real GDP per capita between countries disappear. As a byproduct we test in section 4 whether the growth rates of the the rich and the poor group of countries are the same. Another option is to scale the data by the sum of the real GDP per capita in each year as suggested by Canova and Marcet (1995) or to analyse log transformed data to test for convergence in relative welfare. In section 3 we show that our analysis is not sensitive to scaling the data in each year by a constant. A log transformation makes the data more homogenous and the evidence of bimodality in the data is considerably reduced, see Bianchi (1997). Homogeneity of the data is an attractive feature if one has to meet the assumptions of classical regression models, *e.g.* when testing for β-convergence. Also, one may use data on real GDP per worker instead of real GDP per capita in order to analyse convergence in productivity. In the present paper, we have chosen to focus on testing for convergence in the level of real GDP per capita.

3 Finite Mixture Distributions

We briefly discuss the representation, interpretation and estimation of mixtures distributions. For a good introductory survey of finite mixture distribution reference is made to Everitt and Hand (1981) or Titterington, Smith and Makov (1985). For our purpose it suffices to restrict ourselves to finite mixtures with a multinomial mixing distribution. In this case, the mixture density function g is defined as

$$g(y;\theta_1,...\theta_S,\lambda_1,...\lambda_{S-1}) = \sum_{s=1}^{S}\lambda_s f(y;\theta_s) \text{ with } \lambda_S = 1-\sum_{s=1}^{S-1}\lambda_s, \tag{5.1}$$

where S denotes the number of components in the mixture; $f(y;\theta_s)$, $s=1,...,S$ are probability density functions evaluated at y depending on a parameter vector θ_s; and λ_s, $s=1,...,S-1$ represent the mixing proportions. An example of a finite mixture distribution is a mixture of two normal distributions. The density function g evaluated at y_i is given by

$$g(y_i;\theta_1,\theta_2,\lambda) = \frac{\lambda}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{(y_i-\mu_1)^2}{2\sigma_1^2}\right) + \frac{(1-\lambda)}{\sigma_2\sqrt{2\pi}}\exp\left(-\frac{(y_i-\mu_2)^2}{2\sigma_2^2}\right), \tag{5.2}$$

where $\theta_1 = \{\mu_1,\sigma_1^2\}$ and $\theta_2 = \{\mu_2,\sigma_2^2\}$ denote the mean and the variance of the normal distribution of each component and λ represents the mixing proportion. For suitable chosen parameters, this mixture distribution is bimodal[13].

Interpretation

Representing the bimodal distribution of the data by a mixture of two densities is a convenient and interpretable way of describing the real GDP per capita. The distribution of the real GDP per capita of the poor countries is described by the first component of the mixture and the distribution of the rich countries by the second component. The mixing parameter λ gives the ex-ante probability that a country belongs to the first component of the mixture. Formally, the probability density function for the real GDP per capita for country i, denoted by y_i for $i=1,..,N$ can be written as

$$g(y_i;\theta_1,\theta_2,\lambda) = \lambda f(y_i \mid s_i = 1;\theta_{s_i}) + (1-\lambda)f(y_i \mid s_i = 2;\theta_{s_i}), \tag{5.3}$$

[13] A sufficient condition that a value λ exists such that the mixture of two normal distributions is bimodal is $(\mu_2-\mu_1)^2 < (8\sigma_1^2\sigma_2^2)/(\sigma_1^2+\sigma_2^2)$.

where $\lambda = P[s_i = 1]$ and $1-\lambda = P[s_i = 2]$ are the ex-ante probabilities that country i is poor or rich and where $f(y_i \mid s_i = 1; \theta_{s_i})$ and $f(y_i \mid s_i = 2; \theta_{s_i})$ are conditional probability density functions given that country i is poor or rich. The mean and variance of the conditional distribution of component s can be interpreted as the mean and the variance of the real GDP per capita of countries belonging to component s.

An attractive feature of our approach is that the mixing parameter λ is an *endogenous* parameter which determines the relative importance of each component in the mixture distribution.[14] So, *a priori* we do not impose an absolute borderline between the rich and the poor countries but let the data determine the relative importance of each group. One may interpret a mixture model as an unobserved component model in the following sense. To generate an observation y_i from a mixture, a country is selected to be poor with probability λ or to be rich with probability $(1-\lambda)$, or in other words the value of s_i is determined. Given that the country is poor the value of the real GDP per capita, y_i is generated by the conditional density function $f(y_i \mid s_i = 1; \theta_{s_i})$ (or $f(y_i \mid s_i = 2; \theta_{s_i})$ in case the country is rich). However, we only observe the value of the real GDP per capita y_i and not the value of s_i. Given the realized value of the real GDP per capita y_i and given the values of the parameters θ_1, θ_2 and λ, we can make inference about the value of s_i. The conditional probability that observation y_i is generated by the first component ($s_i = 1$) for the mixture defined in (5.3) is defined as

$$\Pr[s_i = 1 \mid y_i; \theta_1, \theta_2, \lambda] = \frac{\lambda f(y_i \mid s_i = 1; \theta_{s_i})}{\lambda f(y_i \mid s_i = 1; \theta_{s_i}) + (1-\lambda) f(y_i \mid s_i = 2; \theta_{s_i})} \tag{5.4}$$

This conditional probability denotes the ex-post probability that a country is poor and is used for the investigation of mobility in wealth in section 5. Note that the ex-post probability of being rich $\Pr[s_i = 2 \mid y_i; \theta_1, \theta_2, \lambda]$ equals $1 - \Pr[s_i = 1 \mid y_i; \theta_1, \theta_2, \lambda]$ by definition.

In practice, we do not know the true values of the parameters θ_1, θ_2, λ and we have to replace them by their estimates. The estimated λ can be interpreted as the proportion of countries belonging to the first component, *i.e.* the percentage of poor countries, while the probability in (5.4) can be seen as the relative ex-post contribution of country i to the first component. (Note that in case of a mixture of normal densities the estimated mean $\hat{\mu}_1 = \frac{1}{\lambda N}\sum_{i=1}^{N} \Pr[s_1 = 1 \mid y_i; \hat{\theta}_1, \hat{\theta}_2 \hat{\lambda}]\, y_i$, *i.e.* a weighted average of the observations.)

Since countries can switch over time from being poor to being rich and *vice versa* the mixing proportion λ can change through time. The growth in real GDP per capita causes changes in the means of the mixture components through time. Further, countries belonging to a group do not need to have the same growth rates, which implies that the variance does not have to be the same over time. Note that a change in the mean and/or the variance of a component can also be caused by movement of countries between the rich and the poor group.

[14] Durlauf and Johnson (1995) use the regression tree technique to endogenously split the data in multiple regimes.

Estimation

Several methods have been proposed to estimate the parameters of a mixture, *e.g.* maximum likelihood and the methods of moments, see Everitt and Hand (1981). We follow the maximum likelihood approach, which implies maximising the following criterion function

$$\mathscr{L}(Y;\theta_1,\theta_2,\lambda) = \prod_{i=1}^{N} g(y_i;\theta_1,\theta_2,\lambda), \tag{5.5}$$

where the density function g is given by (5.2) and $Y = \{y_1, ..., y_N\}$. From the first order conditions, it is easy to see that maximising the likelihood implies a non-linear optimisation problem. Standard numerical algorithms can be use to maximise the likelihood function. Note that the likelihood function (5.5) for estimation of a mixture of normal densities (5.2) has not a global maximum, since a singularity in the likelihood function arises, whenever one of the components is imputed to have a mean equal to one of the observations (μ_1 equals y_i) with zero variance ($\sigma_1^2 \rightarrow 0$). At that point the value of the likelihood function becomes infinite. Kiefer (1978) shows that if there exists a local maximum in the interior of the parameter region then this maximum yields consistent, asymptotically normal estimators of the parameters. In this case, the ML estimators are not values of the parameters which maximize the likelihood function globally, but are those solutions of the likelihood equations, which yields asymptotically the largest value of the likelihood function. In practice, if a numerical optimisation algorithm gets "stuck" at a singularity, the easiest strategy is to try a different starting value. Another solution is to use a quasi Bayesian approach by multiplying the likelihood function by a prior density to cancel out the singularity problem (see Hamilton, 1991).

A mixture of two normal densities does not suffice to describe our bimodal distributions. It is clear from figure 5.3 that the first component of the mixture distribution is skew. Another point is that real GDP per capita can never be negative, so a mixture of normal densities is, strictly speaking, not appropriate. Possible candidates to describe the distribution of the poor countries (first component) are *e.g.* the Weibull distribution, the gamma distribution and the lognormal distribution. For the distribution of the rich countries a normal distribution (truncated at 0) seems appropriate.

We have estimated several combinations of the proposed distributions and compared the fit to select the best candidates. To analyse the fit of these distribution we divide the data in each of the six subperiods in equally-sized intervals. In each subperiod we compare the number of observations in each interval with the expected number of observations in the interval based on the estimated mixture distribution using a χ^2 goodness of fit test. We note that this strategy is dependent on the number of intervals. We choose 8 through 15 equally spaced intervals to evaluate the estimated mixtures. This means that we perform $(15-7) \times 6 = 48$ goodness of fit tests for each candidate mixture density. Table 5.1 shows the number of rejections for different mixtures in each subperiod using a 5% level of significance. We note that three cases result in four rejections: the mixture of a Weibull or a gamma with a truncated normal density and the mixture of two truncated normal distributions. The other mixtures including the mixtures containing the lognormal distribution perform worse. To choose between the three best fitting mixtures, we look at the number of rejections at the 1% and 10% level. In that case the mixture of a Weibull and a truncated normal distribution produces the best fit.

Table 5.1: The outcomes of χ^2 goodness of fit test for different mixture distributions[1].

components[2]		subperiod					
first	second	1960–64	1965–69	1970–74	1975–79	1980–84	1984–89
normal	normal	0	1	1	1	1	0
gamma	normal	0	0	0	1	3	0
gamma	gamma	0	0	0	2	3	0
lognormal	normal	0	0	0	1	4	0
lognormal	lognormal	0	0	0	1	5	2
Weibull	normal	0	0	0	1	3	0
Weibull	Weibull	0	0	0	1	4	0

[1] The cell denotes the number of rejections at a 5% level out of eight χ^2 goodness of fit test in each subperiod. The data in each subperiod are divided in 8 through 15 equally-sized interval. The χ^2 test compares the number of observations in each interval with the expected number of observations in the interval based on the estimated mixture distribution.
[2] Normal means truncated normal with 0 as point of truncation.

Figure 5.3 shows the fitted density of a mixture of a Weibull and a truncated normal together with the histograms of figure 5.3. The histograms have been normalised such that the area under the bars is equal to one in order to compare them with the density functions. The estimated mixtures fit the histograms reasonably well. Therefore, we decide to consider in this paper a mixture of a Weibull and a truncated normal density. Since a gamma and a truncated normal distribution are also good candidates to describe the first component, we discuss the robustness of our results with respect to the other two mixtures at the end of each section. The density function h of a mixture of a Weibull and a truncated normal evaluated at y_i is given by

$$h(y_i;\beta_1,\alpha_1,\lambda,\mu_2,\sigma_2) = \lambda\frac{\beta_1}{\alpha_1^{\beta_1}}y_i^{\beta_1-1}\exp\left(\left(-\frac{y_1}{\alpha_1}\right)^{\beta_1}\right) + (1-\lambda)\frac{\phi(y_i;\mu_2,\sigma_2^2)}{\Phi(\mu_2/\sigma_2)}, \tag{5.6}$$

where $\phi(y;\mu_2,\sigma_2^2)$ represents the probability density function of a normal distribution with mean μ_2 and variance σ_2^2 and Φ the cumulative density function of a standard normal distribution. The parameters α_1 and β_1 are the scale and location parameters of the Weibull component. The parameters of the mixture $\{\alpha_1,\beta_1,\mu_2,\sigma_2,\lambda\}$ are estimated by maximizing the likelihood function

$$\mathscr{L}(Y;\beta_1,\alpha_1,\mu_2,\sigma_2,\lambda) = \prod_{i=1}^{N} h(y_i;\beta_1,\alpha_1,\mu_2,\sigma_2,\lambda), \tag{5.7}$$

where the density function h is given in (5.6). Here we face, of course, the same problem with the singularity in the likelihood function as in the case of a mixture of two normal densities and we opt for the same solution as before. The numerical algorithm to maximise the likelihood functions (5.7) is Newton-Raphson. A range of starting values is used to find the maximum. In case two or more maxima are found the maximum with the largest value of the likelihood function is chosen. Finally, it can easily be shown that scaling of the data via multiplying by a constant k does not influence the estimated value of the mixing parameter and changes the other parameters in the corresponding way, $k\alpha_1$, $k\mu_2$ and $k\sigma$. Therefore, scaling the data by the sum of the real GDP per capita in a year does not alter the conclusions, since the means and the variances of the components change accordingly.

4 Distribution of Wealth

To describe the cross-section distribution of real GDP per capita over the 120 countries in each year, we estimate a mixture of a Weibull and a truncated normal density. First, we focus on the six subperiods. The first five columns of table 5.2 show the parameter estimates of the fitted mixture distributions in every subperiod. Apart from the mixing proportion λ it is difficult to interpret the estimated scale and location parameters directly, since they do not represent the means and variances of the components. Therefore, the second panel of the table shows the means and the variances of the poor and the rich group based on the parameters estimates together with the mean and variance of all countries. Note that the truncation of the normal component becomes less important in the end of the sample.

From the sixth column of table 5.2 we notice that the mixing proportions indicate an almost constant percentage of poor countries in the first three subperiods followed by a substantially increase after the subperiod 1970–1974. There are 14% more poor countries in the final subperiod than in the first subperiod. A Likelihood Ratio [LR] test for equal mixing proportions in the first and final subperiod equals, however, 2.56 which is not significant at a 5% level (the 95% percentile of the χ^2 distribution with one degree of freedom equals 3.84). The LR test is computed by comparing the sum of the maximum likelihoods of the two unrestricted densities with the maximum likelihood of the mixture densities in the first and final period estimated under the restriction of equal mixing parameters.

The seventh column of table 5.2 shows the mean of all countries in every subperiod. The mean has increased monotonically over time. The same is true for the means of the poor and the rich group. Notice that the mean real GDP per capita of both groups has grown faster than the overall mean. This is possible because the relative number of poor countries has increased over time. The difference between the mean of the poor and rich group is about 4.1 in the first subperiod, while in the

Table 5.2: Estimates of mixture parameters, means and variances of real GDP per capita of the poor and the rich component and of all countries in the six subperiods.[1]

sub-period	$\hat{\beta}_1$	$\hat{\mu}_2$	$\hat{\alpha}_1$	$\hat{\sigma}_2^2$	$\hat{\lambda}$	mean			variance		
						all	poor	rich	all	poor	rich
60–64	2.02	5.24	1.40	6.27	0.70	2.46	1.24	5.35	5.55	0.41	5.66
65–69	1.79	6.68	1.68	6.50	0.73	2.91	1.49	6.71	7.75	0.74	6.27
70–74	1.63	8.29	2.03	7.05	0.73	3.53	1.81	8.30	11.20	1.31	6.98
75–79	1.38	10.66	2.76	3.13	0.82	4.01	2.52	10.66	13.71	3.41	3.13
80–84	1.35	11.77	2.95	2.18	0.82	4.32	2.70	11.78	16.29	4.08	2.18
84–89	1.25	13.32	3.15	3.02	0.84	4.62	2.93	13.33	20.60	5.57	3.02

[1] The mean and variance of the truncated normal component are computed using the formulae in appendix A of Maddala (1986). The mean and the variance of a Weibull distribution are $\alpha_1 \Gamma(1 + 1/\beta_1)$ and $\alpha_1^2 \Gamma(1+2/\beta_1) - (\alpha_1 \Gamma(1+1/\beta_1))^2$ respectively.

final subperiod this difference is 10.4. This indicates that the means of the real GDP per capita of the two groups are diverging. However, the growth rates in the mean of both groups are roughly the same. The mean of real GDP per capita of the rich countries in the final subperiod is two and a half times larger than in the first subperiod. For the poor group this factor is about 2.4. A LR test for equal growth rates equals 0.12, which is not significant at a 5% percent level. This means that although the difference in the mean between the poor and the rich group gets larger over the last 25 years, the growth rates of the means of both groups over this period are not significantly different. To compute the LR test we estimate the mixture distribution in the first period and the final period jointly under the restriction of an equal growth rate.

The final three columns of table 5.2 display the variance of the poor, the variance of rich and the variance of all countries. The total variance has increased monotonically over the last 25 years. The same conclusion can be drawn for the spread of wealth within the poor group, which indicates the absence of convergence within the poor group. For the group of rich countries an increase in the spread of wealth is followed by a decrease after the subperiod 1970–1974.

We have to interpret the results of the diverging means with care. Changes in the mean of each component over time can be caused by two forces. First, the real GDP per capita of countries in a group can increase over time. Second, countries can switch from the poor to the rich group and *vice versa*, which can lead to a change in the ratio of the means of the rich and the poor group. A typical example of the latter occurs when only the very rich countries stay in the rich group. The same kind of reasoning counts for the variances of each component. Changes in the variances of the components can also be caused by changes in the mixing parameter.

To correct for the effect of the decrease in the number of rich countries on the development of the means and variances of the components, we estimate in each period a mixture of a Weibull and a truncated normal with equal mixing proportions. We analyse three different scenarios. First, we determine an optimal mixing parameter for the six subperiods by jointly estimating the mixture densities under the restriction of equal mixing parameters. Next, we set the mixing parameter equal to the estimated mixing parameter in the final subperiod (=0.84) and equal to the estimated parameter in the first subperiod (=0.70). Notice that we theoretically still allow for switches of countries between the poor and the rich group. Using the same techniques that we apply in the next section, we can show that the number of switches between the two groups is low. This means that the rich and the poor group contain almost the same countries in every subperiod.

Table 5.3 shows the means and the variances of each component under the different restrictions on λ. Several conclusions emerge from the results of this table. Not surprisingly, fixing the mixing parameter results in different values for the means of both groups. However, for all three scenarios, the means of the poor and the rich group still diverge, which implies that the change in the number of rich countries is not the driving force in the diverging process. Note that the growth rates in real GDP per capita over the last 25 years of the rich and the poor group are still about the same.

The variances of the components are more sensitive to the value of the mixing parameter. Under equal mixing parameters, the variance of the poor group still increases over time. For the rich group the situation is different. From the lower left panel of table 5.3 we observe that the variance of the countries, which were rich in the beginning of the sample, is increasing over time. This indicates that the decrease in variance, when we allow for a changing mixing parameter, is mainly due to the decrease in the number of rich countries. Hence, a number of countries, which originally were located in a middle group, was not capable of catching-up with the remaining rich countries. The lower right panel of table 5.3 shows the development of the variance of the countries, who ended up rich in the last subperiod. We still notice the decrease in the variance after the period 1970–1974 and the increase after 1980–1984 but the changes in the variances are less pronounced.

The results in tables 5.2 and 5.3 are not suitable to notice short run patterns, since we have considered the average real GDP of five consecutive years. In the remainder of this section we analyse the distribution of the real GDP per capita using a mixture of a Weibull and a truncated normal density for each year from 1960 until 1989. Instead of using tables with parameter estimates, we report the main results in several graphs, which show the interesting aspects of the estimated distributions[15].

Figure 5.4 shows the estimated values of the mixing proportions λ. In 1960 the percentage of poor countries was about 71%. In the first part of our sample there is an overall effect of a decrease in the number of poor countries to 67% in 1973, but after 1973 the number of poor countries has risen especially during the period 1975–1977. At the end of the sample the percentage of poor countries seems to stabilise around 83%. These results match the outcomes of table 5.2.

[15] A detailed outline of the parameter estimates can be obtained from the authors.

Table 5.3: The means and the variances of the poor and the rich component in the six subperiods for different values of the mixing parameters[1].

	unrestricted value of λ				constant value of $\lambda = 0.79$			
	mean		variance		mean		variance	
subperiod	poor	rich	poor	rich	poor	rich	poor	rich
1960-1964	1.24	5.35	0.41	5.66	1.35	6.11	0.57	4.23
1965-1969	1.49	6.71	0.74	6.27	1.59	7.32	0.93	4.70
1970-1974	1.81	8.30	1.31	6.98	1.97	8.99	1.70	4.90
1975-1979	2.52	10.66	3.41	3.13	2.44	10.44	3.09	3.80
1980-1984	2.70	11.78	4.08	2.18	2.68	11.73	3.92	2.30
1984-1989	2.93	13.33	5.57	3.02	2.82	13.08	4.94	3.98

	initial period value of $\lambda = 0.70$				final period value of $\lambda = 0.84$			
	mean		variance		mean		variance	
subperiod	poor	rich	poor	rich	poor	rich	poor	rich
1960-1964	1.24	5.35	0.41	5.66	1.42	6.52	0.70	3.34
1965-1969	1.46	6.44	0.69	6.90	1.66	7.64	1.08	3.94
1970-1974	1.73	7.89	1.14	8.15	2.08	9.34	2.03	4.00
1975-1979	2.04	8.95	1.82	8.40	2.58	10.77	3.67	2.81
1980-1984	2.25	10.18	2.33	7.99	2.72	11.80	4.19	2.13
1985-1989	2.28	10.69	2.34	13.16	2.93	13.33	5.57	3.02

[1] The results in the upper right corner are ased on a joint estimate of the six mixture densities with equal λ parameter. In the lower panel of the table the λ is equal to the estimated λ in the first and the final subperiod respectively, see the sixth column of table 5.2.

Figure 5.5 shows means and variances of the real GDP per capita in each year for the period 1960–1989, which are based on the parameter estimates of the mixtures. The left panel of the figure shows the overall means and the means of each component. The mean of the real GDP per capita of all countries has increased almost monotonically during the whole period. There are small decreases in the periods 1974–1975 and 1980–1983 reflecting the oil crisis and the crisis in the beginning of the eighties. These periods of decrease can also be detected in the mean of the poor group and the mean of the rich group. In 1960 the difference in the means is about 3.8, while in 1989 this difference is 11. The means of both groups are diverging, which leads to a gap between the poor and the rich group. If we however look at the growth rates of both groups we see that for the poor group the real GDP per capita in 1989 is about 2.5 times larger than in 1960, while for the rich group the factor is about 2.8. A LR test for equal growth rates equals 0.37, which is not significant at 5% level of significance. Therefore, this implies again that although the means

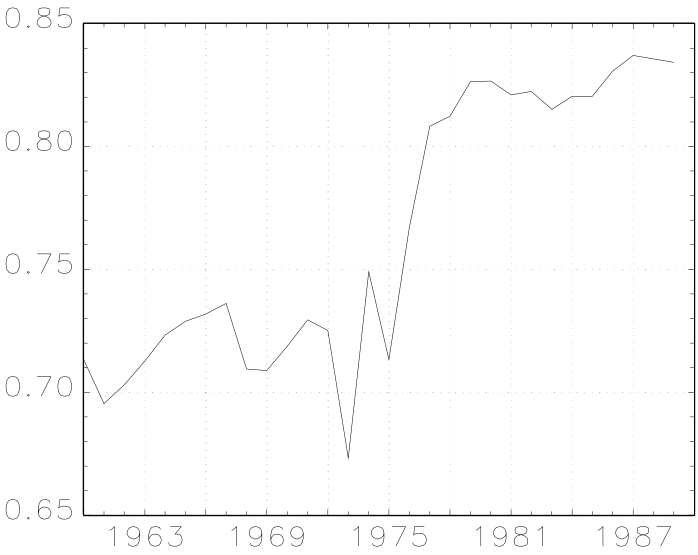


Fig. 5.4: Estimated mixing proportions in each year (1960–1989).

of the poor and the rich group are diverging, the growth rates of the means of both groups are the same over the period 1960–1989. To investigate whether changes in the mixing parameter are responsible for the effects on the means, we estimate the mixture densities under the restriction of equal mixing parameters like in table 5.3. Unreported results show that although we find slightly different values for the means of the poor and the rich group, the means of the two groups are still diverging and the growth rates of the two groups are still about the same.

The right panel of figure 5.5 shows the variance for all countries and for the poor and the rich group in every year. The variance of the real GDP per capita of all countries has risen during 1960–1989 indicating an increase in the spread of wealth between all countries. There are two short periods with a decrease in the variance, *i.e.* 1974–1975 and 1980–1982. The same periods can be found in the variance of the poor group. Unreported estimation results show that the increase in the variance of the pour group remains if we fix the mixing proportion λ. The sharp increase in the variance of the poor group after 1975 is due to the increase in the number of poor countries.

Figure 5.5 shows an increase in the spread of wealth within the rich group until 1973. After the oil crisis the variance has decreased strongly until 1982. In the period

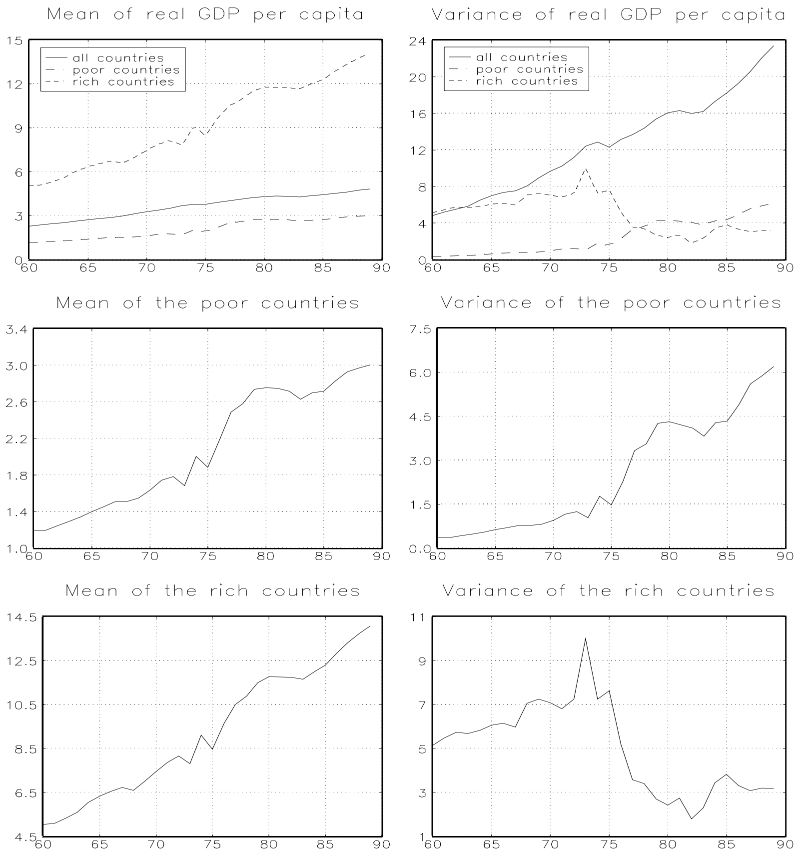


Fig. 5.5: Means and variances of the poor and the rich component and all countries.

1982-1986 there is an increase in the variance. The same analysis as in table 5.3 shows that the variance of the countries, which are rich in the beginning of the sample, is increasing over time and that the variance of the countries, which are rich at the end of the sample, does not decrease. Furthermore, the decrease in the variance of the rich component during the two crises still remains if we fix the mixing parameter, but the decreases are much smaller. In summary, the analysis shows that especially in the middle of the seventies a number of countries was not capable of catching-up with the rich countries and became poor. This has caused a gap between the poor and the rich group. The movement of the poorest rich countries from the rich to the poor group leads to an increase in the variance of the poor countries and a decrease in the variance of the very rich countries.

In this section we have analysed the development of the real GDP per capita over time using a mixture of a Weibull and a truncated normal density. We have seen that the number of poor countries has increased over the last 30 years. The difference in the mean of the real GDP per capita of the poor and the rich group is increasing, indicating no convergence in the level. However, there is no significant difference in the growth rates of both groups, which suggests convergence in growth rates. The spread of wealth within the poor group increases. This is partially caused by the increase in the number of poor countries. For the rich group there is some indication for convergence as the spread of wealth of the rich group has decreased during the two crises in our sample. The largest part of these decreases is however due to the decrease in the number of rich countries. These rich countries were not capable of catching-up with the very rich countries.

In order to investigate the sensitivity of the results with respect with our choice of mixture, we performed the same analysis of cross-section distribution of the real GDP per capita using a mixture of a gamma and a truncated normal density and a mixture of two truncated normal densities, which also produce a reasonable fit according to table 5.1. The results coming out of these analyses are roughly the same. The main difference lies in the estimated mixing proportions before 1974, using a mixture of two truncated normals. The estimated mixing proportions are about 0.10 smaller compared to the mixtures of a Weibull or gamma and a truncated normal. In the next section we analyse the intra-distribution movement of countries within the estimated mixtures. We also consider in more detail the switches of countries between the poor and the rich group.

5 Mobility in Wealth

So far our analysis was limited to describing the development of the distribution of real GDP per capita in each year. In this section we consider the intra-distribution mobility of wealth. The obvious strategy is to look at switches of countries and/or groups of countries from the poor to the rich group and *vice versa*. From figure 5.4 we observe that the mixing proportion has risen during the period 1960–1989 indicating an increase in the number of poor countries. One might conclude that the main mobility between the two groups consists of countries moving from the rich to the poor group. However, even when the mixing parameter is rising over time, there can be switches from poor to rich, when the number of rich countries that become poor is larger than the number of poor countries that become rich. We start analysing mobility in wealth by considering the individual switches of countries between the two groups.

To analyse the mobility between groups, we need to decide whether a country is rich or poor. We can do inference about this question based on the ex-post conditional probability that an observation is generated by one of the components of the mixture, see (5.4). We declare a country poor, if the ex-post conditional probability that a country belongs to the first component of the mixture is larger than 50%, *i.e.* $\Pr[s_i = 1 \mid y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\lambda}] > 0.5$, otherwise the country is labelled as rich. Note that this means that a rich country can become poor even if the level of real GDP per capita of this country does not change or even increases. Such a situation can for instance arise when the other rich and the poor countries grow faster than this country. In summary, switching from rich to poor depends on the relative movement of a country in the distribution with respect to the other countries.

Table 5.4 displays the number of countries that belong to each group based on the ex-post conditional probability. We see that the number of poor countries has risen from 87 in first subperiod to 100 in the last subperiod. The movements from the poor group to the rich group can be summarised as follows. After the first subperiod only Hong Kong moves from the poor group to the rich group and stays in the rich group for two subperiods. However, after 1974 Hong Kong moves back to the rich group. Furthermore, Barbados moves from the poor to the rich group after the second subperiod and stays in the rich group for only one subperiod. The number of movements from the rich group to the poor group is much larger. Especially after the subperiod 1970–1974 many countries have moved from the rich to the poor group including Argentina, Puerto Rico, Iran and Israel, Spain and Ireland. These countries were not able to catch-up with the very rich countries. After 1979 only Venezuela, Trinidad and Saudi Arabia have moved from the rich group to the poor group. Before 1970, Martinique, Barbados, Mexico and Chile have moved from rich to the poor group. There are 19 countries that are rich in every period, *i.e.* Canada, the USA, Japan, Australia, New Zealand, Iceland, Switzerland, Sweden and all countries of the European Union except for Greece, Ireland, Portugal and Spain. There are 86 countries including most of the African and Asian countries that are poor in every period.

The same analysis can be performed using the estimation results in each year. Figure 5.6 shows the number of rich and poor countries in each year based on the ex-post probabilities of the estimated mixtures. In the period 1960–1973 the number of poor countries drops from 88 to 83. After 1975 we see an increase in the number

Table 5.4: The number of rich and poor countries in each subperiod based on ex-post probabilities[1].

subperiod	60–64	65–69	70–74	75–79	80–84	85–89
# poor	87 (→ 86)	90 (→ 89)	90 (→ 90)	98 (→ 97)	98 (→ 98)	100
# rich	33 (→ 29)	30 (→ 29)	30 (→ 22)	22 (→ 21)	22 (→ 20)	20

[1] In parentheses the number of countries that are in the same group the following period.

of poor countries resulting in 99 poor and 21 rich countries in the final year of our sample. The majority of the switches is from the rich to the poor group.

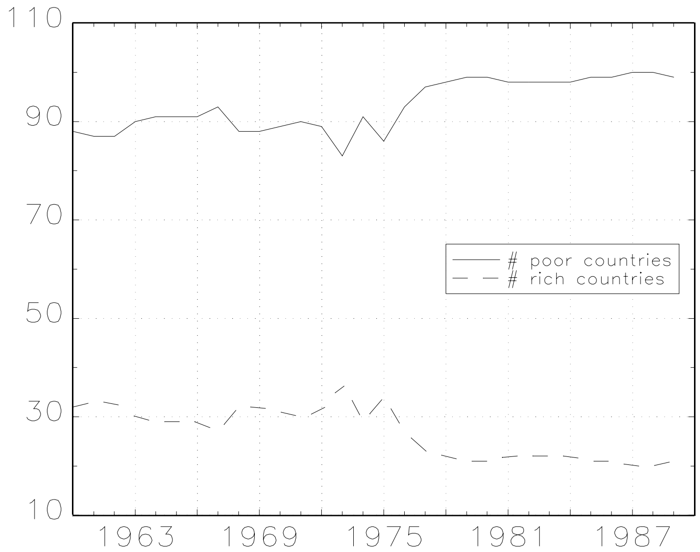


Fig. 5.6: Number of rich and poor countries in each year based on ex-post probabilities.

To investigate the intra-distribution movements of countries we follow the strategy proposed by Quah (1993a). He analyses the intra-distribution dynamics of real GDP per worker over time by a so called fractile Markov Chain. Formally, let F_t denote the distribution of real GDP per worker at time t and suppose that the distribution at time $t+1$ can be written as

$$F_{t+1} = MF_t, \tag{5.8}$$

where M is an operator which maps the distribution F at time t into the distribution at time $t+1$. Iteration of (5.8) gives a prediction for future distributions of the ex-post probabilities

$$F_{t+k} = \underbrace{M...M}_{k} F_t = M^k F_t. \tag{5.9}$$

Quah (1993a) approximates the operator M by an transition matrix by discretising the distribution F_t into intervals. Then M becomes a transition matrix of a

Markov chain. The ergodic probabilities of the Markov chain give insight in the limiting distribution over the states[16]. The transition matrix M is estimated by averaging the total number of switches between the predefined intervals on F. A more technical description of analysing mobility using Markov chains can be found in Shorrocks (1978) and Geweke, Marshall and Zarkin (1986).

In this paper we use the simple framework of Quah (1993a) to analyse the movements of countries between rich and poor. For the distribution F_t we choose the cross-section distribution of the *ex-post probabilities of being poor* in year t (Since the ex-post probability of being poor is equal to one minus the ex-post probability of being rich, we can limit ourselves to analysing the first probabilities.). To estimate the M matrix we divide the the cross-section distribution of ex-post probabilities of being poor at time t, F_t into equally-sized intervals, which is in the line of Quah (1993b). The $[0, 1]$ interval on which F_t is defined, is divided into 2, 3 and 4 equally-sized intervals. In the case of 2 equally-sized intervals, we consider movements from the rich to the poor group and *vice versa*. The division into 3 intervals is useful to analyse whether countries who initially belong to a "middle" group can catch up with the rich countries or fall behind. Movements within the rich and the poor group can be analysed if we use 4 subdivision. The transition matrix M is estimated by averaging the total number of switching between the states over 30 years.

Table 5.5 shows the estimated values of M for the three proposed subdivisions. The transition matrix of the 2-state Markov process shows that the probability of staying poor is larger than the probability of staying rich. The ergodic probabilities of being poor is 0.83, which matches the estimates of the mixing proportions in the last years of our sample period. The transition matrix of the 3-state Markov chain shows the probability of moving from the middle group to the poor group is larger than *vice versa*, which indicates that the probability of catching up is smaller than the probability of falling behind. The ergodic probability of being in the middle group shows that the middle group is vanishing. This matches our earlier findings on the divergence of the levels of the means of the poor and the rich group in section 4 and corresponds with the stylized facts, discussed in section 2. We note that the inconsistency in the ergodic probabilities (0.83 for 2-state, 0.86 for 3-state) is due to the relatively small sample size. The transition matrix of the 4-state Markov process show that if a country is very poor there is almost no chance of becoming rich anymore. The probability to catch up is larger for countries who are in the middle rich group than for countries in the middle poor group. The diagonal elements of the transition matrices are always larger than 0.5, except for the state enabled middle rich in the 4-state Markov process. Further, only sub- and superdiagonal elements differ substantially from zero except for the transition from middle rich to rich, indicating that there are almost no major movements in relative wealth. This implies that the rate at which convergence proceeds, is not large enough for the poorest countries to escape from a poverty trap. Similar findings are reported in Quah (1993a,b).

---

[16] As Quah (1993b) indicates, this framework is much too simple for forecasting. The limiting distribution should be interpreted as an indication for the long-run tendencies in the data rather than a forecast.

There is no need that the transition matrix M is time invariant or that the law of motion for F_t is first order. The former statement is not straightforward to analyse in the present framework. The latter however, can be analysed by considering, for instance, second and higher order Markov chains and compare the estimates of the second order transition matrix with estimates of M from table 5.5 to the power two or to compare the ergodic probabilities. In table 5.6 we show the ergodic probabilities based on a first, second and a third order Markov process. We see that if we increase the order of the chain the ergodic probability of being poor increases. However, the conclusions about the long-run tendencies in the data stay the same.

In this section we have analysed the mobility in wealth using the outcomes of the estimated mixtures of a Weibull and a truncated normal density. The main mobility

Table 5.5: Intra-distribution movements in real GDP per capita analysed using a first order Markov Chain on the ex-post probabilities.

first order Markov process (2-states)

	poor	rich
poor	0.99	0.05
rich	0.01	0.95
ergodic[1]	0.83	0.17

first order Markov process (3-states)

	poor	middle	rich
poor	0.99	0.24	0.01
middle	0.01	0.60	0.03
rich	0.00	0.17	0.96
ergodic[1]	0.86	0.02	0.12

first order Markov process (4-states)

	very poor	middle poor	middle rich	very rich
very poor	0.99	0.24	0.10	0.00
middle poor	0.01	0.52	0.31	0.01
middle rich	0.00	0.22	0.38	0.02
very rich	0.00	0.02	0.21	0.97
ergodic[1]	0.88	0.02	0.02	0.08

[1] Ergodic probabilities of the Markov Chain.

Table 5.6: Ergodic probabilities of a first, second and a third order Markov process with 2, 3 and 4 states, see table 5.5.

	order	poor	rich		
2 subdivisions	1	0.83	0.17		
	2	0.87	0.13		
	3	0.89	0.11		
	order	poor	middle	rich	
3 subdivisions	1	0.86	0.02	0.12	
	2	0.90	0.02	0.08	
	3	0.92	0.02	0.06	
	order	very poor	middle poor	middle rich	very rich
4 subdivisions	1	0.88	0.02	0.02	0.08
	2	0.93	0.01	0.01	0.05
	3	0.94	0.01	0.01	0.04

we have detected is movements of countries from the rich group to the poor group, which have caused the increase in the number of poor countries. The middle group has vanished into the poor group because of the inability of poor countries to catch up with the rich countries. The main results stay the same if we use a Gamma instead of a Weibull distribution to describe the distribution of the poor countries. If we however take a mixture of two truncated normal distributions, we observe a bit more mobility in the beginning of the sample, but after 1975 the results are the same.

6 Conclusion

In this paper we have analysed the distribution of real GDP per capita over 120 countries during the period 1960–1989. The cross-section distribution of the real per capita GDP turns out to be bimodal, displaying a relative large group of poor countries and a small group of rich countries. The analysis is split up in two parts. In the first part we describe the bimodal distributions in each year by a mixture of a Weibull and a truncated normal density and analyse the mixing proportions, the means and variances of the components of the mixture. In the second part we use the estimated mixture distributions for analysing intra-distribution mobility.

The analysis of the cross section distributions shows that the means of the real GDP per capita of the poor and the rich group are diverging, resulting in an increasing gap between the poor and the rich group in terms of levels. However, there is indication of convergence in growth rates between the two groups. The analysis of the mixing proportions shows a large increase in the number of poor countries in the middle of the seventies, which results in an increase in the spread of wealth within the poor group and a decline in the spread of wealth within the rich group. The analysis of the mobility of wealth shows that the main mobility is from rich to poor and the "middle" group between poor and rich disappears. The probability to catch up for the poor countries is smaller than the probability of falling behind. The rate at which convergence proceeds, is not large enough for the poorest countries to escape from a poverty trap.

The results have to be interpreted with care and further research is needed. Specific further research topics are to consider conditioning variables and to link up with endogenous growth models.

Acknowledgements

This research was sponsored by the Economic Research Foundation, part of the Netherlands Organization for Scientific Research (N.W.O.). We thank Suzanne Bijl for preliminary data analysis, Gordon Anderson, John Geweke, Andrew Lewbell, Gary Chamberlain, David Hendry, Mars Cramer and participants of several seminars and of the seventh World Congress of the Econometric Society in Tokyo. We are, in particular, indebted to three anonymous referees for very helpful comments on an earlier version of this paper.

References

Abramowitz, M. (1986) Catching Up, Forging Ahead, and Falling Behind, *Journal of Economic History*, **46**, 385–406.

Barro, R.J. (1991) Economic Growth in a Cross Section of Countries, *Quarterly Journal of Economics*, **106**, 407–443.

Barro, R.J. and X. Sala-i-Martin (1992) Convergence, *Journal of Political Economy*, **100**, 223–251.

Baumol, W.J. (1986) Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show, *American Economic Review*, **76**, 1072–1085.

Bianchi, M. (1997) Testing for Convergence: Evidence from Nonparametric Multimodality Tests, *Journal of Applied Econometrics*, **12**, 393–409.

Canova, F. and A. Marcet (1995) The Poor Stay Poor: Nonconvergence Across Countries and Regions, cEPR Discussion Paper Series, No. 1265.

Desdoigts, A. (1994) Changes in the World Income Distribution: A Non-Parametric Approach to Challenge the Neo-Classical Convergence Argument, Working Paper, Université Paris Dauphine.

Durlauf, S.N. and P. Johnson (1995) Multiple Regimes and Cross-Country Growth Behaviour, *Journal of Applied Econometrics*, **10**, 365–384.

Everitt, B. S. and D. J. Hand (1981) *Finite Mixture Distributions*, Monographs on Applied Probability and Statistics, Chapman and Hall, London.

Friedman, M. (1992) Do Old Fallacies Ever Die?, *Journal of Economic Literature*, **30**, 2129–2132.

Geweke, J., R.C. Marshall and G.A. Zarkin (1986) Mobility Indices in Continuous Time Markov Chains, *Econometrica*, **54**, 1407–1423.

Hamilton, J.D. (1991) A Quasi-Bayesian Approach to Estimating Parameters of Mixtures of Normal Distributions, *Journal of Business and Economic Statistics*, **9**, 27–39.

Jones, C.I. (1997) Convergence Revisited, *Journal of Economic Growth*, **2**, 131–153.

Kiefer, N.M. (1978) Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model, *Econometrica*, **46**, 427–434.

Levine, R. and D. Renelt (1991) Cross Country Studies of Growth and Policy: Some Methodological, Conceptual, and Statistical Problems, World Bank Working Paper Series, No. 608.

Levine, R. and D. Renelt (1992) A Sensitivity Analysis of Cross Country Regressions, *American Economic Review*, **82**, 942–963.

Mankiw, N.G., D. Romer and D.N. Weil (1992) A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics*, **429**, 407–437.

Parente, S.L. and E.C. Prescott (1993) Changes in the Wealth of Nations, *Quarterly Review, Federal Reserve Bank of Minneapolis*, **Spring 1993**, 3–16.

Quah, D. (1993a) Galton's Fallacy and Tests of the Convergence Hypothesis, *Scandinavian Journal of Economics*, **95**, 427–443.

Quah, D. (1993b) Empirical Cross-section Dynamics in Economic Growth, *European Economic Review*, **37**, 426–434.

Quah, D. (1996a) Convergence Empirics Across Economies with (some) Capital Mobility, *Journal of Economic Growth*, **1**, 95–124.

Quah, D. (1996b) Empirics for Economic Growth and Convergence, *European Economic Review*, **40**, 1353–1376.

Sala-i-Martin, X. (1994) Cross-sectional Regression and the Empirics of Economic Growth, *European Economic Review*, **38**, 739–747.

Shorrocks, A.F. (1978) The Measurement of Mobility, *Econometrica*, **46**, 1013–1024.

Summers, R. and A. Heston (1991) The Penn World Table (mark 5): An Expanded Set of International Comparisons, 1950–1988, *Quarterly Journal of Economics*, **106**, 327–368.

Titterington, D.M., A.F.M. Smith and U.E. Makov (1985) *Statistical Analysis of Finite Mixture Distributions*, Wiley and Sons, New York.

Part II

Survey papers on Lorenz functions and the generalizations and extensions of income distributions

CHAPTER 6

A Guide to the Dagum Distributions

Christian Kleiber†

Abstract

In a series of papers in the 1970s, Camilo Dagum proposed several variants of a new model for the size distribution of personal income. This Chapter traces the genesis of the Dagum distributions in applied economics and points out parallel developments in several branches of the applied statistics literature. It also provides interrelations with other statistical distributions as well as aspects that are of special interest in the income distribution field, including Lorenz curves and the Lorenz order and inequality measures. The Chapter ends with a survey of empirical applications of the Dagum distributions, many published in Romance language periodicals.

1 Introduction

In the 1970s, Camilo Dagum embarked on a quest for a statistical distribution closely fitting empirical income and wealth distributions. Not satisfied with the classical distributions used to summarize such data – the Pareto distribution (developed by the Italian economist and sociologist Vilfredo Pareto in the late 19th century (Pareto, 1895, 1896, 1897)) and the lognormal distribution (popularized by the French engineer Robert Gibrat (1931)) – he looked for a model accommodating the heavy tails present in empirical income and wealth distributions as well as permitting an interior mode. The former aspect is well captured by the Pareto but not by the lognormal distribution, the latter by the lognormal but not the Pareto distribution. Experimenting with a shifted log-logistic distribution (Dagum, 1975), a generalization of a distribution previously considered by Fisk (1961), he quickly realized that a further parameter was needed. This led to the Dagum type I

† Correspondence to: Christian Kleiber, Dept. of Statistics and Econometrics, Universität Basel, Petersgraben 51, CH-4051 Basel, Switzerland. E-mail: christian.kleiber@unibas.ch.

distribution, a three-parameter distribution, and two four-parameter generalizations (Dagum, 1977, 1980c).

It took more than a decade until Dagum's proposal began to appear in the English-language economic and econometric literature. The first paper in a major econometrics journal utilizing the Dagum distribution appears to be by Majumder and Chakravarty (1990). In the statistical literature, the situation is more favorable, in that the renowned *Encyclopedia of Statistical Sciences* contains, in Vol. 4 (Kotz *et al.*, 1983), an entry on income distribution models, unsurprisingly authored by Camilo Dagum (Dagum, 1983). In retrospect, the reason for this long delay is fairly obvious: Dagum's (1977) paper was published in *Economie Appliquée*, a French journal with only occasional English-language contributions and fairly limited circulation in English-language countries. In contrast, the paper introducing the more widely known Singh and Maddala (1976) distribution was published in *Econometrica*, just one year before Dagum's contribution. It slowly emerged that the Dagum distribution is, nonetheless, often preferable to the Singh-Maddala distribution in applications to income data.

This Chapter provides a brief survey of the Dagum distributions, including interrelations with several more widely known distributions as well as basic statistical properties and inferential aspects. It also revisits one of the first data sets considered by Dagum and presents a survey of applications in economics.

2 Genesis and Interrelations

Dagum (1977) motivates his model from the empirical observation that the income elasticity $\eta(F,x)$ of the cumulative distribution function (CDF) F of income is a decreasing and bounded function of F. Starting from the differential equation

$$\eta(F,x) = \frac{d\log F(x)}{d\log x} = ap\{1-[F(x)]^{1/p}\}, \quad x \geq 0, \tag{6.1}$$

subject to $p > 0$ and $ap > 0$, one obtains

$$F(x) = [1+(x/b)^{-a}]^{-p}, \quad x > 0. \tag{6.2}$$

This approach was further developed in a series of papers on generating systems for income distributions (Dagum, 1980a,b, 1983, 1990). Recall that the well-known Pearson system is a general-purpose system not derived from observed stable regularities in a given area of application. D'Addario's (1949) system is a translation system with flexible so-called generating and transformation functions built to encompass as many income distributions as possible; see e.g. Kleiber and Kotz (2003) for further details. In contrast, the system specified by Dagum starts from characteristic properties of empirical income and wealth distributions and leads to a generating system specified in terms of

$$\frac{d\log\{F(x)-\delta\}}{d\log x} = \vartheta(x)\phi(F) \leq k, \quad 0 \leq x_0 < x < \infty, \tag{6.3}$$

where $k > 0$, $\vartheta(x) > 0$, $\phi(x) > 0$, $\delta < 1$, and $d\{\vartheta(x)\phi(F)\}/dx < 0$. These constraints ensure that the income elasticity of the CDF is a positive, decreasing and bounded function of F, and therefore of x. Table 6.1 provides a selection of models that can be deduced from Dagum's system for certain specifications of the functions ϑ and ϕ, more extensive versions are available in (Dagum, 1990, 1996). The parameter denoted as α is Pareto's alpha, it depends on the parameters of the underlying distribution and equals a for the Dagum and Fisk distributions and aq in the Singh-Maddala case (see below). The parameter denoted as β also depends on the underlying distribution and equals p in the Dagum case. In addition, signs or values of the parameters β and δ consistent with the constraints of equation (6.3) are indicated. Among the models specified in Table 6.1 the Dagum type II and III distributions are mainly used as models of wealth distribution.

Table 6.1: Dagum's generalized logistic system of income distributions

Distribution	$\vartheta(x)$	$\phi(F)$	(δ,β)	Support
Pareto (I)	α	$(1-F)/F$	$(0,0)$	$0 < x_0 \leq x < \infty$
Fisk	α	$1-F$	$(0,0)$	$0 \leq x < \infty$
Singh-Maddala	α	$\frac{1-(1-F)^\beta}{F(1-F)^{-1}}$	$(0,+)$	$0 \leq x < \infty$
Dagum(I)	α	$1-F^{1/\beta}$	$(0,+)$	$0 \leq x < \infty$
Dagum(II)	α	$1-\left(\frac{F-\delta}{1-\delta}\right)^{1/\beta}$	$(+,+)$	$0 \leq x < \infty$
Dagum(III)	α	$1-\left(\frac{F-\delta}{1-\delta}\right)^{1/\beta}$	$(-,+)$	$0 < x_0 \leq x < \infty$

Dagum (1983) refers to his system as the *generalized logistic-Burr system*. This is due to the fact that the Dagum distribution with $p = 1$ is also known as the log-logistic distribution (the model Dagum (1975) experimented with). In addition, generalized (log-) logistic distributions arise naturally in Burr's (1942) system of distributions, hence the name. The most widely known Burr distributions are the Burr XII distribution – often just called the Burr distribution, especially in the actuarial literature – with CDF

$$F(x) = 1-(1+x^a)^{-q}, \quad x > 0,$$

and the Burr III distribution with CDF

$$F(x) = (1+x^{-a})^{-p}, \quad x > 0.$$

In economics, these distributions are more widely known, after introduction of an additional scale parameter, as the Singh-Maddala and Dagum distributions. Thus the Dagum distribution is a Burr III distribution with an additional scale parameter and therefore a rediscovery of a distribution that had been known for some 30 years prior to its introduction in economics. However, it is not the only rediscovery of this distribution: Mielke (1973), in a meteorological application, arrives at a three-parameter distribution he calls the kappa distribution. It amounts to the Dagum distribution in a different parametrization. Mielke and Johnson (1974) refer to it as the Beta-K distribution. Even in the income distribution literature there is a parallel development: Fattorini and Lemmi (1979), starting from Mielke's kappa distribution but apparently unaware of Dagum (1977), propose (6.2) as an income distribution and fit it to several data sets, mostly from Italy.

Not surprisingly, this multi-discovered distribution has been considered in several parameterizations: Mielke (1973) and later Fattorini and Lemmi (1979) use $(\alpha,\beta,\theta) := (1/p, bp^{1/a}, ap)$, whereas Dagum (1977) employs $(\beta,\delta,\lambda) := (p,a,b^a)$. The parametrization used here follows McDonald (1984), because both the Dagum / Burr III and the Singh-Maddala/Burr XII distributions can be nested within a four-parameter generalized beta distribution of the second kind (hereafter: GB2) with density

$$f(x) = \frac{a x^{ap-1}}{b^{ap} B(p,q)[1+(x/b)^a]^{p+q}}, \quad x > 0,$$

where $a,b,p,q > 0$. Specifically, the Singh-Maddala is a GB2 distribution with shape parameter $p = 1$, while the Dagum distribution is a GB2 with $q = 1$ and thus its density is

$$f(x) = \frac{ap x^{ap-1}}{b^{ap}[1+(x/b)^a]^{p+1}}, \quad x > 0. \tag{6.4}$$

It is also worth noting that the Dagum distribution (D) and the Singh-Maddala distribution (SM) are intimately connected, specifically

$$X \sim D(a,b,p) \iff \frac{1}{X} \sim SM(a,1/b,p) \tag{6.5}$$

This relationship permits to translate several results pertaining to the Singh-Maddala family into corresponding results for the Dagum distributions, it is also the reason for the name *inverse Burr distribution* often found in the actuarial literature for the Dagum distribution (e.g., Panjer (2006)).

Dagum (1977, 1980c) introduces two further variants of his distribution, hence the previously discussed standard version will be referred to as the Dagum type I distribution in what follows. The Dagum type II distribution has the CDF

$$F(x) = \delta + (1-\delta)[1+(x/b)^{-a}]^{-p}, \quad x \geq 0,$$

where as before $a,b,p > 0$ and $\delta \in (0,1)$. Clearly, this is a mixture of a point mass at the origin with a Dagum (type I) distribution over the positive halfline. The type II distribution was proposed as a model for income distributions with null and negative incomes, but more particularly to fit wealth data, which frequently presents a large number of economic units with null gross assets and with null and negative net assets.

There is also a Dagum type III distribution, like type II defined as

$$F(x) = \delta + (1-\delta)[1+(x/b)^{-a}]^{-p},$$

with $a,b,p > 0$. However, here $\delta < 0$. Consequently, the support of this variant is now $[x_0,\infty)$, $x_0 > 0$, where $x_0 = \{b[(1-1/a)^{1/p}-1]\}^{-1/a}$ is determined implicitly from the constraint $F(x) \geq 0$.

As mentioned above, both the Dagum type II and the type III are members of Dagum's generalized logistic-Burr system.

Investigating the relation between the functional and the personal distribution of income, Dagum (1999) also obtained the following bivariate CDF when modeling the joint distribution of human capital and wealth

$$F(x_1,x_2) = (1+b_1x_1^{-a_1}+b_2x_2^{-a_2}+b_3x_1^{-a_1}x_2^{-a_2})^{-p}, \quad x_i > 0,\ i = 1,2.$$

If $b_3 = b_1b_2$,

$$F(x_1,x_2) = (1+b_1x_1^{-a_1})^{-p}(1+b_2x_2^{-a_2})^{-p},$$

hence the marginals are independent. There do not appear to be any empirical applications of this multivariate Dagum distribution at present.

The remainder of this paper will mainly discuss the Dagum type I distribution.

3 Basic Properties

The parameter b of the Dagum distribution is a scale while the remaining two parameters a and p are shape parameters. Nonetheless, these two parameters are not on an equal footing: This is perhaps most transparent from the expression for the distribution of $Y := \log X$, a generalized logistic distribution with PDF

$$f(y) = \frac{ap\,e^{ap(y-\log b)}}{[1+e^{a(y-\log b)}]^{p+1}}, \quad -\infty < y < \infty.$$

Here, only p is a shape (or skewness) parameter while a and $\log b$ are scale and location parameters, respectively.

Figure 6.1 illustrates the effect of variations of the shape parameters: for $ap < 1$, the density exhibits a pole at the origin, for $ap = 1$, $0 < f(0) < \infty$, and for $ap > 1$ there exists an interior mode. In the latter case, this mode is at

$$x_{mode} = b\left(\frac{ap-1}{a+1}\right)^{1/a}.$$

This built-in flexibility is an attractive feature in that the model can approximate income distributions, which are usually unimodal, and wealth distributions, which

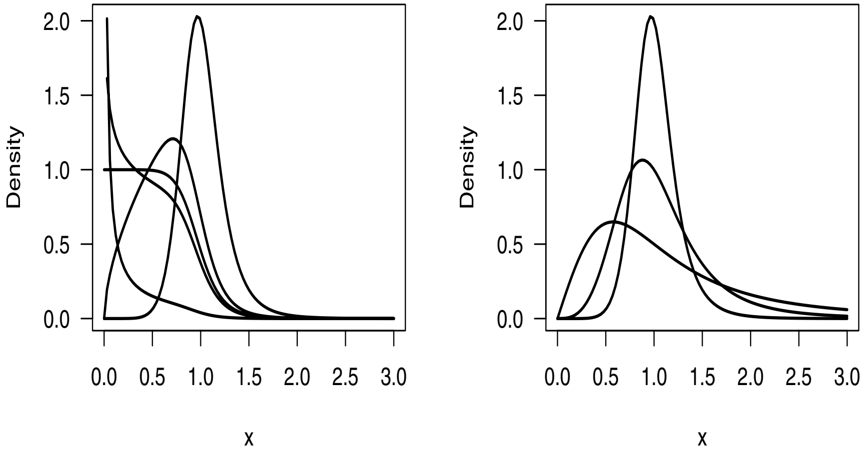


Fig. 6.1: Shapes of Dagum distributions. Left panel: variation of p ($a = 8$, $p = 0.01, 0.1, 0.125, 0.2, 1$, from top left to bottom left). Right panel: variation of a ($p = 1$, $a = 2, 4, 8$, from left to right).

are zeromodal. It should be noted that ap and a determine the rate of increase (decrease) from (to) zero for $x \to 0$ ($x \to \infty$), and thus the probability mass in the tails. It should also be emphasized that, in contrast to several popular distributions used to approximate income data, notably the lognormal, gamma and GB2 distributions, the Dagum permits a closed-form expression for the CDF. This is also true of the quantile function,

$$F^{-1}(u) = b[u^{-1/p} - 1]^{-1/a}, \quad \text{for } 0 < u < 1, \tag{6.6}$$

hence random numbers from a Dagum distribution are easily generated via the inversion method.

The kth moment exists for $-ap < k < a$ and equals

$$E(X^k) = \frac{b^k B(p + k/a, 1 - k/a)}{B(p, 1)} = \frac{b^k \Gamma(p + k/a)\Gamma(1 - k/a)}{\Gamma(p)}, \tag{6.7}$$

where $\Gamma()$ and $B()$ denote the gamma and beta functions. Specifically,

$$E(X) = \frac{b\Gamma(p + 1/a)\Gamma(1 - 1/a)}{\Gamma(p)}$$

and

$$Var(X) = \frac{b^2\{\Gamma(p)\Gamma(p+2/a)\Gamma(1-2/a) - \Gamma^2(p+1/a)\Gamma^2(1-1/a)\}}{\Gamma^2(p)}.$$

Moment-ratio diagrams of the Dagum and the closely related Singh-Maddala distributions, presented by Rodriguez (1983) and Tadikamalla (1980) under the names of Burr III and Burr XII distributions, reveal that both models allow for various degrees of positive skewness and leptokurtosis, and even for a considerable degree of negative skewness although this feature does not seem to be of particular interest in applications to income data. (A notable exception is an example of faculty salary distributions presented by Pocock *et al.* (2003).) Tadikamalla (1980, p. 342) observes "that although the Burr III [= Dagum] distribution covers all of the region ... as covered by the Burr XII [= Singh-Maddala] distribution and more, much attention has not been paid to this distribution." Kleiber (1996) notes that, ironically, the same has happened independently in the econometrics literature.

An interesting aspect of Dagum's model is that it admits a mixture representation in terms of generalized gamma (GG) and Weibull (Wei) distributions. Recall that the generalized gamma and Weibull distributions have PDFs

$$f_{GG}(x) = \frac{a}{\theta^{ap}\Gamma(p)} x^{ap-1} e^{-(x/\theta)^a}, \quad x > 0,$$

and

$$f_{Wei}(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} e^{-(x/b)^a}, \quad x > 0,$$

respectively. The Dagum distribution can be obtained as a compound generalized gamma distribution whose scale parameter follows an inverse Weibull distribution (i.e., the distribution of $1/X$ for $X \sim Wei(a,b)$), symbolically

$$GG(a,\theta,p) \bigwedge_{\theta} InvWei(a,b) = D(a,b,p).$$

Note that the shape parameters a must be identical. Such representations are useful in proofs (see, e.g., Kleiber (1999)), they also admit an interpretation in terms of unobserved heterogeneity.

Further distributional properties are presented in Kleiber and Kotz (2003). In addition, a rather detailed study of the hazard rate is available in Domma (2002).

4 Measuring Inequality using Dagum Distributions

The most widely used tool for analyzing and visualizing income inequality is the Lorenz curve (Lorenz (1905); see also Kleiber (2008) for a recent survey), and several indices of income inequality are directly related to this curve, most notably the Gini index (Gini, 1914).

Since the quantile function of the Dagum distribution is available in closed form, its normalized integral, the Lorenz curve

$$L(u) = \frac{1}{E(X)} \int_0^u F^{-1}(t)dt, \quad u \in [0,1],$$

is also of a comparatively simple form, namely (Dagum, 1977)

$$L(u) = I_z(p+1/a, 1-1/a), \quad 0 \leq u \leq 1, \tag{6.8}$$

where $z = u^{1/p}$ and $I_z(x,y)$ denotes the incomplete beta function ratio. Clearly, the curve exists iff $a > 1$.

For the comparison of estimated income distributions it is of interest to know the parameter constellations for which Lorenz curves do or do not intersect. The corresponding stochastic order, the Lorenz order, is defined as

$$F_1 \geq_L F_2 \quad \Longleftrightarrow \quad L_1(u) \leq L_2(u) \quad \text{for all } u \in [0,1].$$

First results were obtained by Dancelli (1986) who found that inequality is decreasing to zero (i.e., the curve approaches the diagonal of the unit square) if $a \to \infty$ or $p \to \infty$ and increasing to one if $a \to 1$ or $p \to 0$, respectively, keeping the other parameter fixed. A complete analytical characterization is of more recent date. Suppose $F_i \sim \mathrm{D}(a_i, b_i, p_i)$, $i = 1,2$. The necessary and sufficient conditions for Lorenz dominance are

$$L_1 \leq L_2 \quad \Longleftrightarrow \quad a_1 p_1 \leq a_2 p_2 \quad \text{and} \quad a_1 \leq a_2. \tag{6.9}$$

This shows that the less unequal distribution (in the Lorenz sense) always exhibits lighter tails. This was derived by Kleiber (1996) from the corresponding result for the Singh-Maddala distribution using (6.5), for a different approach see Kleiber (1999). Figure 6.2 provides an illustration of (6.9).

Apart from the Lorenz order, stochastic dominance of various degrees has been used when ranking income distributions, hence it is of interest to study conditions on the parameters implying such orderings. A distribution F_1 first-order stochastically dominates F_2, denoted as $F_1 \geq_{FSD} F_2$, iff $F_1 \leq F_2$. This criterion was suggested by Saposnik (1981) as a ranking criterion for income distributions. Klonner (2000) presents necessary as well as sufficient conditions for first-order stochastic dominance within the Dagum family. The conditions $a_1 \geq a_2$, $a_1 p_1 \leq a_2 p_2$ and $b_1 \geq b_2$ are sufficient for $F_2 \geq_{FSD} F_1$, whereas the conditions $a_1 \geq a_2$ and $a_1 p_1 \leq a_2 p_2$ are necessary.

As regards scalar measures of inequality, the most widely used of all such indices, the Gini coefficient, takes the form (Dagum, 1977)

$$G = \frac{\Gamma(p)\Gamma(2p+1/a)}{\Gamma(2p)\Gamma(p+1/a)} - 1. \tag{6.10}$$

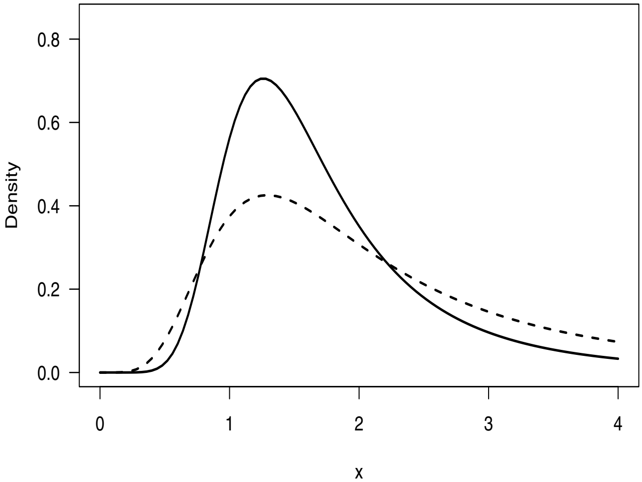


Fig. 6.2: Tails and the Lorenz order for two Dagum distributions: $X_1 \sim D(2,1,3)$ (dashed), $X_2 \sim D(3,1,3)$ (solid), hence $F_1 \geq_L F_2$.

For generalized Gini indices see Kleiber and Kotz (2003). From (6.7), the coefficient of variation (CV) is

$$CV = \sqrt{\frac{\Gamma(p)\Gamma(p+2/a)\Gamma(1-2/a)}{\Gamma^2(p+1/a)\Gamma^2(1-1/a)} - 1}. \tag{6.11}$$

Recall that the coefficient of variation is a monotonic transformation of a measure contained in the generalized entropy class of inequality measures, e.g., Kleiber and Kotz (2003). All these measures are functions of the moments and thus easily derived from (6.7). The resulting expressions are somewhat involved, however, as are expressions for the Atkinson (1970) measures of inequality. Recently, Jenkins (2007) provided formulae for the generalized entropy measures for the more general GB2 distributions, from which the Dagum versions are also easily obtained.

Some 20 years ago, an alternative to the Lorenz curve emerged in the Italian language literature. Like the Lorenz curve the Zenga curve (Zenga, 1984) can be introduced via the first-moment distribution

$$F_{(1)}(x) = \frac{\int_0^x t f(t) dt}{E(X)}, \quad x \geq 0,$$

thus it exists iff $E(X) < \infty$. The Zenga curve is now defined in terms of the quantiles $F^{-1}(u)$ of the income distribution itself and of those of the corresponding first-moment distribution, $F_{(1)}^{-1}(u)$: for

$$Z(u) = \frac{F_{(1)}^{-1}(u) - F^{-1}(u)}{F_{(1)}^{-1}(u)} = 1 - \frac{F^{-1}(u)}{F_{(1)}^{-1}(u)}, \quad 0 < u < 1, \tag{6.12}$$

the set $\{(u, Z(u)) | u \in (0,1)\}$ is the Zenga concentration curve. Note that $F_{(1)} \leq F$ implies $F^{-1} \leq F_{(1)}^{-1}$, hence the Zenga curve belongs to the unit square. It follows from (6.12) that the curve is scale-free.

It is then natural to call a distribution F_2 less concentrated than another distribution F_1 if its Zenga curve is nowhere above the Zenga curve associated with F_1 and thus to define an ordering via

$$F_1 \geq_Z F_2 \qquad :\Longleftrightarrow \qquad Z_1(u) \geq Z_2(u) \text{ for all } u \in (0,1).$$

Zenga ordering within the family of Dagum distributions was studied by Polisicchio (1990) who found that $a_1 \leq a_2$ implies $F_1 \geq_Z F_2$, for a fixed p, and analogously that $p_1 \leq p_2$ implies $F_1 \geq_Z F_2$, for a fixed a. Under these conditions it follows from (6.9) that the distributions are also Lorenz ordered, specifically $F_1 \geq_L F_2$. Recent work of Kleiber (2007) shows that the conditions for Zenga ordering coincide with those for Lorenz dominance within the class of Dagum distributions.

5 Estimation and Inference

Dagum (1977), in a period when individual data were rarely available, minimized

$$\sum_{i=1}^{n} \{F_n(x_i) - [1 + (x_i/b)^{-a}]^{-p}\}^2,$$

a non-linear least-squares criterion based on the distance between the empirical CDF F_n and the CDF of a Dagum approximation. A further regression-type estimator utilizing the elasticity (6.1) was later considered by Stoppa (1995).

Most researchers nowadays employ maximum likelihood (ML) estimation. Two cases need to be distinguished, grouped data and individual data. Until fairly recently, only grouped data were available, and here the likelihood $L(\theta)$, where $\theta = (a, b, p)^\top$, is a multinomial likelihood with (assuming independent data)

$$L(\theta) = \prod_{j=1}^{m} \{F(x_j) - F(x_{j-1})\}, \qquad x_0 = 0,\ x_m = \infty.$$

By construction this likelihood is always bounded from above.

In view of the 30th anniversary of Dagum's contribution it seems appropriate to revisit one of his early empirical examples, the US family incomes for the year 1969.

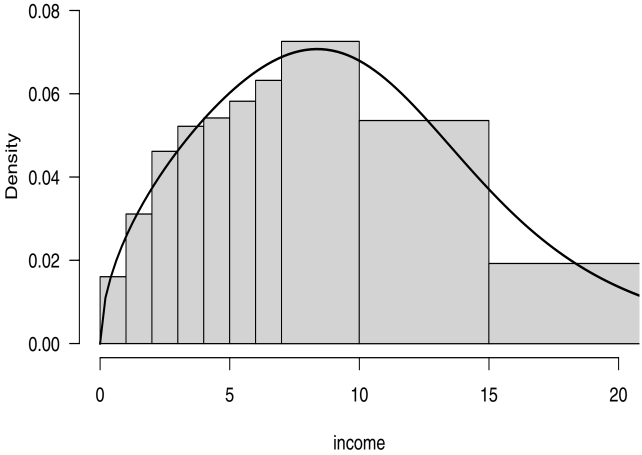


Fig. 6.3: Dagum distribution fitted to the 1969 US family incomes.

The data are given in Dagum (1980c, p. 360). Figure 6.3 plots the corresponding histogram along with a Dagum type I approximation estimated via grouped maximum likelihood. The resulting estimates are $\hat{a} = 4.273$, $\hat{b} = 14.28$ and $\hat{p} = 0.36$, and are in good agreement with the values estimated by Dagum via nonlinear least squares.

With the increasing availability of microdata, likelihood estimation from individual observations attracts increasing attention, and here the situation is more involved: the log-likelihood $\ell(\boldsymbol{\theta}) \equiv \log L(\boldsymbol{\theta})$ for a complete random sample of size n is

$$\begin{aligned} \ell(a,b,p) =& n \log a + n \log p + (ap-1) \sum_{i=1}^{n} \log x_i - nap \log b \\ & - (p+1) \sum_{i=1}^{n} \log\{1 + (x_i/b)^a\} \end{aligned} \tag{6.13}$$

yielding the likelihood equations

$$\frac{n}{a} + p\sum_{i=1}^{n}\log(x_i/b) = (p+1)\sum_{i=1}^{n}\frac{\log(x_i/b)}{1+(b/x_i)^a}, \tag{6.14}$$

$$np = (p+1)\sum_{i=1}^{n}\frac{1}{1+(b/x_i)^a}, \tag{6.15}$$

$$\frac{n}{p} + a\sum_{i=1}^{n}\log(x_i/b) = \sum_{i=1}^{n}\log\{1+(x_i/b)^a\} \tag{6.16}$$

which must be solved numerically. However, likelihood estimation in this family is not without problems: considering the distribution of $\log X$, a generalized logistic distribution, Shao (2002) shows that the MLE may not exist, and if it does not, the so-called embedded model problem occurs. That is, letting certain parameters tend to their boundary values, a distribution with fewer parameters emerges. Implications are that the behavior of the likelihood should be carefully checked in empirical work. It would be interesting to determine to what extent this complication arises in applications to income data where the full flexibility of the Dagum family is not needed.

Apparently unaware of these problems, Domański and Jedrzejczak (1998) provide a simulation study for the performance of the MLEs. It turns out that rather large samples are required until estimates of the shape parameters a, p can be considered as unbiased, while reliable estimation of the scale parameter seems to require even larger samples.

The Fisher information matrix

$$I(\theta) = \left[-E\left(\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}\right)_{i,j}\right] =: \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}$$

takes the form

$$\begin{aligned}
I_{11} &= \frac{1}{a^2(2+p)}\left[p[\{\psi(p)-\psi(1)-1\}^2+\psi'(p)+\psi'(1)]+2\{\psi(p)-\psi(1)\}\right] \\
I_{21} = I_{12} &= \frac{p-1-p\{\psi(p)-\psi(1)\}}{b(2+p)} \\
I_{22} &= \frac{a^2 p}{b^2(2+p)} \\
I_{23} = I_{32} &= \frac{a}{b(1+p)} \\
I_{31} = I_{13} &= \frac{\psi(2)-\psi(p)}{a(1+p)} \\
I_{33} &= \frac{1}{p^2}
\end{aligned}$$

where ψ is the digamma function.

It should be noted that there are several derivations of the Fisher information in the statistical literature, a detailed one using Dagum's parameterization due to Latorre (1988) and a second one due to Zelterman (1987). The latter article considers the distribution of $\log X$, a generalized logistic distribution, using the parameterization $(\theta, \sigma, \alpha) = (\log b, 1/a, p)$.

As regards alternative estimators, an inspection of the scores (6.14)–(6.16) reveals that $\sup_x \|\partial\ell/\partial\theta\| = \infty$, where $\|.\|$ stands for the Euclidean norm, thus the score function is unbounded in the Dagum case. This implies that the MLE is rather sensitive to single observations located sufficiently far away from the majority of the data. There appears, therefore, to be some interest in more robust procedures. For a robust approach to the estimation of the Dagum model parameters using an optimal B-robust estimator (OBRE) see Victoria-Feser (1995, 2000).

Income distributions have always been popular with Italian authors, and the Dagum distribution is no exception. Cheli *et al.* (1995) study mixtures of Dagum distributions and their estimation via the EM algorithm. Distributions of the sample median and the sample range were obtained by Domma (1997). In addition, Latorre (1988) provides delta-method standard errors for several inequality measures derived from MLEs for the Dagum model.

6 Software

As regards available software, Camilo Dagum started to develop routines for fitting his distributions fairly early. A stand-alone package named "EPID" (Econometric Package for Income Distribution) (Dagum and Chiu, 1991) written in FORTRAN was available from the Time Series Research and Analysis Division of Statistics Canada for some time. The program fitted Dagum type I–III distributions and computed a number of associated statistics such as Lorenz and Zenga curves, the Gini coefficient and various goodness of fit measures. More recently, Jenkins (1999) provided Stata routines for fitting Dagum and Singh-Maddala distributions by (individual) maximum likelihood (current versions are available from the usual repositories), while Jenkins and Jäntti (2005, Appendix) present Stata code for estimating Dagum mixtures. Yee (2007) developed a rather large R (R Development Core Team, 2007) package named VGAM (for "vector generalized additive models") that permits fitting nearly all of the distributions discussed in Kleiber and Kotz (2003) – notably the Dagum type I – conditional on covariates by means of flexible regression methods. The computations for Figure 6.3 were also carried out in R, Version 2.5.1, but along different lines, namely via modifying the `fitdistr()` function from the MASS package, the package accompanying Venables and Ripley (2002).

7 Applications of Dagum Distributions

Although the Dagum distribution was virtually unknown in the major English language economics and econometrics journals until well into the 1990s there are several early applications to income and wealth data, most of which appeared in French, Italian and Latin American publications. Examples include Fattorini and Lemmi (1979) who consider Italian data, Espinguet and Terraza (1983) who study French earnings and Falcão Carneiro (1982) with an application to Portuguese data. Even after 1990 there is a noticable bias towards Romance language contributions. Fairly recent examples include Blayac and Serra (1997), Dagum *et al.* (1995) and Martín Reyes *et al.* (2001).

Table 6.2: Selected applications of Dagum distributions

Country	Source
Argentina	Dagum (1977), Botargues and Petrecolla (1999a,b)
Australia	Bandourian *et al.* (2003)
Belgium	Bandourian *et al.* (2003)
Canada	Dagum (1977, 1985), Dagum and Chiu (1991), Bandourian *et al.* (2003), Chotikapanich and Griffiths (2006)
Czech Republik	Bandourian *et al.* (2003)
Denmark	Bandourian *et al.* (2003)
Finland	Bandourian *et al.* (2003), Jenkins and Jäntti (2005)
France	Espinguet and Terraza (1983), Dagum *et al.* (1995), Bandourian *et al.* (2003)
Germany	Bandourian *et al.* (2003)
Hungary	Bandourian *et al.* (2003)
Ireland	Bandourian *et al.* (2003)
Israel	Bandourian *et al.* (2003)
Italy	Fattorini and Lemmi (1979), Dagum and Lemmi (1989), Bandourian *et al.* (2003)
Mexico	Bandourian *et al.* (2003)
Netherlands	Bandourian *et al.* (2003)
Norway	Bandourian *et al.* (2003)
Philippines	Bantilan *et al.* (1995)
Poland	Domański and Jedrzejczak (2002), Bandourian *et al.* (2003), Łukasiewicz and Orłowski (2004)
Portugal	Falcão Carneiro (1982)
Russia	Bandourian *et al.* (2003)
Slovakia	Bandourian *et al.* (2003)
Spain	Bandourian *et al.* (2003)
Sri Lanka	Dagum (1977)
Sweden	Fattorini and Lemmi (1979), Bandourian *et al.* (2003)
Switzerland	Bandourian *et al.* (2003)
Taiwan	Bandourian *et al.* (2003)
United Kingdom	Victoria-Feser (1995, 2000)
USA	Dagum (1977, 1980c, 1983), Fattorini and Lemmi (1979), Majumder and Chakravarty (1990), Campano (1991), McDonald and Mantrala (1995), McDonald and Xu (1995), Bandourian *et al.* (2003)

Table 6.2 lists selected applications of Dagum distributions to some 30 countries. Only works containing parameter estimates are included. There exist several further studies mainly concerned with goodness of fit that do not provide such information. A recent example is Azzalini *et al.* (2003) who fit the distribution to the 1997 data for 13 countries from the European Community Household Panel.

Of special interest are papers fitting several distributions to the same data, with an eye on relative performance. From comparative studies such as McDonald and Xu (1995), Bordley *et al.* (1996), Bandourian *et al.* (2003) and Azzalini *et al.* (2003) it emerges that the Dagum distribution typically outperforms its competitors, apart from the GB2 which has an extra parameter. Bandourian *et al.* (2003), find that, in a study utilizing 82 data sets, the Dagum is the best 3-parameter model in no less than 84% of the cases. From all these studies it would seem that empirically relevant values of the Dagum shape parameters are $a \in [2,7]$ and $p \in [0.1,1]$, approximately. Hence the implied income distributions are heavy-tailed admitting moments $E(X^k)$ for $k \leq 7$ while negative moments may exist up to order 7 in some examples.

For reasons currently not fully understood, the Dagum often provides a better fit to income data than the closely related Singh-Maddala distribution. Kleiber (1996) provides a heuristic explanation arguing that in the Dagum case the upper tail is determined by the parameter a while the lower tail is governed by the product ap, for the Singh-Maddala distribution the situation is reversed. Thus the Dagum distribution has one extra parameter in the region where the majority of the data are, an aspect that may to some extent explain the excellent fit of this model.

The previously mentioned works typically consider large populations, say households of particular countries. In an interesting contribution, Pocock *et al.* (2003) estimate salary distributions for different professions (specifically, the salaries of statistics professors at different levels) from sparse data utilizing the Dagum distribution (under the name of Burr III). This is of interest for competitive salary offers as well as for determining financial incentives for retaining valued employees. One of the few applications to wealth data, and at the same time one of the few applications of the Dagum type III distributions, is provided by Jenkins and Jäntti (2005) who estimate mixtures of Dagum distributions using wealth data for Finland.

Researchers have also begun to model conditional distributions in a regression framework, recent examples are Biewen and Jenkins (2005) and Quintano and D'Agostino (2006).

During the last decade, Camilo Dagum furthermore attempted to obtain information on the distribution of human capital, an example utilizing US data is Dagum and Slottje (2000) while the paper by Martín Reyes *et al.* (2001) mentioned above considers Spanish data.

In addition to all these empirical applications, the excellent fit provided by the distribution has also led to an increasing use in simulation studies. Recent examples include Hasegawa and Kozumi (2003), who consider Bayesian estimation of Lorenz curves, and Cowell and Victoria-Feser (2006), who study the effects of trimming on distributional dominance, both groups of authors utilize Dagum samples for illustrations. Also, Palmitesta *et al.* (1999, 2000) investigate improved finite-sample confidence intervals for inequality measures using Gram-Charlier series and bootstrap methods, respectively. Their methods are illustrated using Dagum samples. There even exist occasional illustrations in economic theory such as Glomm and Ravikumar (1998). Finally, there are numerous applications of this multi-discovered distribution in many fields of science and engineering (typically under the name of Burr III distribution), a fairly recent example from geophysics explicitly citing Dagum (1977) is Clark *et al.* (1999).

8 Concluding Remarks

This Chapter has provided a brief introduction to the Dagum distributions and their applications in economics. Given that the distribution only began to appear in the English language literature in the 1990s, it is safe to predict that there will be many further applications. On the methodological side, there are still some unresolved issues including aspects of likelihood inference. When the distribution celebrates its golden jubilee in economics, these problems no doubt will be solved.

Acknowledgements

I am grateful to Duangkamon Chotikapanich for her invitation to contribute to this memorial volume and to the late Camilo Dagum for providing me, for almost ten years, with a large number of reprints, many of which were used in the preparation of this Chapter.

References

Atkinson, A. B. (1970) On the Measurement of Inequality, *Journal of Economic Theory*, **2**, 244–263.

Azzalini, A. T., T. dal Cappello and S. Kotz (2003) Log-skew-normal and Log-skew-t Distributions as Models for Family Income Data, *Journal of Income Distribution*, **11**, 12–20.

Bandourian, R., J. B. McDonald and R. S. Turley (2003) A Comparison of Parametric Models of Income Distribution across Countries and over Time, *Estadística*, **55**, 135–152.

Bantilan, M. C. S., N. B. Bernal, M. M. de Castro and J. M. Pattugalan (1995) Income Distribution in the Philippines, 1957-1988: An Application of the Dagum Model to the Family Income and Expenditure Survey (FIES) Data, in C. Dagum and A. Lemmi (eds.) *Research on Economic Inequality, Vol. 6: Income Distribution, Social Welfare, Inequality and Poverty*, Greenwich, CT: JAI Press, 11–43.

Biewen, M. and S. P. Jenkins (2005) A Framework for the Decomposition of Poverty Differences with an Application to Poverty Differences between Countries, *Empirical Economics*, **30**, 331–358.

Blayac, T. and D. Serra (1997) Tarifs Publics et Redistribution Spatiale. Une application aux Transports Ferroviaires, *Revue d'Economie Régionale et Urbaine*, **4**, 603–618.

Bordley, R. F., J. B. McDonald and A. Mantrala (1996) Something New, Something Old: Parametric Models for the Size Distribution of Income, *Journal of Income Distribution*, **6**, 91–103.

Botargues, P. and D. Petrecolla (1999a) Funciones de Distribución del Ingreso y Afluencia Económica Relativa para Ocupados Según nivel de Educación en GBA, Argentina, 1992-1996, in M. Cardenas Santa Maria and N. Lustig (eds.) *Pobreza y desigualdad en América Latina*, Santafé de Bogotá, D.C., Fedesarrollo, Lacea, Colciencias, Tercer Mundo. Also Documento de Trabajo Instituto Torcuato Di Tella, DTE 216.

Botargues, P. and D. Petrecolla (1999b) Estimaciones Paramétricas y no Paramétricas de la Distribución del Ingreso de los Ocupados del Gran Buenos Aires, 1992-1997, *Economica (National University of La Plata)*, **XLV (No 1)**, 13–34.

Burr, I. W. (1942) Cumulative Frequency Functions, *Annals of Mathematical Statistics*, **13**, 215–232.

Campano, F. (1991) Recent Trends in US Family Income Distribution: A Comparison of All, White, and Black Families, *Journal of Post-Keynesian Economics*, **13**, 337–350.

Cheli, B., A. Lemmi and C. Spera (1995) An EM Algorithm for Estimating Mixtures of Dagum's Models, in C. Dagum and A. Lemmi (eds.) *Research on Economic Inequality, Vol. 6: Income Distribution, Social Welfare, Inequality and Poverty*, Greenwich, CT: JAI Press, 131-142.

Chotikapanich, D. and W. E. Griffiths (2006) Bayesian Assessment of Lorenz and Stochastic Dominance in Income Distributions, in J. Creedy and G. Kalb (eds.) *Research on Economic Inequality, Vol 13: Dynamics of Inequality and Poverty*, pp. 297–321, Elsevier, Amsterdam.

Clark, R. M., S. J. D. Cox and G. M. Laslett (1999) Generalizations of Power-law Distributions Applicable to Sampled Fault-trace Lengths: Model Choice, Parameter Estimation and Caveats, *Geophysical Journal International*, **136**, 357–372.

Cowell, F. A. and M.-P. Victoria-Feser (2006) Distributional Dominance with Trimmed Data, *Journal of Business & Economic Statistics*, **24**, 291–300.

D'Addario, R. (1949) Richerche sulla Curva dei Redditi, *Giornale degli Economisti e Annali di Economia*, **8**, 91–114.

Dagum, C. (1975) A Model of Income Distribution and the Conditions of Existence of Moments of Finite Order, *Bulletin of the International Statistical Institute*, **46**, 199–205, Proceedings of the 40th Session of the ISI, Warsaw, Contributed Papers.

Dagum, C. (1977) A New Model for Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, **30**, 413–437.

Dagum, C. (1980a) Generating Systems and Properties of Income Distribution Models, *Metron*, **38**, 3–26.

Dagum, C. (1980b) Sistemas Generadores de Distribución de Ingreso y la ley de Pareto, *El Trimestre Economico*, **47**, 877–917, reprinted in *Estadística*, 35 (1981), 143–183.

Dagum, C. (1980c) The Generation and Distribution of Income, the Lorenz Curve and the Gini Ratio, *Economie Appliquée*, **33**, 327–367.

Dagum, C. (1983) Income Distribution Models, in S. Kotz, N. L. Johnson and C. Read (eds.) *Encyclopedia of Statistical Sciences*, vol. 4, John Wiley, New York.

Dagum, C. (1985) Analysis of Income Distribution and Inequality by Education and Sex in Canada, in R. L. Basmann and G. F. Rhodes, Jr (eds.) *Advances in Econometrics*, vol. 4, pp. 167–227.

Dagum, C. (1990) Generation and Properties of Income Distribution Function, in C. Dagum and M. Zenga (eds.) *Income and Wealth Distribution, Inequality and Poverty: Proceedings of the Second International Conference on Income Distribution by Size: Generation, Distribution, Measurement and Applications.*, pp. 1–17, Springer, New York - Berlin - London - Tokyo.

Dagum, C. (1996) A Systemic Approach to the Generation of Income Distribution Models, *Journal of Income Distribution*, **6**, 105–126.

Dagum, C. (1999) Linking the Functional and Personal Distributions of Income, in J. Silber (ed.) *Handbook on Income Inequality Measurement*, pp. 101–128, Kluwer, Boston - Dordrecht - London.

Dagum, C. and K. Chiu (1991) Users Manual for the Program "EPID" (Econometric Package for Income Distribution) for Personal Computers. Statistics Canada/Statistique Canada: Time Series Research and Analysis Division.

Dagum, C., F. Guibbaud-Seyte and M. Terraza (1995) Analyse Interrégionale des Distributions des Salaires Français, *Economie Appliquée*, **48**, 103–133.

Dagum, C. and A. Lemmi (1989) A Contribution to the Analysis of Income Distribution and Income Inequality and a Case Study: Italy, *Research on Economic Inequality*, **1**, 123–157.

Dagum, C. and D. J. Slottje (2000) A New Method to Estimate the Level and Distribution of Household Human Capital with Application, *Structural Change and Economic Dynamics*, **11**, 67–94.

Dancelli, L. (1986) Tendenza alla Massima ed alla Minima Concentrazione nel Modello di Distribuzione del Reddito Personale di Dagum, in *Scritti in Onore di Francesco Brambilla*, vol. 1, Ed. Bocconi Comunicazioni, Milano, 249–267.

Domański, C. and A. Jedrzejczak (1998) Maximum Likelihood Estimation of the Dagum Model Parameters, *International Advances in Economic Research*, **4**, 243–252.

Domański, C. and A. Jedrzejczak (2002) Income Inequality Analysis in the Period of Economic Transformation in Poland, *International Advances in Economic Research*, **8**, 215–220.

Domma, F. (1997) Mediana e Range Campionario per il Modello di Dagum, *Quaderni di Statistica e Matematica Applicata alle Scienze Economico-Sociali*, **19**, 195–204.

Domma, F. (2002) L'andamento della *Hazard Function* nel Modello di Dagum a tre Parametri, *Quaderni di Statistica*, **4**, 1–12.

Espinguet, P. and M. Terraza (1983) Essai d'Extrapolation des Distributions de Salaires Français, *Economie Appliquée*, **36**, 535–561.

Falcão Carneiro, J. (1982) Modelo de Dagum de Distribuição Pessoal do Rendimento: Uma Aplicação às Receitas Familiares em Portugal, *Análise Social*, **18**, 231–243.

Fattorini, L. and A. Lemmi (1979) Proposta di un modello Alternativo per L'analisi della Distribuzione Personale del Reddito, *Atti Giornate di Lavoro AIRO*, **28**, 89–117.

Fisk, P. R. (1961) The Graduation of Income Distributions, *Econometrica*, **29**, 171–185.

Gibrat, R. (1931) *Les Inégalités Économiques*, Librairie du Recueil Sirey, Paris.

Gini, C. (1914) Sulla Misura della Concentrazione e della Variabilità dei Caratteri, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, **73**, 1203–1248.

Glomm, G. and B. Ravikumar (1998) Opting out of Publicly Provided Services: A Majority Voting Result, *Social Choice and Welfare*, **15**, 187–199.

Hasegawa, H. and H. Kozumi (2003) Estimation of Lorenz Curves: A Bayesian Nonparametric Approach, *Journal of Econometrics*, **115**, 277–291.

Jenkins, S. P. (1999) Fitting Singh-Maddala and Dagum Distributions by Maximum Likelihood, in *Stata Technical Bulletin, 48, 19-25*, also in *Stata Technical Bulletin Reprints*, vol. 8, 261-268. College Station, TX: Stata Press.

Jenkins, S. P. (2007) Inequality and the GB2 Income Distribution. Working Paper 2007-12. Colchester: Institute for Social and Economic Research, University of Essex.

Jenkins, S. P. and M. Jäntti (2005) Methods for Summarizing and Comparing Wealth Distributions, ISER Working Paper 2005-05. Colchester: University of Essex, Institute for Social and Economic Research.

Kleiber, C. (1996) Dagum vs. Singh-Maddala Income Distributions, *Economics Letters*, **53**, 265–268.

Kleiber, C. (1999) On the Lorenz Order within Parametric Families of Income Distributions, *Sankhyā B*, **61**, 514–517.

Kleiber, C. (2007) On the Zenga Order within Parametric Families of Income Distributions, Working paper, Universität Basel, Switzerland.

Kleiber, C. (2008) The Lorenz Curve in Economics and Econometrics, in G. Betti and A. Lemmi (eds.) *Advances on Income Inequality and Concentration Measures. Collected Papers in Memory of Corrado Gini and Max O. Lorenz*, Routledge, London.

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

Klonner, S. (2000) The First-Order Stochastic Dominance Ordering of the Singh-Maddala Distribution, *Economics Letters*, **69**, 123–128.

Kotz, S., N. L. Johnson and C. Read (eds.) (1983) *Encyclopedia of Statistical Sciences*, vol. 4, John Wiley, New York.

Latorre, G. (1988) Proprietà Campionarie del Modello di Dagum per la Distribuzione dei Redditi, *Statistica*, **48**, 15–27.

Lorenz, M. O. (1905) Methods of Measuring the Concentration of Wealth, *Quarterly Publications of the American Statistical Association*, **9 (New Series, No. 70)**, 209–219.

Łukasiewicz, P. and A. Orłowski (2004) Probabilistic Models of Income Distributions, *Physica A*, **344**, 146–151.

Majumder, A. and S. R. Chakravarty (1990) Distribution of Personal Income: Development of a New Model and Its Application to US Income Data, *Journal of Applied Econometrics*, **5**, 189–196.

Martín Reyes, G., A. Fernández Morales and E. Bárcena Martí (2001) Estimación de una Función Generadora de la Renta Mediante un Modelo de Variables Latentes, *Estadística Española*, **43**, 63–87.

McDonald, J. B. (1984) Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52**, 647–663.

McDonald, J. B. and A. Mantrala (1995) The Distribution of Income: Revisited, *Journal of Applied Econometrics*, **10**, 201–204.

McDonald, J. B. and Y. J. Xu (1995) A Generalization of the Beta Distribution with Applications, *Journal of Econometrics*, **66**, 133–152, Erratum: Journal of Econometrics, **69**: 427-428.

Mielke, P. W. (1973) Another Family of Distributions for Describing and Analyzing Precipitation Data, *Journal of Applied Meteorology*, **12**, 275–280.

Mielke, P. W. and E. S. Johnson (1974) Some Generalized Beta Distributions of the Second Kind having Desirable Application Features in Hydrology and Meteorology, *Water Resources Research*, **10**, 223–226.

Palmitesta, P., C. Provasi and C. Spera (1999) Approximated Distributions of Sampling Inequality Indices, *Computational Economics*, **13**, 211–226.

Palmitesta, P., C. Provasi and C. Spera (2000) Confidence Interval Estimation for Inequality Indices of the Gini Family, *Computational Economics*, **16**, 137–147.

Panjer, H. H. (2006) *Operational Risks*, John Wiley, Hoboken, NJ.

Pareto, V. (1895) La Legge della Domanda, *Giornale degli Economisti*, 10, 59-68. English Translation in Rivista di Politica Economica, 87 (1997), 691–700 .

Pareto, V. (1896) La Courbe de la Répartition de la Richesse, Reprinted 1965 in G. Busoni (ed.): Œeuvres Complètes de Vilfredo Pareto, Tome 3: Écrits sur la Courbe de la Répartition de la Richesse, Geneva: Librairie Droz. English translation in *Rivista di Politica Economica*, 87 (1997), 645–700.

Pareto, V. (1897) *Cours d'Economie Politique*, Rouge, Lausanne.

Pocock, M. L., J. B. McDonald and C. L. Pope (2003) Estimating Faculty Salary Distributions: An Application of Order Statistics, *Journal of Income Distribution*, **11**, 43–51.

Polisicchio, M. (1990) Sulla Interpretazione dei Parametri di Modelli Analitici per la Distribuzione del Reddito Personale, *Statistica*, **50**, 383–397.

Quintano, C. and A. D'Agostino (2006) Studying Inequality in Income Distribution of Single-Person Households in Four Developed Countries, *Review of Income and Wealth*, **52**, 525–546.

R Development Core Team (2007) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rodriguez, R. N. (1983) Burr Distributions, in S. Kotz and N. L. Johnson (eds.) *Encyclopedia of Statistical Sciences, Vol 1*, 335-340, John Wiley, New York.

Saposnik, R. (1981) Rank Dominance in Income Distributions, *Public Choice*, **36**, 147–151.

Shao, Q. (2002) Maximum Likelihood Estimation for Generalised Logistic Distributions, *Communications in Statistics: Theory and Methods*, **31**, 1687–1700.

Singh, S. K. and G. S. Maddala (1976) A Function for the Size Distribution of Incomes, *Econometrica*, **44**, 963–970.

Stoppa, G. (1995) Explicit Estimators for Income Distributions, in C. Dagum and A. Lemmi (eds.) *Research on Economic Inequality, Vol. 6: Income Distribution, Social Welfare, Inequality and Poverty*, Greenwich, CT: JAI Press, 393-405.

Tadikamalla, P. R. (1980) A Look at the Burr and Related Distributions, *International Statistical Review*, **48**, 337–344.

Venables, W. N. and B. D. Ripley (2002) *Modern Applied Statistics with S*, 4th ed., Springer, New York.

Victoria-Feser, M.-P. (1995) Robust Methods for Personal Income Distribution Models with Applications to Dagum's Model, in C. Dagum and A. Lemmi (eds.) *Research on Economic Inequality, Vol. 6: Income Distribution, Social Welfare, Inequality and Poverty*, Greenwich, CT: JAI Press, 225-239.

Victoria-Feser, M.-P. (2000) Robust Methods for the Analysis of Income Distribution, Inequality and Poverty, *International Statistical Review*, **68**, 277–293.

Yee, T. W. (2007) *VGAM: Vector Generalized Linear and Additive Models*, R package version 0.7-5. `http://www.stat.auckland.ac.nz/~yee/VGAM`.

Zelterman, D. (1987) Parameter Estimation in the Generalized Logistic Distribution, *Computational Statistics & Data Analysis*, **5**, 177–184.

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*, **48**, 301–326.

Chapter 7

Pareto and Generalized Pareto Distributions

Barry C. Arnold†

Abstract

More than one hundred years after its introduction, Pareto's proposed model for fitting income distributions continues to be heavily used. A variety of generalizations of this model have been proposed including discrete versions, together with natural multivariate extensions. Several stochastic scenarios can be used to justify the prevalence of income distributions exhibiting approximate Paretian behavior. This chapter will provide a survey of results related to these Pareto-like models including discussion of related distributional and inferential questions. Topics will include the classical Pareto models and its generalizations, stochastic income models leading to Paretian income distributions, distributional properties of generalized Pareto distributions, related discrete distributions, inequality measures for Paretian models, inferential issues and multivariate extensions.

1 The Classical Pareto Model

Pareto (1897) observed that in many populations the income distribution was one in which the number of individuals whose income exceeded a given level x could be approximated by $Cx^{-\alpha}$ for some choice of C and α. More specifically, he observed that such an approximation seemed to be appropriate for large incomes, i.e. for x's above a certain threshold. If one, for various values of x, plots the logarithm of the income level against the number of individuals whose income exceeds that level, Pareto's insight suggests that an approximately linear plot will be encountered. Empirical experience over the last century has buttressed this belief. It is however remarkable how scanty was the actual evidence published by Pareto to justify this claim. Income distributions with upper tails decreasing at a polynomial rate are

† Department of Statistics, University of California, Riverside, U.S.A., email: barry.arnold@ucr.edu

indeed commonly encountered. Consequently variants of Pareto's model are well accepted as default models for income distributions absent strong evidence to the contrary. The ubiquitous role of the Pareto laws in the study of income and other size distributions is somewhat akin to the central role played by the normal distribution in many experimental sciences. In both settings, plausible stochastic arguments can be advanced in favor of the models, but probably the deciding factor is that the models are analytically tractable and do seem to adequately "fit" observed data in many cases. In Section 3, we will mention a few of the stochastic income models that have been suggested as possible explanations of Pareto-tail phenomena. Whether any of these models are compelling or not will be a subjective decision on the part of the reader. Nevertheless, history has confirmed a central role for Pareto models in income theory, thus justifying a careful study of the distributional and inferential aspects of the popular variants of Vilfredo Pareto's models. To this we now turn.

2 Variations on the Basic Pareto Theme

We will follow Arnold (1983) in setting up a hierarchy of Pareto models of increasing complexity, beginning with the classical Pareto model which will be called a Pareto (I) distribution.

A random variable X will be said to have a Pareto (I) distribution if its survival function is of the form

$$\bar{F}_X(x) = P(X > x) = (x/\sigma)^{-\alpha}, \qquad x > \sigma \tag{7.1}$$

where σ is a (positive) scale parameter and α is a positive slope parameter. The parameter α is sometimes called Pareto's index. It corresponds to the negative of the slope of what is called a Pareto income chart, a plot of $\log \overline{F}_X(x)$ vs. $\log x$. If a random variable X has a survival function of the form (7.1) we will write $X \sim P(I)(\sigma, \alpha)$.

Remark: If a random variable X has survival function (7.1) then it is readily verified that the random variable $Y = \log X$ has a translated exponential distribution, i.e. $\log X$ admits the representation $\log \sigma + U$ where U has an exponential (α) distribution. This observation allows us to borrow freely from the rich distributional and inferential literature regarding exponential random variables when discussing properties of the classical Pareto or Pareto (I) model. See for example, the discussion following equation (7.13) below.

The introduction of a location parameter leads us to the Pareto (II) model. Thus X has a Pareto (II) distribution and we write $X \sim P(II)(\mu, \sigma, \alpha)$ if

$$\bar{F}_X(x) = P(X > x) = [1 + (\frac{x - \mu}{\sigma})]^{-\alpha}, \qquad x > \mu \tag{7.2}$$

where $\mu \in \mathbf{R}, \sigma \in \mathbf{R}^+$ and $\alpha \in \mathbf{R}^+$. Frequently μ is assumed to be positive but this is not essential. Note that parameter α controls the shape of the Pareto (II) density. It is sometimes described as an index of inequality. When $\mu = 0$, it does indeed qualify as a suitable parameter for ordering populations with regard to inequality. In that case most of the popular analytic measures of inequality are indeed monotone functions of α.

A distribution whose survival function mimics the tail behavior of the Pareto (II) survival function is

$$\bar{F}_X(x) = [1 + (\frac{x-\mu}{\sigma})^{1/\gamma}]^{-1}, \qquad x > \mu. \tag{7.3}$$

Here $\mu \in \mathbf{R}$ and $\sigma, \gamma \in \mathbf{R}^+$. If X has (7.3) as its survival function we will say that it has a Pareto (III) distribution and we will write $X \sim P(III)(\mu, \sigma, \gamma)$. We will call γ the index of inequality. When $\mu = 0$, γ is exactly equal to the Gini index of inequality.

Even more general models are available. The Pareto (IV) distribution has a survival function of the form

$$\bar{F}_X(x) = [1 + (\frac{x-\mu}{\sigma})^{1/\gamma}]^{-\alpha}, \qquad x > \mu \tag{7.4}$$

where $\mu \in \mathbf{R}$ and $\sigma, \gamma, \alpha \in \mathbf{R}^+$. All the models (7.1)-(7.4) exhibit the slowly varying tail behavior that Pareto noted as a characteristic of many empirical income data sets, i.e. in all cases $\bar{F}_X(x) \sim x^{-\delta}$ as $x \to \infty$.

The Pareto (III) distribution is sometimes called the log-logistic distribution since, if X has a logistic distribution then e^X has a Pareto (III) distribution (with $\mu = 0$). The Pareto (IV) density is included in Burr's (1942) catalog of frequency functions (his Type XII).

For resolving distributional questions regarding the Pareto (I) - (IV) distributions it is sometimes convenient to recognize them as all being special cases of what Arnold and Laguna (1977) call a Feller-Pareto distribution. They arrive at this distribution by using the following construction. Begin with a random variable Y which has a beta distribution with parameters γ_1 and γ_2, i.e.

$$f_Y(y) = y^{\gamma_1 - 1}(1-y)^{\gamma_2 - 1}/B(\gamma_1, \gamma_2), \qquad 0 < y < 1. \tag{7.5}$$

Feller (1971) then considered the distribution of the random variable $Y^{-1} - 1$ and called it the Pareto distribution. Arnold and Laguna then raised this random variable to a power and introduced a location and a scale parameter. Instead we will use an equivalent construction of a Feller-Pareto variable by beginning with two independent gamma random variables. Thus we take U_1 and U_2 to be independent random variables with $U_1 \sim \Gamma(\delta_1, 1)$ and $U_2 \sim \Gamma(\delta_2, 1)$. Then define

$$W = \mu + \sigma(\frac{U_1}{U_2})^{\gamma}. \tag{7.6}$$

A random variable defined as in (7.6) will be said to have a Feller-Pareto distribution and we will write $W \sim FP(\mu, \sigma, \gamma, \delta_1, \delta_2)$. (Note that in Arnold and Laguna (1977) and Arnold (1983) a slightly different parameterization is used, i.e. δ_1 is replaced by γ_2 and δ_2 by γ_1). Observe that all of the distributions in our hierarchy of Pareto models are special cases of the Feller-Pareto model. Thus:

$$\begin{aligned} P(I)(\sigma,\alpha) &= FP(\sigma,\sigma,1,1,\alpha) \\ P(II)(\mu,\sigma,\alpha) &= FP(\mu,\sigma,1,1,\alpha) \\ P(III)(\mu,\sigma,\gamma) &= FP(\mu,\sigma,\gamma,1,1) \\ P(IV)(\mu,\sigma,\gamma,\alpha) &= FP(\mu,\sigma,\gamma,1,\alpha) \end{aligned} \tag{7.7}$$

The representation (7.6) will prove to be very useful for deriving properties of the various Pareto models, I-IV. Note that (U_1/U_2) in (7.6) has a beta distribution of the second kind and, since a random variable with an F distribution has a scaled beta distribution of the second kind, it is reasonable to dub the Feller-Pareto distribution as a generalized F distribution. Indeed in the survival literature it is so christened usually with $\mu = 0$, see Kalbfleisch and Prentice (2002).

In the context of extreme value theory, especially in the study of peaks over thresholds, Pickands (1975) introduced what he called a generalized Pareto distribution. The corresponding density is

$$f(x;\sigma,k) = \frac{1}{\sigma}(1-\frac{kx}{\sigma})^{(1-k)/k} I(x>0, \frac{kx}{\sigma}<1) \tag{7.8}$$

where $\sigma > 0$ and $k \in \mathbf{R}$. The density corresponding to $k = 0$ is obtained by taking the limit as $k \uparrow 0$ in (7.8). In fact (7.8) includes three kinds of densities. When $k < 0$, it yields a Pareto (II) density (with $\mu = 0$), when $k = 0$ it yields an exponential density while for $k > 0$ it corresponds to a scaled beta distribution (of the first kind). The density (7.8) thus unifies 3 models that are of interest in peaks over threshold analysis. However for income modeling it will generally be true that only the case $k < 0$ (and perhaps $k = 0$) will be of interest. Nevertheless, the literature on Pickands' generalized Pareto distribution has become quite extensive and provides a useful source of information about the Pareto (II) distribution.

Before investigating distributional properties of the various Paretian models introduced in this section, we will digress and describe some of the stochastic models which have been suggested as possible explanations for the ubiquity of regularly-varying survival functions that Pareto drew to the attention of the world.

3 Some Income Distribution Models

Gibrat (1931) argued that the income of an individual is a stochastic process subject to small multiplicative fluctuations. This line of argument leads to a proposal of the lognormal distribution as a suitable income model. The major advantage of such a log-normal model is that, by a simple transformation, the enormous existing body of normal theory inference procedures becomes immediately available for the study of income. Indeed, the log-normal model continues to be a popular alternative to Pareto models for fitting income distributions. Recalling Pareto's observation that survival functions that are regularly varying at ∞ are typically encountered in such settings, it is appropriate to study the log-normal model from this viewpoint. The regularly varying survival function is characterized by the linearity of the Pareto chart (a plot of $\log \bar{F}_X(x)$ vs. $\log x$). Such plots for the lognormal distribution are curvilinear, but they might well be judged to be close enough to linear to satisfy Pareto (who was often quite easily satisfied in this regard).

In a series of publications Champernowne (1937, 1953, 1973) discussed a discrete time model for income in which income was restricted to take on values in some grid (say in units of hundreds of pounds, euros or dollars). Under the assumption that transitions from one income level to the next were governed by a Markov chain with steps down limited to no more than N income levels, he was able to show that the long run distribution was a discretized Pareto distribution, and he consequently argued that the continuous classical Pareto model could then be reasonably used as a simple approximation for fitting income data sets.

Ericson (1945) used a coin-shower mechanism to justify the use of an exponential distribution as a model for income data. Assuming that real populations are mixtures of relatively homogeneous sub-populations, the coin shower model can be used to argue for the use of mixtures of exponential distributions as income models. Since the Pareto (II) model (with $\mu = 0$) can be viewed as a scale mixture of exponential distributions (see (7.6) and (7.7)), the Ericson approach might be used to argue for use of the Pareto (II) model.

An excellent survey of various other income distribution models may be found in Kleiber and Kotz (2003). We mention only two more in the interest of brevity. Lydall (1959) suggested a pyramid structure of employment organization. Each individual at level i is viewed as supervising n employees at the next lowest level. His salary is assumed to be proportional to the aggregate income of his supervisees. This is shown to lead to a discrete Pareto distribution of income which will be well approximated by the classical model.

Arnold and Laguna (1976, 1977) use a model of competitive bidding for employment in which a random number of individuals apply for job openings and employers hire the individual requesting the lowest salary. The long run income distribution in this case is of the Pareto (II) (or log-logistic) form. It is interesting that, historically, the major competitor to Pareto models for fitting income data has been provided by the log-normal distribution. Typically, data that are well fitted by Pareto models are found to also be quite well described by a log-normal model. The obvious similarity between the normal and the logistic distributions, will of course imply an analogous similarity between the log-normal and the Paretian log-logistic distributions. From this viewpoint we will not be surprised by the difficulty of making a selection between Pareto and log-normal models.

While some would argue that we can live comfortably using log-normal models and do not need to use the various Pareto models, the same arguments can be used to justify using Pareto models and discarding log-normal models as being inessential. The Pareto models form a considerably richer family than do log-normal models and this may be a strong argument in their favor.

4 Distributional Properties of Pareto Distributions

We will refer to the hierarchy of Pareto distributions introduced in Section 2 as generalized Pareto distributions (with apologies to Pickands who limited use of this term to densities of the form (7.8)).

The hierarchy of generalized Pareto models to be discussed is

$$\begin{array}{lll} P(I) & \bar{F}(x;\sigma,\alpha) = (x/\sigma)^{-\alpha}, & \mu > \sigma, \\ P(II) & \bar{F}(x;\mu,\sigma,\alpha) = [1+(\frac{x-\mu}{\sigma})]^{-\alpha}, & x > \mu, \\ P(III) & \bar{F}(x;\mu,\sigma,\gamma) = [1+(\frac{x-\mu}{\sigma})^{1/\gamma}]^{-1}, & x > \mu, \\ P(IV) & \bar{F}(x;\mu,\sigma,\gamma,\alpha) = [1+(\frac{x-\mu}{\sigma})^{1/\gamma}]^{-\alpha}, & x > \mu, \end{array}$$

all of which can be viewed as special cases of the Feller-Pareto $(\mu,\sigma,\gamma,\delta_1,\delta_2)$ model given by

$$W = \mu + \sigma(\frac{U_1}{U_2})^{\gamma} \tag{7.9}$$

where U_1, U_2 are independent gamma random variables with shape parameters δ_1 and δ_2 respectively. The parametric values in (7.9) which lead to Pareto (I) - (IV) models were listed in (7.7).

Feller-Pareto random variables do not have simple expressions available for their moment generating functions though it is not difficult to obtain expressions for their moments using the representation (7.9). Feller-Pareto densities are always unimodal. The mode will be at μ if $\gamma > \delta_1$ and will be at $\mu + \sigma[\frac{\delta_1-\gamma}{\delta_2+\gamma}]^{\gamma}$ if $\gamma \le \delta_1$.

Recall that if $U \sim \Gamma(\alpha,1)$ then the τ-th moment of U is given by

$$E(U^{\tau}) = \Gamma(\tau+\alpha)/\Gamma(\alpha) \tag{7.10}$$

provided that $\tau + \alpha > 0$. Using this result we can readily obtain moments for a Feller-Pareto distributed random variable with, for simplicity, $\mu = 0$. Thus if W is as defined in (7.9), with $\mu = 0$, we have

$$E(W^{\tau}) = \sigma^{\tau}\frac{\Gamma(\delta_1+\tau\gamma)\Gamma(\delta_2-\tau\gamma)}{\Gamma(\delta_1)\Gamma(\delta_2)} \tag{7.11}$$

provided that $-\frac{\delta_1}{\gamma} < \tau < \frac{\delta_2}{\gamma}$. From this expression we may obtain moments for the Pareto (II) - (IV) distributions by making the substitutions indicated in (7.7).

Moments for the classical Pareto model (i.e. Pareto (I) (σ,α)) cannot be obtained from (7.11) (since for the Pareto (I) model, $\mu = \sigma \neq 0$) but elementary computations yield

$$\text{Pareto (I)} \quad E(X^\tau) = \sigma^\tau (1 - \tfrac{\tau}{\alpha})^{-1}, \qquad \tau < \alpha\,. \tag{7.12}$$

Convolutions of Pareto densities do not generally admit closed form expressions; at best series expansions can be exhibited. However such convolutions will have regularly varying tails and in this sense are asymptotically Paretian.

Products of independent classical Pareto (I) random variables do have tractable distributions. To verify this we recall an alternative description of the classical Pareto model that will prove useful in several subsequent discussions related to that distribution. If $X \sim P(I)(\sigma,\alpha)$ then

$$Y = \log X =^d \log\sigma + V/\alpha \tag{7.13}$$

where $V \sim \exp(1)(= \Gamma(1,1))$. Thus distributional (and inferential) questions about Pareto (I) variables can be reinterpreted as dealing with translated exponential random variables (or two parameter exponential variables, as they are sometimes called). Using this representation and well known results for sums of independent exponential random variables we can obtain the density for products of independent Pareto (I) random variables in two cases.

Case (i): If $X_1,\ldots,X_n$ are i.i.d. $P(I)(\sigma,\alpha)$ random variables and $Y = \prod_{i=1}^n X_i$, then

$$f_Y(y) = \frac{(\sigma \log(\frac{y}{\sigma}))^{n-1}(\frac{y}{\sigma})^{-\alpha}(\frac{\alpha}{y})}{\Gamma(n)} I(y > \sigma), \tag{7.14}$$

where $\sigma = \prod_{i=1}^n \sigma_i$.

Case (ii): If $X_1,\ldots,X_n$ are independent random variables with $X_i \sim P(I)(\sigma_i,\alpha_i)$ in which all the α_i's are distinct then, again with $Y = \prod_{i=1}^n X_i$, we have

$$f_Y(y) = \sum_{i=1}^n \frac{\alpha_i}{\sigma}(\frac{y}{\sigma})^{-(\alpha_i+1)} \prod_{k=1,k\neq i}^n (\frac{\alpha_k}{\alpha_i - \alpha_k}) I(y > \sigma), \tag{7.15}$$

with $\sigma = \prod_{i=1}^n \sigma_i$.

Inspection of the Pareto (IV) survival function in (7.4) makes it clear that minima of independent Pareto (IV) variables (with the same μ,σ and γ) will again have distributions of the Pareto (IV) form. Thus if $X_1,X_2,\ldots,X_n$ are independent with $X_i \sim P(IV)(\mu,\sigma,\gamma,\alpha_i), i = 1,2,\ldots,n$ then

$$\min(X_1,X_2,\ldots,X_n) \sim P(IV)(\mu,\sigma,\gamma,\sum_{i=1}^n \alpha_i). \tag{7.16}$$

4.1 Order Statistics

For a sample of size n from a Feller-Pareto distribution, we will denote the corresponding order statistics by $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$. Generally speaking, elegant expressions for the densities of these order statistics are not obtainable. One exception, previously noted, involves $P(IV)$ samples. If $X_1, X_2, \ldots, X_n$ are i.i.d. $P(IV)(\mu, \sigma, \gamma, \alpha)$ then $X_{1:n} \sim P(IV)(\mu, \sigma, \gamma, n\alpha)$. If we begin with a sample from a Pareto (III) (μ, σ, γ) distribution then the distribution of the i'th order statistic is tractable. We have $X_{i:n} \sim FP(\mu, \sigma, \gamma, i, n-i+1)$.

If we have observations from a classical Pareto (I) (σ, α) distribution then distributional properties of the corresponding order statistics can be readily derived using well known results for exponential random variables. We use the fact that for a sample from a translated exponential distribution ($Y_i = \beta_0 + \beta_1 Z_i$ in which the Z_i's are i.i.d. $\exp(1)$) the scaled spacings $(n-i+1)(Y_{i:n} - Y_{i-1:n})$ are themselves i.i.d. exponential random variables (here $Y_{0:n} = \beta_0$ by definition) (Sukhatme, 1937). This result allows us to derive distributional properties of ratios of classical Pareto order statistics. Thus for order statistics $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ from a Pareto (I) (σ, α) distribution we may consider, for $k_1 < k_2$, the ratio

$$R_{k_1,k_2:n} = X_{k_2:n}/X_{k_1:n}. \tag{7.17}$$

Using the Sukhatme (1937) result we have

$$R_{k_1,k_2:n} =^d \prod_{j=k_1+1}^{k_2} W_j \tag{7.18}$$

where the W_j's are independent random variables with $W_i \sim P(I)(1, (n-j+1)\alpha)$.

Referring to (7.15) it is then possible to write down the corresponding density of $R_{k_1,k_2:n}$. By introducing $X_{0:n} = \sigma$, by definition, the above analysis can be used to obtain the survival function (and hence the density function) of $X_{i:n}$ based on a classical Pareto sample. Since the Pareto (II) distribution is a simple translation of the classical Pareto model we can then obtain the distribution for $X_{i:n}$, the ith order statistic, in a sample from a $P(II)(\mu, \sigma, \alpha)$ distribution. Thus, in that case,

$$P(X_{i:n} > x) = \sum_{j=1}^{i} (1 + \frac{x-\mu}{\sigma})^{-(n-j+1)\alpha} \prod_{\ell=1,\ell\neq j}^{i} (\frac{n-\ell+1}{\ell-j}), \qquad x > \mu, \tag{7.19}$$

a linear combination of $P(II)$ survival functions.

Moments of order statistics can be computed in some special cases as follows.
(i) Pareto (I): If the X_i's are i.i.d. $P(I)(\sigma, \alpha)$ then $X_{i:n} =^d \sigma \prod_{j=1}^{i} W_{j:n}$ where the $W_{j:n}$'s are independent random variables with $W_{j:n} \sim P(I)(1, (n-j+1)\alpha)$. Consequently

$$E(X_{i:n}^{\tau}) = \sigma^{\tau} \prod_{j=1}^{i} (1 - \frac{\tau}{\alpha(n-j+1)})^{-1}$$
$$= \sigma^{\tau} \frac{n!}{(n-j)!} \frac{\Gamma(n-i+1-\tau\alpha^{-1})}{\Gamma(n+1-\tau\alpha^{-1})}. \qquad (7.20)$$

Means, variances and covariances can then be readily obtained from this expression. (ii) Pareto (II): If the X_i's are i.i.d. $P(II)(\mu, \sigma, \alpha)$ then $X_{i:n} =^d (\mu - \sigma) + Y_{i:n}$ where the $Y_{i:n}$'s are i.i.d. $P(I)(\sigma, \alpha)$ so that means and variances and covariances of $X_{i:n}$ can be obtained using (7.20).
(iii) Pareto (III): If the X_i's are i.i.d. $P(III)(\mu, \sigma, \gamma)$ Then $X_i =^d \mu + \sigma Z_i^{\gamma}$ where Z_i has what we may call a standard Pareto distribution with survival function

$$\bar{F}_Z(z) = (1+z)^{-1}, \qquad z > 0. \qquad (7.21)$$

The corresponding quantile function (or inverse distribution function) is

$$F_Z^{-1}(y) = y/(1-y), \qquad 0 < y < 1. \qquad (7.22)$$

It is then a straightforward matter to obtain the τ'th moment of the i'th order statistic from a sample of size n from the standard Pareto distribution (7.21). Thus

$$E(Z_{i:n}^{\tau}) = \int_0^1 (\frac{z}{1-z})^{\tau} z^{i-1} (1-z)^{n-i} dz / B(i, n+i+1)$$
$$= \frac{\Gamma(i+\tau)\Gamma(n-i-\tau+1)}{\Gamma(i)\Gamma(n-i+1)} \qquad (7.23)$$

provided that $i + \tau < n+1$. Using (7.23) and the representation $X_{i:n} =^d \mu + \sigma Z_{i:n}^{\gamma}$ we can obtain expressions for the means, variances and covariances of order statistics from a $P(III)(\mu, \sigma, \gamma)$ population.
Pareto IV: Although general expressions for the moments of Pareto (IV) order statistics are not available, there is a straightforward approach available for generating moments of such order statistics. We make use of the following identity (see e.g. Arnold *et al.* (1992, p. 112)).

$$E(X_{i:n}^{\tau}) = \sum_{r=n-i+1}^{n} (-1)^{r-n+i-1} \binom{n}{r} \binom{r-1}{n-i} E(X_{1:r}^{\tau}). \qquad (7.24)$$

Thus it is only necessary to compute moments of sample minima. If $X_1, \ldots, X_r$ are i.i.d. $P(IV)(\mu, \sigma, \gamma, \alpha)$ then $X_{1:r} \sim P(IV)(\mu, \sigma, \gamma, r\alpha)$, as was observed earlier, and consequently

$$E(X_{1:r}^{\tau}) = E((\mu + \sigma Z_{1:r})^{\tau}) \qquad (7.25)$$

where $Z_{1:r} \sim P(IV)(0, 1, \gamma, r\alpha)$. Referring to (7.11) we find

$$E(Z_{1:r}^{\tau}) = \frac{\Gamma(1+\tau\gamma)\Gamma(r\alpha - \tau\gamma)}{\Gamma(r\alpha)},$$

from which we may immediately obtain $E(X_{1:r}^{\tau})$ when $\mu = 0$, and also obtain readily the mean and variance of $X_{1:r}$ when $\mu \neq 0$.

4.2 Truncation From Below

For any distribution function F_X we may define the corresponding truncated version of F, truncated from below at x_0 and denoted by $F_X^{(x_0)}$, by

$$F_X^{(x_0)}(x) = P(X \leq x | X > x_0). \tag{7.26}$$

Such truncated distributions are naturally of interest in income distribution studies since much of the available data is so truncated. We will denote the corresponding survival function of $F_X^{(x_0)}$ by $\bar{F}_X^{(x_0)}$, i.e. $\bar{F}_X^{(x_0)}(x) = 1 - F_X^{(x_0)}(x)$. The Pareto (II) family of distributions is closed under truncation from below. To see this consider, for $X \sim P(II)(\mu, \sigma, \alpha)$ and $x > x_0 \geq \mu$

$$\begin{aligned} P(X > x | X > x_0) &= \frac{(1 + \frac{x-\mu}{\sigma})^{-\alpha}}{(1 + \frac{x_0-\mu}{\sigma})^{-\alpha}} \\ &= (1 + \frac{x - x_0}{\sigma + x_0 - \mu})^{-\alpha}. \end{aligned}$$

Thus if $X \sim P(II)(\mu, \sigma, \alpha)$ then $X | X > x_0 \sim P(II)(x_0, \sigma + x_0 - \mu, \alpha))$. It follows readily that

$$P(II) \qquad E(X | X > x_0) = x_0 + \frac{\sigma + x_0 - \mu}{\alpha - 1} \tag{7.27}$$

provided $x_0 \geq \mu$. In a reliability context, one would say that X has a linear mean residual life function (as a function of x_0).

The truncated versions of more general Pareto models (such as $P(III), P(IV)$ and FP) will not have recognizable forms. This fact together with the ubiquity of truncated data sets could be used to argue for use of only the $P(II)$ (including the classical Pareto) model for fitting income data. However the limited flexibility of the Pareto (II) model must be recognized. Arguments against use of the Pareto (II) as a model might be based on the fact that the corresponding Gini index (see Section 6) is somewhat limited in its range.

4.3 Record Values

Suppose that $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables with common continuous distribution function F. An observation X_j will be called a record value if it is larger than all X's that precede it in the sequence. By convention, X_1 is a record value (the zero'th record value). The corresponding sequence of record values will be denoted by $X_{(0)}, X_{(1)}, \ldots$. It is convenient to denote with asterisks a sequence $X_1^*, X_2^*, \ldots$ of i.i.d. exponential (1) random variables with corresponding record values $X_{(0)}^*, X_{(1)}^*, \ldots$. We define a transformation Ψ_F by

$$\Psi_F(u) = F^{-1}(1 - e^{-u}). \tag{7.28}$$

It is not difficult to verify that

$$X_{(n)} =^d \Psi_F(X_{(n)}^*), n = 0, 1, \ldots. \tag{7.29}$$

In the particular case in which the X's have a Pareto(II) (μ, σ) distribution we will have

$$F^{-1}(u) = (\mu - \sigma) + \sigma(1 - u)^{-1/\alpha}$$

and so the n'th record value can be represented in the form

$$X_{(n)} =^d (\mu - \sigma) + \sigma e^{X_{(n)}^*/\alpha}. \tag{7.30}$$

However it is readily verified that $X_{(n)}^* =^d \sum_{j=0}^n X_i^*$ and that $e^{X_1^*/\alpha} \sim P(I)(1, \alpha)$. Consequently, $X_{(n)}$ admits the representation

$$X_{(n)} =^d (\mu - \sigma) + \sigma \prod_{j=0}^{n} Y_j \tag{7.31}$$

where the Y_j's are i.i.d. $P(I)(1, \alpha)$ random variables. Distributional questions about Pareto (II) record values are usually most simply resolved by using the representation (7.30) or (7.31).

If one considers the Pareto (IV) distribution, it transpires that only in the case when $\gamma = 1$ (i.e. the Pareto (II) case) does the inverse distribution function assume a convenient form, and thus there is no attractive available representation analogous to (7.30), except in the Pareto (II) case.

4.4 Characterizations

Because of the intimate relationship between the classical Pareto distribution and the standard exponential distribution, there are numerous characterizations of the Pareto distribution (in particular of the classical Pareto distribution) available in the literature. We will not attempt a complete list but will briefly mention some of the characterizations which might be meaningful in an income distribution context.

4.4.1. Truncation equivalent to rescaling

If $X \sim P(I)(\sigma,\alpha)$ then for $x_0 > \sigma$ we have $X|X > x_0 \sim P(I)(x_0,\alpha)$ so that truncation is equivalent to rescaling i.e. $X|X > x_0 =^d \frac{x_0}{\sigma}X$. If this is assumed to hold true for every $x_0 > \sigma$, the result can be rephrased as being equivalent to a lack of memory property for $\log X$ which implies an exponential distribution for $\log X$ and thus a $P(I)$ distribution for X.

4.4.2. An underreported income scenario

Let Y denote actual income and X denote reported income. Let us assume a multiplicative reporting error model so that

$$X = RY \tag{7.32}$$

where R and Y are independent and $0 \le R \le 1$. Also assume that $P(RX > x_0) > 0$. Suppose that the distribution of $Y = RX$ truncated below at x_0 is the same as the distribution of X. Krishnaji (1970) in this setting assumed that R has density

$$f_R(r) = \delta r^{\delta-1}, \qquad 0 < r < 1 \tag{7.33}$$

and was led to a differential equation for $\bar{F}_X$ whose solution corresponds to a Pareto (II) distribution for X. In fact, the assumption of the density (7.33) for R is not crucial. The result holds for a quite general class of distributions for R. The problem can be translated to one involving the integrated Cauchy functional equation (Huang, 1978).

4.4.3. Geometric minimization

If $X_1, X_2, \ldots$ are i.i.d. $P(III)(\mu,\sigma,\gamma)$ and if N, independent of the X_i's, has a geometric distribution (i.e. $P(N = n) = p(1-p)^{n-1}, n = 1,2,\ldots$) and if we define

$$Y = \min_{1 \le i \le N} X_i \tag{7.34}$$

then we may verify that, for $y > \mu$,

$$\begin{aligned} P(Y > y) &= \sum_{n=1}^{\infty} P(Y > y|N = n)P(N = n) \\ &= \sum_{n=1}^{\infty} [1 + (\frac{y-\mu}{\sigma})^{1/\gamma}]^{-n} p(1-p)^{n-1} \\ &= [1 + (\frac{y-\mu}{\sigma p^{\gamma}})^{1/\gamma}]^{-1}. \end{aligned}$$

Thus

$$Y \sim P(III)(\mu, \sigma p^{\gamma}, \gamma). \tag{7.35}$$

In particular if $\mu = 0$, we have

$$p^{-\gamma} Y =^{d} X_1. \tag{7.36}$$

Thus if the X_i's have a $P(III)(0, \sigma, \gamma)$ distribution then Y defined by (7.34) has a distribution that is the same as X_1 except for a change of scale. In this setting, the condition $Y =^{d} cX_1$ plus a mild regularity condition that $\lim_{x\to 0} x^{-\lambda} F(x) = \eta > 0$, will guarantee that the X_i's must have a Pareto (III) distribution with $\mu = 0$. (Arnold and Laguna (1976)). Without the regularity condition, the X_i's can have what is known as a semi-Pareto distribution (such distributions are discussed in Pillai (1991)).

4.4.4. Some failed "characterizations"

First we discuss a failure involving order statistics. If $X_1, \ldots, X_n$ are i.i.d. Pareto (I) (σ, α) random variables then for any $k \in \{1, 2, \ldots, n-1\}$ we may readily verify that

$$X_{k+1:n}/X_{k:n} \sim P(I)(1, \alpha). \tag{7.37}$$

Is this a characteristic property of the classical Pareto distribution? If (7.37) is assumed to hold for just one value of k, then we cannot conclude that the X_i's have a Pareto (I) distribution. To see this let Y_1, Y_2, Y_3, Y_4 be i.i.d. $\Gamma(\frac{1}{2}, 1)$ random variables and define U_1, U_2 and V_1, V_2 by

$$\begin{aligned} U_1 &= Y_1 + Y_2, \qquad & U_2 &= Y_3 + Y_4 \\ V_1 &= Y_1 - Y_2, \qquad & V_2 &= Y_3 - Y_4. \end{aligned}$$

Now U_1 and U_2 are exponential random variables while V_1 and V_2 are not. Define $X_1 = e^{U_1}, X_2 = e^{U_2}, \widetilde{X}_1 = e^{V_1}, \widetilde{X}_2 = e^{V_2}$. Now X_1, X_2 are i.i.d. $P(I)(1,1)$ random variables while $\widetilde{X}_1$ and $\widetilde{X}_2$ are clearly not Pareto variables. However

$$\begin{aligned} \widetilde{X}_{2:2}/\widetilde{X}_{1:2} &= e^{V_{2:2} - V_{1:2}} =^{d} e^{|V_1 - V_2|} \\ &=^{d} e^{|U_1 - U_2|} =^{d} e^{U_{2:2} - U_{1:2}} = X_{2:2}/X_{1:2} \\ &\sim P(I)(1,1). \end{aligned}$$

Now let us turn to a putative characterization based on record values. We have observed that if $X_1, X_2, \ldots$ are i.i.d. Pareto (I) (σ, α) random variables then the "geometric" record value spacings $X_{(k)}/X_{(k-1)}$ have the property that for $k = 1, 2, \ldots$

$$X_{(k)}/X_{(k-1)} \sim P(I)(1, \alpha). \tag{7.38}$$

If (7.38) holds for one value $k \in \{1,2,\dots\}$, can we conclude that the X_i's are Pareto (I) variables? Here too, the answer is no. Houchens (1984) showed that if we begin with $U_1, U_2, \dots$ as a sequence of i.i.d. Gumbel random variables, i.e.

$$U_i =^d \mu + \sigma \log X_i^*$$

Where, as usual, the X_i^*'s are i.i.d. exponential(1) variables, then the corresponding record values admit the representation

$$U_{(n)} =^d \mu + \sigma \log(\sum_{i=0}^{n} X_i^*) \tag{7.39}$$

and consequently the record spacing $U_{(n)} - U_{(n-1)}$ can be shown to have an exponential distribution. So, if we define $\widetilde{X}_n = e^{U_n}$ we will find that although the $\widetilde{X}_n$'s do not have a Pareto distribution, it is indeed true that $\widetilde{X}_{(k)}/\widetilde{X}_{(k-1)} \sim P(I)(\sigma, \alpha_k)$.

4.5 Asymptotics

Intermediate order statistics from Pareto (IV) samples will be asymptotically normal, thus

$$P(IV) \;\; X_{k:n} \sim AN(F^{-1}(\tfrac{k}{n}), [\tfrac{k}{n}(1-\tfrac{k}{n})/nf^2(F^{-1}(\tfrac{k}{n}))]) \tag{7.40}$$

where f and F denote the Pareto (IV) density and distribution function respectively and so

$$F^{-1}(u) = \mu + \sigma[(1-u)^{-1/\alpha} - 1]^{\gamma}. \tag{7.41}$$

In the classical Pareto (I) (σ, α) case we have

$$X_{k:n} \sim AN(\sigma(1-\frac{k}{n})^{-1/\alpha}, \frac{k\sigma^2}{n^2\alpha^2}(1-\frac{k}{n})^{-(1+2\alpha^{-1})}) \tag{7.42}$$

The asymptotic distributions for the extremes of Pareto (IV) samples are readily obtainable. We find, for the minimum,

$$\lim_{n\to\infty} P((\alpha n)^{\gamma}(X_{1:n} - \mu)/\sigma > z) = e^{-z^{1/\gamma}}, z > 0 \tag{7.43}$$

(a Weibull limit distribution). For the maximum,

$$\lim_{n\to\infty} P(n^{-\gamma/\alpha}(X_{n:n} - \mu)/\sigma \le z) = e^{-z^{(-\alpha/\gamma)}}, z > 0 \tag{7.44}$$

(a Frechet extreme value limit distribution).

If $S_n = \sum_{i=1}^{n} X_i$ where the X_i's are i.i.d. $P(IV)(\mu, \sigma, \gamma, \alpha)$ random variables then, provided $\alpha/\gamma > 2$, the central limit theorem will apply and

$$(S_n - E(S_n))/\sigma(S_n) \to^d N(0,1)$$

If $\alpha/\gamma = 2$ asymptotic normality is also encountered if $\sigma(S_n)$ (which is not defined when $\alpha/\gamma = 2$) is replaced by $n\log n$. For $\alpha/\gamma < 2$, suitable normalization of S_n will lead to a stable limit distribution with characteristic exponent α/γ.

The Pareto (III) distribution can itself arise as a limit distribution under repeated geometric minimization (Arnold and Laguna (1976)) while the Pareto (II) distribution can arise as a limiting distribution when one considers residual life at great age (Balkema and de Haan, 1974).

5 A Note on Related Discrete Distributions

Quite frequently income data is available only in grouped form. For this reason it is of interest to study discretized versions of the Pareto distributions discussed in this chapter. Without loss of generality (simply redefine the units of measurement if necessary) we can focus on random variables that can be identified as the integer parts of Pareto variables. Arnold (1983) suggests that these be called Zipf distributions in honor of George Zipf who, in Zipf (1949), described a wide variety of settings in which such distributions appear to arise naturally. We will say that a random variable X has a Zipf(IV) $(k_0, \sigma, \gamma, \alpha)$ distribution if it has possible values $k_0, k_0+1, k_0+2, \ldots$ and satisfies

$$P(X \geq k) = [1 + (\frac{k-k_0}{\sigma})^{1/\gamma}]^{-\alpha}, k = k_0, k_0+1, ... \qquad (7.45)$$

Typically k_0 will be an integer, but it does not have to be. Analogously we may define Zipf(I), (II) and (III) distributions by suitable parametric constraints in (7.45). Minima of Zipf (IV) samples again have Zipf (IV) distributions and as a direct parallel to the Pareto (III) result, the Zipf (III) family is closed under geometric minimization. The standard Zipf distribution may be taken to be the Zipf (IV) $(0,1,1,1)$ distribution. If U has such a distribution then

$$P(U \leq k) = \frac{k}{k+1}, k = 0,1,2,\ldots. \qquad (7.46)$$

This distribution (translated by 1) arises as the waiting time till the first record value in a sequence of i.i.d. continuous random variables. It also admits a representation as a uniform mixture of geometric distributions.

6 Inequality Measures

Pareto (1897) suggested that the (negative of the) slope of a Pareto chart (the plot of $\log \overline{F}_X(x)$ vs. $\log x$) could be used as a suitable indication of inequality in a population. In fact, for a classical Pareto distribution, the slope will be $-\alpha$ and small values of α will be associated with a large amount of inequality. Most of the popular inequality measures are, in the classical Pareto case, decreasing functions of α. For the more general Pareto models, (II) - (IV), the Pareto chart will be approximately linear with a slope given by $-\alpha/\gamma$ and for these distributions also, most of the commonly utilized inequality measures again will be decreasing functions of the negative of this slope. A pot-pourri of inequality measures have been proposed since Pareto proposed his graphical measure.

Arguments can be advanced in favor of inequality measures that are scale invariant. From this viewpoint the Gini index introduced in Gini (1914) becomes an attractive candidate. For a distribution F supported in $(0,\infty)$ we can define the Gini index to be

$$G(F) = E(|X_1 - X_2|)/2E(X_1) \tag{7.47}$$

where X_1, X_2 are i.i.d. with common distribution F. Here we assume that $E|X_1| < \infty$.

A somewhat earlier competitor of the Gini index is the coefficient of variation which can be defined by

$$CV(F) = \sqrt{E[(X_1 - X_2)^2]}/\sqrt{2}E(X_1) \tag{7.48}$$

where again X_1, X_2 are i.i.d. with common distribution F. For this we must assume that $E(X_1^2) < \infty$.

However, it has transpired that another graphical indicator of inequality has taken over center stage, especially in income distribution contexts. This is the celebrated Lorenz curve introduced by Lorenz (1905). The Lorenz curve $L_F(u)$ associated with a distribution function F supported in $(0,\infty)$ (with quantile function F^{-1}) is conveniently defined by (following Gastwirth (1971))

$$L(u) = \int_0^u F^{-1}(y)dy \Big/ \int_0^1 F^{-1}(y)dy, \qquad 0 \le u \le 1 \tag{7.49}$$

where it is assumed that the mean of F (equal to the denominator in (7.49)) exists. Lorenz proposed to measure inequality by the degree to which the bow-shaped Lorenz curve is "bent". Thus if F_1 and F_2 have corresponding Lorenz curves L_1 and L_2, we will say that F_1 exhibits at least as much inequality as does F_2 if $L_1(u) \le L_2(u), \forall u \in [0,1]$ and we will write $F_1 \ge_L F_2$. Arguments based on Dalton's (1920) transfer principles suggest that inequality measures should respect this Lorenz ordering. Many, though not all, inequality measures do so, and indeed several have attractive graphical interpretations related to Lorenz curves. For example, the Gini index (7.47) corresponding to a distribution F is equal to two times the area between the Lorenz curve $L(F)$ and the egalitarian line (which corresponds to the Lorenz curve for a degenerate random variable $X = c > 0$, representing the situation where everyone has the same income). Thus

$$G(F) = 2\int_0^1 [u - L_F(u)]du. \tag{7.50}$$

The Pietra index which can be defined as

$$E(|X-E(X)|)/2E(X) \tag{7.51}$$

also has a Lorenz curve interpretation (Pietra, 1932). It can be viewed either as the maximal deviation between $L_F(u)$ and u or, equivalently, as two times the area of the maximal triangle that can be inscribed between $L(u)$ and the egalitarian line. For the hierarchy of Pareto distribution (I) - (IV) we may summarize these inequality measures as follows

Pareto (I) (σ,α) $\bar{F}(x)=(x/\sigma)^{-\alpha}, x\geq\sigma$. Assume $\alpha>1$ to ensure that $E|X|<\infty$.

$$\begin{aligned}
\text{Lorenz curve}: &\quad L(u)=1-(1-u)^{(\alpha-1)/\alpha}\\
\text{Gini index}: &\quad (2\alpha-1)^{-1}\\
\text{Pietra index}: &\quad (\alpha-1)^{\alpha-1}/\alpha^{\alpha}\\
\text{Coefficient of variation}: &\quad (\alpha^2-2\alpha)^{-1/2} \text{ if } \alpha>2
\end{aligned}$$

Pareto (II) $(0,\sigma,\alpha)$ ($\mu=0$ for simplicity)

$$\begin{aligned}
\text{Gini index}: &\quad \alpha/(2\alpha-1)\\
\text{Pietra index}: &\quad [\frac{\alpha-1}{\alpha}]^{\alpha-1}\\
\text{Coefficient of variation}: &\quad \alpha^{1/2}(\alpha-2)^{-1/2} \text{ if } \alpha>2
\end{aligned}$$

Pareto (III) $(0,\sigma,\gamma)$

$$\text{Gini index}: \quad \gamma$$

Pareto (IV) $(\mu,\sigma,\gamma,\alpha)$

$$\begin{aligned}
\text{Lorenz curve}: &\quad L(u)=\frac{\mu u+\alpha\{B(\alpha-\gamma,\gamma+1)-I_{(1-u)^{1/\alpha}}(\alpha-\gamma,\gamma+1)\}}{\mu+\sigma\alpha B(\alpha-\gamma,\gamma+1)}\\
\text{Gini index}: &\quad 1-\frac{\mu+2\sigma\alpha B(2\alpha-\gamma,\gamma+1)}{\mu+\sigma\alpha B(\alpha-\gamma,\gamma+1)}
\end{aligned}$$

where $I_z(\alpha,\beta)=\int_0^z v^{\alpha-1}(1-v)^{\beta-1}dv$ denotes the incomplete beta function.

7 Inference

Parameter estimation based on samples from the classical Pareto (I) distribution has been the subject of a large number of papers. By a simple transformation this problem reduces to the problem of estimation for data from a translated exponential distribution, though few authors have explicitly taken advantage of this possible approach. If we have a sample $X_1, X_2, \ldots, X_n$ from a Pareto (I) (σ, α) population, the corresponding likelihood function is of the form

$$L(\sigma,\alpha) = \alpha^n \sigma^{n\alpha} (\prod_{i=1}^{n} X_i)^{-(\alpha+1)} I(\sigma \le X_{1:n}). \tag{7.52}$$

A complete minimal sufficient statistic based on $X_1, \ldots, X_n$ is $(X_{1:n}, \prod_{i=1}^{n} X_i)$. It is not difficult to determine the maximum likelihood estimate of (σ, α). Observe that for any fixed α, the likelihood is monotone increasing in σ and so is maximized when $\sigma = X_{1:n}$. Thus we have

$$P(I) \qquad \hat{\sigma}_n = X_{1:n} \tag{7.53}$$

$$\hat{\alpha}_n = [\frac{1}{n} \sum_{i=1}^{n} \log(X_i / X_{1:n})]^{-1} \tag{7.54}$$

$(\hat{\alpha}_n, \hat{\sigma}_n), n = 1, 2, \ldots$ is a strongly consistent sequence of estimates of (α, σ). In addition, $\hat{\alpha}_n$ is asymptotically normal. For a fixed sample size n we can explicitly describe the distribution of$(\hat{\alpha}_n, \hat{\sigma}_n)$. Using Basu's lemma or results on ratios of adjacent order statistics for Pareto (I) samples, we can verify that $\hat{\alpha}_n$ and $\hat{\sigma}_n$ are independent random variables. The corresponding marginal densities are available when we observe that

$$\hat{\sigma}_n \sim P(I)(\sigma, n\alpha) \tag{7.55}$$

and

$$(\hat{\alpha}_n)^{-1} \sim \Gamma(n-1, (\alpha n)^{-1}). \tag{7.56}$$

From (7.55) and (7.56) we obtain

$$E(\hat{\sigma}_n) = \sigma(1 - \frac{1}{n\alpha})^{-1} \tag{7.57}$$

$$\text{var}(\hat{\sigma}_n) = \sigma^2 n\alpha (n\alpha - 1)^{-2} (n\alpha - 2)^{-1} \tag{7.58}$$

$$E(\hat{\alpha}_n) = \alpha n / (n-2) \tag{7.59}$$

$$\text{var}(\hat{\alpha}_n) = \alpha_n^2 (n-2)^{-2} (n-3)^{-1} \tag{7.60}$$

Both $\hat{\alpha}_n$ and $\hat{\sigma}_n$ are thus positively biased.

To obtain minimum variance unbiased estimates of σ and α, we need only to identify the (unique) functions of the minimal sufficient statistics which are unbiased for σ and α. In this way we are led to the MVUE's:

$$\hat{\sigma}_n^{(U)} = [1 - (n-1)^{-1}\,\hat{\alpha}_n^{-1}]\,\hat{\sigma}_n \tag{7.61}$$

and

$$\hat{\alpha}_n^{(U)} = \frac{n-2}{n}\,\hat{\alpha}_n\,. \tag{7.62}$$

An alternative estimator of α, which has smaller mean squared error than $\hat{\alpha}_n$ and $\hat{\alpha}_n^{(U)}$ is given by

$$\hat{\alpha}_n^{(J)} = \frac{n-3}{n}\,\hat{\alpha}_n \tag{7.63}$$

where the "J" refers to A.M. Johnson who first derived this estimate (see Saksena (1978)).

Minor modifications of $\hat{\sigma}_n^{(U)}$ have been suggested with a goal of reducing the mean squared error, but no uniformly better estimate has been identified. Note that which of the estimates $\hat{\sigma}_n^{(U)}$ and $\hat{\sigma}_n$ has smallest mean squared error depends on the true value of α.

Since we have a complete sufficient statistic available it is not difficult to identify minimum variance unbiased estimates of many functions of (σ, α). For example one can find the MVUE of $P(X > t) = (t/\sigma)^{-\alpha}$ (Lwin (1972)) or the Gini index $(2\alpha - 1)^{-1}$. Of course, maximum likelihood estimates of functions of the form $h(\sigma, \alpha)$ will be obtained by "plugging in" $\hat{\sigma}_n$ and $\hat{\alpha}_n$, i.e. $\widehat{h(\sigma,\alpha)} = h(\hat{\sigma}_n, \hat{\alpha}_n)$. Likewise one can determine the best unbiased predictor of a future observation X.

Several alternative estimation methods have been proposed. One simple approach is based on the fact that the Pareto chart (a plot of $\log \bar{F}_X(x)$ against $\log x$) for a Pareto distribution is linear, with a slope $-\alpha$ and intercept $\log \sigma$. So we may simply find a least squares line of best fit to the empirical Pareto chart of the sample and use the fitted slope and intercept to estimate σ and α in the obvious way. Such estimates are consistent (Quandt (1966)). Similarly it is possible to obtain a consistent estimate of α by a least squares fitting of the empirical Lorenz curve.

Variants of the method of moments can instead be used for parameter estimation. Perhaps the most popular of these techniques is one suggested by Quandt (1966). He equated the sample mean and the sample minimum to their expected values and solved for σ and α. Using an (M) to remind us that a method of moments was used, we can express these estimates as

$$\hat{\alpha}_n^{(M)} = \frac{n\bar{X} - X_{1:n}}{n(\bar{X} - X_{1:n})} \tag{7.64}$$

and

$$\hat{\sigma}_n^{(M)} = \frac{n\widehat{\alpha}_n^{(M)} - 1}{n\widehat{\alpha}_n^{(M)}} X_{1:n} \tag{7.65}$$

(here we must assume $\alpha > 1$ for the existence of $E(\bar{X})$). The estimates in (7.64) and (7.65) are consistent. If we are unhappy with the assumption $\alpha > 1$, we may use a fractional moment of $\bar{X}$ instead of its mean, to derive alternative estimates. This will extend the range of values for which α can be consistently estimated.

Quantile estimation is a viable alternative. For this Quandt (1966) proposed selection of two probabilities $p_1 > p_2$. We may then obtain estimates by equating sample quantiles to population quantiles. Thus we solve the pair of equations

$$p_i = 1 - \left(\frac{X_{[np_i]:n}}{\sigma}\right)^{\alpha}, \qquad i = 1,2 \tag{7.66}$$

to obtain our estimates.

By transforming our Pareto (I) data to exponential data, several authors have discussed best linear unbiased estimates based on chosen order statistics (see Johnson *et al.* (1994, pp. 584-587) for an introduction to this area of investigation).

In the case in which σ is known, Brazauskas and Serfling (2000) make a strong case for the use of a generalized median estimate of α constructed as follows. For each subset of k of the n random variables $X_1, X_2, ..., X_n$ define

$$h(X_{i_1}, X_{i_2}, ..., X_{i_k}) = \left(\frac{1}{k}\sum_{j=1}^{k} \log X_{i_j} - \log\sigma\right)^{-1}.$$

The generalized median estimate of α is then defined to be the median of these quantities that can be viewed as subsample based estimates. If σ is not known, it is reasonable to substitute $X_{1:n}$ for it in the generalized median. A strong argument in favor of such estimates is their robustness against outliers. Detailed discussion of generalized median estimates, together with comparisons with competing estimates in terms of asymptotic relative efficiency and breakdown points may be found in Brazauskas and Serfling (2000).

Instead of using the frequentist estimation approaches discussed above, we may consider a Bayesian formulation of the problem. If σ is known, then the problem, after a logarithmic transformation reduces to one involving estimation of the reciprocal of the scale parameter from a sample from a gamma distribution. A gamma prior will be conjugate here and a routine analysis is possible. If α is known, then a power function prior density for σ ($f(\sigma) \propto \delta\sigma^{\delta-1}, 0 < \sigma < \sigma_0$) is conjugate and again a routine analysis is possible. More interesting is the case in which both σ and α are unknown. A "natural" conjugate prior for this case was identified by Lwin (1972). Arnold and Press (1989) suggested that independent priors be used for σ and α. Instead we will approach the problem from the viewpoint of conditionally conjugate priors in the sense of Arnold *et al.* (1998). This approach subsumes the Lwin prior and the independent marginals prior but provides more flexibility.

Before we begin it is convenient to reparameterize our Pareto density in terms of a shape parameter α and a precision parameter $\tau(= 1/\sigma)$. With this parameterization the likelihood function becomes

$$f_{\underline{X}}(\underline{x}; \alpha, \tau) = \alpha^n \tau^{-n\alpha} \left(\prod_{i=1}^{n} x_i\right)^{-(\alpha+1)} I(\tau x_{1:n} > 1). \tag{7.67}$$

If τ were known, it would be natural to take a prior for α to be in the gamma family. If α were known, the natural conjugate prior for τ would be in the Pareto family.

It is then reasonable to use a prior density for (α, τ) which has gamma and Pareto conditionals. The corresponding 6 parameter family of priors is then given by

$$\begin{aligned} f(\alpha, \tau) \quad &\propto \exp[m_{01} \log \tau + m_{21} \log \alpha \log \tau] \\ &\times \exp[m_{10}\alpha + m_{20} \log \alpha + m_{11} \alpha \log \tau] I(\tau c > 1) \end{aligned} \tag{7.68}$$

It is readily verified that (7.68) is a conjugate family and that it does have gamma and Pareto conditionals. When (7.68) is combined with (7.67) the resulting posterior density is again of the form (7.68) with posterior hyper-parameters (indicated by primes) related to prior hyper-parameters (without primes) as below

$$m'_{01} = m_{01}$$
$$m'_{21} = m_{21}$$

$$m'_{10} = m_{10} - \sum_{i=1}^{n} \log x_i \tag{7.69}$$

$$\begin{aligned} m'_{20} &= m_{20} + n \\ m'_{11} &= m_{11} - n \\ c' &= \min(x_{1:n}, c). \end{aligned}$$

Because the posterior distribution has gamma and Pareto conditionals, it is easy to simulate realizations from the posterior density using a Gibbs sampler. The family of priors (7.68) includes the Lwin (1972) priors. For them, set $m_{01} = m_{21} = 0$. It also includes the "independent marginals" priors of Arnold and Press (1989). They correspond to the choice $m_{11} = m_{21} = 0$ in (7.68).

More detailed discussion of such conditionally conjugate priors may be found in Arnold *et al.* (1999, ch. 13).

Censored data pose no additional difficulties for maximum likelihood or Bayesian inference in the Pareto (I) family. If our data consists of n precisely observed X_i's say $X_1 = x_1, \ldots, X_n = x_n$ and m imprecisely observed observations $X_1' > c_1, X_2' > c_2, \ldots, X_m' > c_m$ where $c_1, c_2, \ldots, c_m$ are known, then our likelihood function (reverting once more to the (σ, α) parameterization) will be, provided $c_{1:n} > x_{1:n}$,

$$\begin{aligned} L(\sigma, \alpha) &= \prod_{i=1}^{n} \frac{\alpha}{\sigma} \left(\frac{X_i}{\sigma}\right)^{-(\alpha+1)} \prod_{j=1}^{m} \left(\frac{c_i}{\sigma}\right)^{-\alpha} I(X_{1:n} \geq \sigma) \\ &\propto \alpha^n \sigma^{(n+m)\alpha} \left(\prod_{i=1}^{n} X_i \prod_{j=1}^{m} c_j\right)^{-\alpha} I(X_{1:n} \geq \sigma) \end{aligned} \tag{7.70}$$

which is of the same form as (7.52) so that maximum likelihood estimates are readily obtainable and the same classes of prior densities can be used for Bayesian analysis as were used when all of the X_i's were precisely observed. If the data are grouped, life becomes more complicated. See Arnold and Press (1986) for some discussion of Bayesian inference in such a situation.

Turning now to estimation problems for the more general Pareto (II) - (IV) distributions, we predictably find estimation to be a little more difficult. Suppose that we have n observations $X_1, X_2, \ldots, X_n$ from a Pareto (IV) $(\mu, \sigma, \gamma, \alpha)$ distribution. The log-likelihood is of the form

$$P(IV) \qquad l(\mu,\sigma,\gamma,\alpha) = (\frac{1}{\gamma}-1)\sum_{i=1}^{n}\log(\frac{X_i-\mu}{\sigma}) - (\alpha+1)\sum_{i=1}^{n}\log(1+(\frac{X_i-\mu}{\sigma})^{1/\gamma}) - n\log\gamma - n\log\sigma + n\log\alpha. \tag{7.71}$$

It is reasonable to use $X_{1:n}$ as an estimate of μ. If we do this and then subtract $X_{1:n}$ from the other observations we may act as if we have a sample of size $(n-1)$ from a $P(IV)(0, \sigma, \gamma, \alpha)$ distribution. The likelihood equations, even in this simplified setting, must be solved numerically but we will be guaranteed asymptotic normality of the resulting m.l.e.'s and so we will have

$$\begin{pmatrix} \hat{\sigma}_n \\ \hat{\gamma}_n \\ \hat{\alpha}_n \end{pmatrix} \sim N\left(\begin{pmatrix} \sigma \\ \gamma \\ \alpha \end{pmatrix}, \frac{1}{n}J\right) \tag{7.72}$$

where J is the information matrix. Brazauskas (2003) discusses this information matrix in some detail. In an earlier paper Brazauskas (2002) provided the information matrix for the Feller-Pareto distribution.

It is of course possible to use the method of moments for Pareto (IV) samples. Again it is simplest to use $X_{1:n}$ as an estimate of μ and then we can set up 3 moment equations for a Pareto $(0, \sigma, \gamma, \alpha)$ sample using expressions for the moments in (7.11). Note that an iterative search for a solution may be necessary since the expressions for the moments involve gamma functions. More attractive is the possibility of equating 4 sample quantiles to the corresponding population quantiles, although even here, the solution of the 4 equations in 4 unknowns (or 3 equations if we set $\mu = X_{1:n}$) will be a non-trivial numerical exercise.

Bayesian inference for $P(II) - P(IV)$ populations will be hindered by the absence of a convenient minimal sufficient statistic. It is of course possible to use diffuse, reference or non-informative priors for the parameters but analysis of the resulting posterior density will be computer intensive.

8 Multivariate Pareto Distributions

The first k-dimensional Pareto distribution to appear in the literature was introduced by Mardia (1962). It has $P(I)$ marginals. Its joint survival function is of the form

$$\bar{F}_{\underline{X}}(\underline{x}) = P(\underline{X} > \underline{x}) = [\sum_{i=1}^{k} \frac{x_i}{\sigma_i} - k + 1]^{-\alpha}, \quad x_i > \sigma_i, \quad i = 1, 2, \ldots, k \tag{7.73}$$

where $\alpha > 0$ and $\sigma_i > 0$, $i = 1, 2, \ldots, k$. If $\underline{X}$ has such a survival function we will write

$$\underline{X} \sim P^{(k)}(I)(\underline{\sigma}, \alpha).$$

Let us partition $\underline{X} = (\underline{X}^{(1)}, \underline{X}^{(2)})$ where $\underline{X}^{(1)}$ is of dimension $k_1 < k$ and $\underline{X}^{(2)}$ is of dimension $k - k_1$. Analogously we partition $\underline{\sigma} = (\underline{\sigma}^{(1)}, \underline{\sigma}^{(2)})$. Using this notation we may verify that the marginals of the $P^{(k)}(I)$ distribution are again multivariate Pareto and conditional distributions have translated multivariate Pareto distributions. Specifically we have

$$\underline{X}^{(1)} \sim P^{(k_1)}(I)(\underline{\sigma}^{(1)}, \alpha) \tag{7.74}$$

and

$$\underline{X}^{(1)} | \underline{X}^{(2)} = \underline{x}^{(2)} \sim c(\underline{x}^{(2)}) P^{(k_1)}(\underline{\sigma}^{(1)}, \alpha + k - k_1) + (1 - c(\underline{x}^{(2)}))\underline{\sigma}^{(1)} \tag{7.75}$$

where $c(\underline{x}^{(2)}) = [\sum_{1=k_1+1}^{k} (\frac{x_i}{\sigma_i}) - k + k_1 + 1]$.

It is possible to define multivariate Pareto (II) and Pareto (III) distributions but of course they may be viewed as special cases of the following multivariate Pareto (IV) distribution

$$\bar{F}_{\underline{X}}(\underline{x}) = [1 + \sum_{i=1}^{k} (\frac{x_i - \mu_i}{\sigma_i})^{1/\gamma_i}]^{-\alpha}, \quad x_i > \mu_i, \quad i = 1, 2, \ldots, k \tag{7.76}$$

where $\underline{\mu} \in \mathbf{R}^k$, $\underline{\sigma} > \underline{0}$, $\underline{\gamma} > \underline{0}$ and $\alpha > 0$. If $\underline{X}$ has a survival function corresponding to (7.76) we will write

$$\underline{X} \sim P^{(k)}(IV)(\underline{\mu}, \underline{\sigma}, \underline{\gamma}, \alpha).$$

This family is "closed" under marginalization and conditioning. We have

$$\underline{X}^{(1)} \sim P^{(k_1)}(IV)(\underline{\mu}^{(1)}, \underline{\sigma}^{(1)}, \underline{\gamma}^{(1)}, \alpha) \tag{7.77}$$

and

$$\underline{X}^{(1)} | \underline{X}^{(2)} = \underline{x}^{(2)} \sim P^{(k_1)}(IV)(\underline{\mu}^{(1)}, \underline{\tau}^{(1)}, \underline{\gamma}^{(1)}, \alpha + k - k_1) \tag{7.78}$$

where

$$\tau_i = \sigma_i[1 + \sum_{j=k_1+1}^{k} (\frac{x_i - \mu_i}{\sigma_i})^{1/\gamma_j}]^{\gamma_i}, \;\; i = 1, 2, \ldots, k.$$

The following stochastic representation of a $P^{(k)}(IV)$ random vector $\underline{X}$ is convenient for determining the regression functions and the variance covariance structure of the distribution

$$X_i = \mu_i + \sigma_i(W_i/Z)^{\gamma_i}, \;\; i = 1, 2, \ldots, k \tag{7.79}$$

where the W_i's are i.i.d. standard exponential random variables (i.e. $W_i \sim \Gamma(1,1)$) and Z, independent of the W_i's, has a $\Gamma(\alpha,1)$ distribution.
Remark: A multivariate Feller-Pareto distribution can be defined by using a representation analogous to (7.79), in which the W_i's are independent gamma variables with $W_i \sim \Gamma(\delta_i,1), i = 1, 2, \ldots, k$ independent of $Z \sim \Gamma(\alpha,1)$.

There are several other ways in which multivariate Pareto (IV) random vectors can be constructed. One can take gamma mixtures of dependent Weibull random variables. One can begin with $m > k$ independent Pareto(IV) variables, $Y_1, \ldots, Y_m$, and define $X_1, \ldots, X_k$ by postulating that each of the X_i's is a minimum of some subset of the Y_i's. Arnold (1983, pp. 260-263) discusses several multivariate Pareto (III) distributions which are defined in terms of geometric minimization. Finally we remark that Arnold *et al.* (1999, pp. 182-183) describe several multivariate distributions with Pareto conditional distributions. A requirement that the conditional distributions should be Paretian, rather than positing that marginal densities should be Paretian, might actually be more compelling as a modeling assumption. As an example of a bivariate density with Pareto (II) conditionals (with $\mu = 0$) consider

$$f_{X_1,X_2}(x_1,x_2,\underline{\lambda}) \propto (\lambda_{00} + \lambda_{10}x_1 + \lambda_{01}x_2 + \lambda_{11}x_1x_2)^{-(\alpha+1)}. \tag{7.80}$$

This model includes Mardia's bivariate Pareto model and also includes distributions with independent Pareto (II) marginals. Higher dimensional versions of conditionally specified distributions such as (7.80) can potentially involve enormous numbers of parameters so that simplified sub-models need to be considered to hopefully avoid the curse of over-flexibility of these models. They are mathematically identifiable, but based on a finite sample, many parameter configurations will seem to be equally suitable for approximately matching the empirical data configurations.

9 Envoi

Multivariate models with surfeits of parameters and univariate models such as the Feller-Pareto model with 5 parameters for univariate data, will cause enormous headaches when we try to estimate parameters, construct confidence intervals, test hypotheses or try to decide among hierarchical (and non-hierarchical) families of submodels. It may be cause for wishing for the good old days when Pareto's simple two parameter model held the field without competition. Life was easier when all you had to do was roughly sketch an empirical Pareto chart and read off the slope and intercept. Then one could quickly get down to the enjoyable activity of arguing with colleagues about the economic and social implications of the degree of inequality exhibited by various populations and subpopulations. But even Pareto recognized that his simple model wasn't always adequate. Bells and whistles had to be added. For example, Pareto proposed to sometimes use the following model

$$\bar{F}_X(x) = C(x+b)^{-\alpha}e^{-\beta x}, \qquad x > d \tag{7.81}$$

The hierarchy of generalized Pareto models introduced in section 2 may very well not be the final answer. There is still room for suggestions of other suitable models with Paretian tail behavior.

References

Arnold, B. C. (1983) *Pareto Distributions*, International Cooperative Publishing House, Fairland, Maryland USA.

Arnold, B. C., N. Balakrishnan and H. N. Nagaraja (1992) *A First Course in Order Statistics*, Wiley, New York.

Arnold, B. C., E. Castillo and J. M. Sarabia (1998) Bayesian Analysis for Classical Distributions Using Conditionally Specified Priors, *Sankya B*, **60**, 228–245.

Arnold, B. C., E. Castillo and J. M. Sarabia (1999) *Conditional Specification of Statistical Models*, Springer, New York.

Arnold, B. C. and L. Laguna (1976) A Stochastic Mechanism Leading to Asymptotically Paretian Distributions, Proceedings of the American Statistical Association. Business and Economic Statistics Section.

Arnold, B. C. and L. Laguna (1977) *On Generalized Pareto Distributions with Applications to Income Data*, International Studies in Economics, Monograph No. 10, Department of Economics, Iowa State University.

Arnold, B. C. and S. J. Press (1986) Bayesian Analysis of Censored or Grouped Data from Pareto Populations, in *Studies in Bayesian Econometrics and Statistics, 6*, pp. 157–173, Amsterdam, North Holland.

Arnold, B. C. and S. J. Press (1989) Bayesian Estimation and Prediction for Pareto Data, *Journal of American Statistical Association*, **84**, 1079–1084.

Balkema, A. A. and L. de Haan (1974) Residual Life Time at Great Age, *Annals of Probability*, **2**, 792–804.

Brazauskas, V. (2002) Fisher Information Matrix for the Feller-Pareto Distribution, *Statistics and Probability Letters*, **59**, 159–167.

Brazauskas, V. (2003) Information Matrix for Pareto (IV), Burr and Related Distributions, *Communications in Statistics: Theory and Methods*, **32**, 315–325.

Brazauskas, V. and R. Serfling (2000) Robust and Efficient Estimation of the Tail Index of a Single-Parameter Pareto Distribution, *North American Actuarial Journal*, **4**, 12–27.

Burr, I. W. (1942) Cumulative Frequency Functions, *Annals of Mathematical Statistics*, **13**, 215–232.

Champernowne, D. G. (1937) The Theory of Income Distribution, *Econometrica*, **5**, 379–381.

Champernowne, D. G. (1953) A Model of Income Distribution, *Economic Journal*, **63**, 318–351.

Champernowne, D. G. (1973) *The Distribution of Income*, Cambridge University Press, Cambridge.

Dalton, H. (1920) The Measurement of the Inequality of Income, *Economic Journal*, **30**, 348–361.

Ericson, O. (1945) Un Probleme de la Repartition des Revenus, *Skand. Aktuar*, **28**, 247–257.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, vol. 2, 2nd ed., Wiley, New York.

Gastwirth, J. L. (1971) A General Definition of the Lorenz Curve, *Econometrica*, **39**, 1037–1039.

Gibrat, R. (1931) *Les Inégalités Économiques*, Librairie du Recueil Sirey, Paris.

Gini, C. (1914) Sulla Misura della Concentrazione e della Variabilità dei Caratteri, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, **73**, 1203–1248.

Houchens, R. L. (1984) *Record Value Theory and Inference*, Ph.D. Dissertation, University of California, Riverside, California.

Huang, J. S. (1978) On a “Lack of Memory” Property, Tech. Rep. 84, Statistical Series, University of Guelph, Guelph, Ontario, Canada.

Johnson, N. L., S. Kotz and N. Balakrishnan (1994) *Continuous Univariate Distributions*, vol. 1, 2nd ed., Wiley, New York.

Kalbfleisch, J. D. and R. L. Prentice (2002) *The Statistical Analysis of Failure Time Data*, 2nd ed., Wiley, New York.

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

Krishnaji, N. (1970) Characterization of the Pareto Distribution through a Model of Under-Reported Incomes, *Econometrica*, **38**, 251–255.

Lorenz, M. O. (1905) Methods of Measuring the Concentration of Wealth, *Quarterly Publications of the American Statistical Association*, **9 (New Series, No. 70)**, 209–219.

Lwin, T. (1972) Estimation of the Tail of the Paretian Law, *Skand. Aktuar*, **55**, 170–178.

Lydall, H. F. (1959) The Distribution of Employment Income, *Econometrica*, **27**, 110–115.

Mardia, K. V. (1962) Multivariate Pareto Distributions, *Annals of Mathematical Statistics*, **33**, 1008–1015.

Pareto, V. (1897) *Cours d’Economie Politique*, Rouge, Lausanne.

Pickands, J. (1975) Statistical Inference Using Extreme Order Statistics, *Annals of Statistics*, **3**, 119–131.

Pietra, G. (1932) Nuovi Contributi Alla Methodologia Degli Indici di Variabilita e di Concentrazione, *Atti del Reale Instituto Veneto di Scienze, Lettere ed Arti*, **91**, 989–1008.

Pillai, R. N. (1991) Semi-Pareto Processes, *Journal of Applied Probability*, **28**, 461–465.

Quandt, R. E. (1966) Old and New Methods of Estimation and the Pareto Distribution, *Metrika*, **10**, 55–82.

Saksena, S. K. (1978) *Estimation of Parameters in a Pareto Distribution and Simultaneous Comparison of Estimation*, Ph.D. thesis, Louisiana Tech. University, Ruston (Louisiana).

Sukhatme, P. V. (1937) Tests of Significance for Samples of the Chi Squared Population with Two Degrees of Freedom, *Annals of Eugenics*, **8**, 52–56.

Zipf, G. (1949) *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge.

CHAPTER 8

The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality*

James B. McDonald† and Michael Ransom‡

Abstract

The generalized beta (GB) is considered as a model for the distribution of income. It is well known that its special cases include Dagum's distribution along with the Singh-Maddala distribution. Related measures of inequality such as the Gini Coefficient, Pietra Index, or Theil Index are expressed in terms of the parameters of the generalized beta. This paper also explores the use of numerical integration techniques for calculating inequality indexes. Numerical integration may be useful since in some cases it may be computationally very difficult to evaluate the equations that have been derived or the equations are not available. We provide examples from the distribution of family income in the United States for the year 2000.

* The research assistance of Sam Dastrup and Brad Larsen is greatly appreciated. Comments by the referees are gratefully acknowledged.

† Department of Economics, Brigham Young University, Provo, Utah, USA,
email: james_mcdonald@byu.edu

‡ Department of Economics, Brigham Young University, Provo, Utah, USA

1 Introduction

Parametric and nonparametric methods have been considered in describing the size distribution of income. This paper provides a survey of the generalized beta (GB) as an income distribution, derives some inequality measures for some previously unknown cases, and explores the use of numerical methods to evaluate inequality measures to a fitted distribution. These issues are important because poor fitting functional forms or inaccurate measures of inequality can lead to inappropriate economic policy.

Pareto first proposed a model of income distribution in 1895 which was found to accurately model the upper tail of the distribution, but did a poor job describing the lower tail (Pareto, 1895). Pareto's analysis generated a debate on the effect of economic growth on income inequality. Gini disagreed with Pareto's opinion that economic growth leads to less inequality. Gini proposed a unit-free measure of income inequality known as the Gini coefficient that is still commonly used today (Gini, 1912).

Gibrat's (1931) law of proportionate effect provided a theoretical basis for the two-parameter lognormal distribution to be considered as a model for the size distribution of income. The lognormal was further examined by Aitchison and Brown (1969). Another two-parameter distribution, the gamma, was proposed by Ammon (1895)[17] and was more recently reintroduced and fit to US income data by Salem and Mount (1974). Bartels and van Metelen (1975) suggested the two-parameter Weibull distribution. While these two-parameter models provide increased flexibility in fitting empirical data, they do not allow for intersecting Lorenz curves sometimes observed with income data.

The introduction of a third parameter allows for intersecting Lorenz curves. Some three-parameter models which have been used to model the size distribution of income include the generalized gamma (Amoroso, 1924-1925; Taillie, 1981) and beta (Thurow, 1970) as well as two closely related models which are members of the Burr family of distributions: the Singh and Maddala (1976), known in statistics literature as the Burr 12, and the Dagum (1977), known as the Burr 3.

The generalized beta of the first and second kind (GB1 and GB2) are four-parameter distributions which have not only been very successful in fitting the data, but also include all of the previously mentioned distributions as special or limiting cases, McDonald (1984). The empirical success of the GB2 was complemented by Parker's (1999) theoretical model of income generation, showing earnings to follow a GB2 distribution. Bordley *et al.* (1996) found that the GB2 distribution generally provided a significantly better fit than its nested distributions when fit to income

[17] Dagum (1980) states that, "Ammon appears to be the first to propose the gamma pdf as a descriptive model of income distribution. It was applied by L. March (1898) to fit the wage distribution of several professional categories. . . ." One of the referees to this paper indicated that he had been unable to verify that Ammon proposed the gamma in any of the three editions of his book. Having access to the third edition, the referee observed a histogram with a superimposed smooth curve, which Ammon states is an eyeball estimate. This curve resembles a gamma pdf, but this does not prove that Ammon actually used a gamma distribution.

data from the United States. Bandourian *et al.* (2003) applied the generalized beta distributions to income data from 23 countries and various years from the mid 70's to the mid 90's. They found that the Weibull, Dagum, and generalized beta of the second kind were generally the best fitting models with two, three, and four parameters when using earnings income data. Furthermore, estimated measures of income inequality increased over time (1979-2000) for most countries. Dastrup *et al.* (2007) investigated the impact of taxes and transfer payments on the distribution of income over time for a number of countries.

Section 2 summarizes the statistical models, related estimation issues, and measures of inequality to be considered in this paper. Nonparametric and parametric estimators of the Gini, Pietra, and Theil estimators are reviewed in Section 3, with applications being considered in Section 4. Section 5 includes a summary of the conclusions. The Appendices (I and II) contain the derivation of expressions for the Pietra and Theil Indices as a function of the distributional parameters.

2 Statistical Models for the Size Distribution of Income

2.1 *The generalized beta distribution family*

The generalized beta (GB) distribution is defined by its probability density function (pdf),

$$GB(y;a,b,c,p,q) = \frac{|a|\, y^{ap-1}(1-(1-c)(y/b)^a)^{q-1}}{b^{ap}\mathrm{B}(p,q)(1+c(y/b)^a)^{p+q}} \text{ for } 0 < y^a < \frac{b^a}{1-c} \tag{8.1}$$

and zero otherwise, where $0 \le c \le 1$; $b,p,q > 0$; and $B(p,q)$ denotes the beta function.

The GB pdf has an inverted "*U*" shape if *ap <1* and *q <1* with vertical asymptotes at $y = 0$ and $y = b/(1-c)^{1/a}$. The GB includes all of the distributions mentioned in Section 1 as special or limiting cases, McDonald and Xu (1995). The four- parameter GB1 and GB2 correspond to the GB with the *c* parameter set equal to zero and one, respectively:

$$\mathrm{GB1}(y;a,b,p,q) = \frac{|a|\, y^{ap-1}\,(1-(y/b)^a)^{q-1}}{b^{ap}B\,(p,q)} = GB(y;a,b,c=0,p,q) \tag{8.2}$$

$$\mathrm{GB2}(y;a,b,p,q) = \frac{|a|\, y^{ap-1}}{b^{ap}\mathrm{B}\,(p,q)\,(1+(y/b)^a)^{p+q}} = \mathrm{GB}(y;a,b,c=1,p,q). \tag{8.3}$$

Thurow (1970) used the beta of the first kind (B1)

$$\mathrm{B1}\,(y;b,p,q) = \frac{y^{p-1}(b-y)^{q-1}}{b^p B\,(p,q)} = \mathrm{GB1}\,(y;a=1,b,p,q) \;\; ; \;\; 0<y<b \tag{8.4}$$

to analyze factors contributing to income inequality for whites and blacks.

One of the first distributions used to model income was the Pareto distribution

$$Pareto\,(y;b,p) = \frac{py^{-p-1}}{b^{-p}} = GB1\,(y;a=-1,b,p,q=1) \quad \text{for } \; b<y. \tag{8.5}$$

The three-parameter Dagum distribution corresponds to the case

$$DAGUM(y;a,b,p) = \frac{|a|\,py^{ap-1}}{b^{ap}\left(1+(y/b)^a\right)^{p+1}} = \mathrm{GB2}\;(y;a,b,p,q=1). \tag{8.6}$$

This is actually a Dagum Type 1 distribution. Dagum's more general form has the cumulative probability function (cdf): $F\,(y) = \alpha + (1-\alpha)\left(1+(y/b)^{-a}\right)^{-p}$. Dagum's Types 1, 2, and 3 correspond to α=0, 0<α<1, and α<0, respectively. Dagum's Type 2 model allows for non-positive values of Y with $F(0) = \alpha$. Type 3 is associated with a positive lower bound for Y, y_0. A generalization of this formulation is given by $F(y) = \alpha + (1-\alpha)F^*(y)$ where $F^*(y)$ could denote any cdf for positive Y, such as a GB1, GB2, or GB. An alternative formulation could be viewed as arising from a "translation of the origin" to y_0 where y_0 can be negative, zero, or positive. The value of y_0 can be estimated from other information such as the fraction of negative and zero observations for Dagum's Type 2 model or can be estimated as a parameter. Bandourian *et al.* (2003) include an example of a translated origin in the estimation of the models considered in this paper.

The Singh-Maddala (SM) distribution is also a special case of the GB2

$$\mathrm{SM}(y;a,b,q) = \frac{|a|\,qy^{a-1}}{b^a\left(1+(y/b)^a\right)^{q+1}} = \mathrm{GB2}(y;a,b,p=1,q) \tag{8.7}$$

and the generalized gamma (GG) distribution is a limiting case of the GB2 defined as

$$\begin{aligned}\mathrm{GG}(y;a,\beta,p) &= \lim_{q\to\infty} \mathrm{GB}(y;a,b=q^{1/a}\beta,c=1,p,q) \\ &= \frac{y^{ap-1}e^{-(y/\beta)^a}}{\beta^a\Gamma\,(p)}\end{aligned} \tag{8.8}$$

where $\Gamma\,(p)$denotes the gamma function. The cumulative distribution functions (cdf) for the Dagum and Singh-Maddala distributions have closed form representations, but the cdf for the generalized gamma involves an infinite series.

The GB2 can be expressed as a mixture of a generalized gamma and an inverse generalized gamma distribution

$$\mathrm{GB2}(y;a,b\beta,p,q)=\int_0^\infty \mathrm{GG}(y/\lambda;a,\beta,p)\,\mathrm{IGG}(\lambda;a,b,q)\,d\lambda$$

where the IGG distribution is a GG with a negative value of the parameter a. This mixture interpretation can be used as a model for a multiplicative measurement error model where (λ) denotes the multiplicative measurement error and true income is distributed as a GG, Israelsen and McDonald (2003).

The two-parameter gamma (GA) and lognormal (LN) distributions used by Salem and Mount (1974) are both special cases of the generalized gamma pdf and are defined by

$$\mathrm{GA}(y;p,\beta)=\frac{y^{p-1}e^{-(y/\beta)}}{\beta\Gamma(p)}=\mathrm{GG}(y;a=1,\beta,p) \tag{8.9}$$

$$\begin{aligned}\mathrm{LN}(y;\mu,\sigma)&=\left(\frac{e^{-\left(\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right)}}{y\sqrt{2\pi}\sigma}\right)\\&=\lim_{a\to 0}\mathrm{GG}\left(y;a,\beta=\left(\sigma^2a^2\right)^{1/a},p=(a\mu+1)/\sigma^2a^2\right).\end{aligned} \tag{8.10}$$

The Weibull distribution is also a special case of the generalized gamma

$$\mathit{Weibull}(y;\beta,q)=\frac{y^{a-1}e^{-(y/\beta)^a}}{\beta^a}=\mathrm{GG}(y;a,\beta,p=1,q). \tag{8.11}$$

A convenient way to visualize these relationships and some other special cases mentioned in the introduction is the distribution tree in figure 8.1 where *Beta*1 and *Beta*2 are the beta of the first and second kind, respectively. Additional details can be found in McDonald and Xu (1995).[18]

[18] Some generalizations of this model could include situations in which the distributional parameters are functions of explanatory variables or regression models for positive random variables. Examples of the former could include the distributional parameters of the distribution of income being functions of age and education level, $\theta=\theta(x)$ where θ denotes a vector valued function of explanatory variables which needs to be estimated. The result would be a different income distribution for each age and education level. Examples of regression models with positive random variables could take the form $y_i=\ln(T_i)=X_i\beta+\varepsilon_i$ where T_i, $(0< T_i)$, is the dependent variable to be modeled and could be selected to be a distribution from the distributional family depicted in Figure 8.1. McDonald and Butler (1990) give some additional details related to this regression formulation.

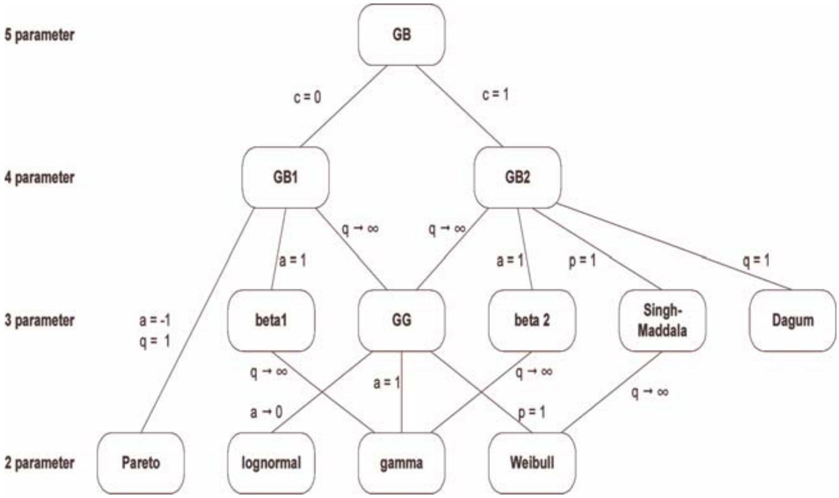


Fig. 8.1: Distribution tree

2.2 Parameter estimation and measures of goodness of fit

Maximum likelihood estimation (MLE) is probably the most common method of estimating the distributional parameters for these models. For individual observations $(y_i : i = 1, 2, ..., n)$ and for data reported in a grouped format, respectively, the MLE of θ are obtained by maximizing

$$\ell(\theta) = \sum_{i=1}^{N} \ell n\left(f_d\left(y_i : \theta\right)\right) \qquad (8.12\, a-b)$$

$$\ell(\theta) = \ell n\,(N!) + \sum_{i=1}^{g} \left\{n_i \ell n\left[p_i(\theta)\right] - \ell n\left(n_i!\right)\right\}$$

over θ, with $p_i(\theta) = F_d(Y_i : \theta) - F_d(Y_{i-1} : \theta)$ where$f_d()$ and $F_d()$ denote the pdf and cdf for distribution type d, θ is a vector containing the distributional parameters, Y_i and Y_{i-1} are the upper and lower bounds of the ith of g data groups, n_i is the number of observations in the ith group, and N is the total number of observations. Two alternatives to MLE for grouped data are the minimum chi-squared and minimum modified chi-squared estimators which correspond to minimizing

$$\sum_{i=1}^{g} \left(\frac{\left(n_i - Np_i(\theta)\right)^2}{Np_i(\theta)}\right) \qquad (8.13\, a-b)$$

$$\sum_{i=1}^{g}\left(\frac{(n_i - Np_i(\theta))^2}{n_i}\right)$$

over θ. The MLE, minimum chi-squared and minimum modified chi-squared estimators of θ are asymptotically efficient, Cox and Hinkley (1974, p.306). The chi-square values can be considered as goodness of fit indices which also provide the basis for statistical tests. The χ^2 statistic is asymptotically distributed as a chi-square with degrees of freedom equal to the one less than the difference between the number of groups and the number of estimated parameters, Cox and Hinkley (1974, p.316). The derivation of any of these estimators generally involves the use of numerical optimization algorithms.

Two other goodness of fit indices are the sum of squared errors (SSE) and sum of absolute errors (SAE), respectively defined by

$$\text{SSE} = \sum_{i=1}^{g}\left(\frac{n_i}{N} - p_i(\theta)\right)^2 \quad \text{and} \qquad (8.14\,a-b)$$

$$\text{SAE} = \sum_{i=1}^{g}\left|\frac{n_i}{N} - p_i(\theta)\right|.$$

Minimizing SSE or SAE over θ could be thought of as yielding a least squares or a least absolute errors estimator; however, these estimators will not be efficient.

Testing nested hypotheses, such as H_0: GB2 = Dagum, can be performed using the likelihood ratio test statistic, defined by

$$LR = 2\left[\hat{\ell} - \hat{\ell}^*\right] \sim_a \chi^2(r), \qquad (8.15)$$

where $\hat{\ell}$ and $\hat{\ell}^*$ respectively represent the optimized log-likelihood values corresponding to the unconstrained (GB2) and nested (Dagum) models and r (the degrees of freedom for the asymptotic chi-square) is the difference in the number of estimated parameters in the two model specifications. Thus, degrees of freedom of the chi-square test statistic in testing the equivalence of the GB2 and Dagum distributions would be one. Nested models on the boundary of the parameter space may compromise the appropriateness of $\chi^2(r)$.

2.3 Measures of Inequality

While the cumulative distribution function of income uniquely characterizes distributional characteristics of income, alternative functions can facilitate a comparison of the relative inequality of two distributions of income. For example, the Lorenz curve depicts the relationship between the percent of income received by different percentages of a given population. The Lorenz curve can be formally defined by

$$L(p) = \mu^{-1}\int_0^p F^{-1}(t)\,dt \qquad 0 \le p \le 1 \qquad (8.16)$$

where $L(p)$ denotes the fraction of total income that holders of the lowest pth fraction of income possess and μ is the mean income, Gastwirth (1971). In comparing two populations using Lorenz curves, population 2 is said to be more egalitarian than population 1 if $L_1(p) \leq L_2(p)$ for all p, $0 < p < 1$. Sarabia *et al.* (2002) consider Lorenz orderings for distributions in the GB2 family.

Numerous scalar measures of inequality have been considered in the literature, including the coefficient of variation (CV), the Pietra index (P), the standard deviation of logarithms (H), Theil's entropy measure (T), and the Gini coefficient (G). In this paper we will consider the use of numerical and analytic methods to evaluate G, P, and T defined by

$$G = \left(\frac{1}{2\mu}\right) E\left(|Y-X|\right) = \left(\frac{1}{2\mu}\right) \int_0^\infty \int_0^\infty |x-y| f(x) f(y)\, dx dy \qquad (8.17\ a-c)$$

$$= 1 - \frac{\int_0^\infty (1-F(y))^2 dy}{\int_0^\infty (1-F(y))\, dy} \text{ (Dorfman, 1979)}$$

$$P = \left(\frac{1}{2\mu}\right) E\left(|Y-\mu|\right) = \left(\frac{1}{2\mu}\right) \int_0^\infty |y-\mu| f(y)\, dy$$

$$T = E\left(\ln\left((Y/\mu)^{Y/\mu}\right)\right) = \int_0^\infty \left(\frac{y}{\mu}\right) \ln\left(\frac{y}{\mu}\right) f(y)\, dy.$$

The Gini Coefficient can be interpreted as twice the area of concentration between the Lorenz curve and the 45 degree line of perfect equality and the Pietra Index is twice the area of the largest triangle which can be inscribed in the area of concentration; thus, P $\leq$ G. Each of the income inequality measures can be expressed in terms of the underlying distributional parameters (θ). Alternative measures of inequality are frequently compared on the basis of such characteristics as their sensitivity to changes in units of measurement or to the impact of income transfers from one group to another. For example, the Pigou-Dalton Principle states that an income transfer from a richer person to a poorer person always reduces inequality and is satisfied by the Gini and Theil Coefficients, but not by the Pietra Index which is invariant to income transfers on the same side of the mean. The known equations expressing G in terms of the distributional parameters (θ) are taken from Dagum (1977) and McDonald (1984) and are summarized in Dastrup *et al.* (2007). Some similar results for the Pietra and Theil indices are summarized in McDonald (1981) for the Pareto, lognormal, gamma, beta, and Singh-Maddala distributions. This paper provides equations for the Pietra and Theil derived from the GG, GB1 and GB2 distributions; hence, for all of their special cases.

3 Evaluation of the Gini, Pietra, and Theil Measures

Given the results reported in this paper, analytic expressions for G, P, and T corresponding to the GB1 or GB2 or any of their special cases such as shown in figure 8.1 can be readily obtained from the LN, GG, GB1 or GB2 results. Similar results for the GB have not been derived, but numerical estimates can be obtained using numerical integration.

3.1 Analytic expressions for G, P, and T

Expressions for the Gini coefficient corresponding to the LN, GG, GB1 and GB2 were derived and/or reported in McDonald (1984) and can be written as

$$G_{LN} = 2\,\mathrm{N}\left[\sigma/\sqrt{2};0,1\right] - 1 \qquad (8.18\; a-d)$$

$$G_{GG} = \frac{1}{2^{2p+1/a}B(p,p+1/a)} \times \left\{ \left(\frac{1}{p}\right) {}_2F_1\left[\begin{matrix} 1,\, 2p+1/a; \frac{1}{2} \\ p+1; \end{matrix}\right] - \left(\frac{1}{p+1/a}\right) {}_2F_1\left[\begin{matrix} 1,\, 2p+1/a\; ; \frac{1}{2} \\ p+1/a+1; \end{matrix}\right] \right\}$$

$$G_{GB1} = \frac{B(2p+1/a,q)}{B(p,q)\,B(p+1/a,q)\,p\,(ap+1)} \times {}_4F_3\left[\begin{matrix} 2p+1/a,\, p,\, p+1/a,\, 1-q; 1 \\ 2p+q+1/a,\, p+1,\, p+1/a+1; \end{matrix}\right]$$

$$G_{GB2} = \frac{B(2q-1/a,2p+1/a)}{B(p,q)\,B(p+1/a,q-1/a)} \times \left\{ \left(\frac{1}{p}\right) {}_3F_2\left[\begin{matrix} 1,\, p+q,\, 2p+1/a; 1 \\ p+1,\, 2(p+q); \end{matrix}\right] - \left(\frac{1}{p+1/a}\right) {}_3F_2\left[\begin{matrix} 1,\, p+q,\, 2p+1/a; 1 \\ p+1/a+1,\, 2(p+q); \end{matrix}\right] \right\}$$

where $N[x;0,1] = .5 + \frac{xe^{-x^2/2}}{\sqrt{2\pi}}\; {}_1\mathrm{F}_1\left[\begin{matrix} 1;\; x^2/2 \\ \frac{3}{2}; \end{matrix}\right]$ denotes the cdf for a standard normal evaluated at x, ${}_pF_q\left[\begin{matrix} a_1\; \dots\; a_p;\, x \\ b_1\; \dots\; b_q; \end{matrix}\right] = \sum_{i=0}^{\infty} \frac{(a_1)_i \cdots (a_p)_i x^i}{(b_1)_i \cdots (b_q)_i i!}$ represents the generalized hypergeometric series with $(a)_i = (a)(a+1)\cdots\;(a+i-1) = \Gamma(a+i)/\Gamma(a)$, and $\Gamma(x)$ denotes the gamma function, Rainville (1960). The hypergeometric series in G_{LN} , G_{GG} , $\mathrm{G}_{\mathrm{GB1}}$, and $\mathrm{G}_{\mathrm{GB2}}$ involve approximating infinite series and can present computational problems.

Derivations for the Pietra indices are outlined in Appendix I and can be shown to be

$$P_{LN} = 2\,\mathrm{N}[\sigma/2;\ 0,1] - 1 \qquad (8.19\,a-d)$$

$$P_{GG} = \frac{e^{-\left(\frac{\mu}{\beta}\right)^a}\left(\frac{\mu}{\beta}\right)^{ap}}{\Gamma(p)} \times$$

$$\left\{\frac{1}{p}\,{}_1F_1\left[\begin{matrix}1\ ; \\ p+1;\end{matrix}\ \left(\frac{\mu}{\beta}\right)^a\right] - \left(\frac{1}{p+1/a}\right)\,{}_1F_1\left[\begin{matrix}1\ ; \\ p+1/a+1;\end{matrix}\ \left(\frac{\mu}{\beta}\right)^a\right]\right\}$$

$$P_{GB1} = \frac{\left(\frac{\mu}{b}\right)^{ap}}{B(p,q)} \times$$

$$\left\{\frac{1}{p}\,{}_2F_1\left[\begin{matrix}p, & 1-q; \\ p+1; & \end{matrix}\ \left(\frac{\mu}{b}\right)^a\right] - \left(\frac{1}{p+1/a,}\right)\,{}_2F_1\left[\begin{matrix}p+1/a, & 1-q; \\ p+1/a+1; & \end{matrix}\ \left(\frac{\mu}{b}\right)^a\right]\right\}$$

$$P_{GB2} = \left(\frac{z^{ap}}{B(p,q)}\right) \times$$

$$\left\{\left(\frac{1}{p}\right)\,{}_2F_1\left[\begin{matrix}p, & 1-q; \\ p+1; & \end{matrix}\ z\right] - \left(\frac{1}{p+1/a}\right)\,{}_2F_1\left[\begin{matrix}p+1/a, & 1+1/a-q; \\ p+1/a+1; & \end{matrix}\ z\right]\right\},$$

where $z = \left(\frac{(\mu/b)^a}{1+(\mu/b)^a}\right)$ and μ denotes the respective means for the GG, GB1 and GB2.

The equations for the Theil indices, derived in the Appendix II, can be written as

$$T_{LN} = \frac{\sigma^2}{2} \qquad (8.20\,a-d)$$

$$T_{GG} = \frac{1}{a}\psi(p+1/a) + \ell n(\beta/\mu)$$

$$T_{GB1} = \left(\frac{1}{a}\right)[\psi(p+1/a) - \psi(p+q+1/a)] + \ell n(b/\mu)$$

$$T_{GB2} = \left(\frac{1}{a}\right)[\psi(p+1/a) - \psi(q-1/a)] + \ell n(b/\mu)$$

where $\psi(x)$ denotes the digamma function which can be defined in terms of the gamma function as $\psi(x) = \frac{d\ell n\Gamma(x)}{dx}$.

The Theil, Pietra, and Gini indices potentially involve infinite series as does the digamma function. In these cases there are potential computational problems involved in evaluating the corresponding results. An alternative approach is to use numerical integration routines to evaluate the integrals (8.17 $a-c$) associated with the expectations defining the inequality index .

The LN, GG, GB1, and GB2 results given above can be used to obtain results for their special cases. For example, we note from figure 8.1 that letting q = 1 in the GB2 yields the Dagum distribution; hence, substituting q=1 into the GB2 equations ((8.18 d), (8.19 d), and (8.20 d)) for G, P, and T and simplifying yields the Dagum results:

$$G_{Dagum} = \frac{B(p,p)}{B(p,p+1/a)} - 1 \qquad (8.21\ a-c)$$

$$P_{Dagum} = (z^{ap}) \left\{ 1 - \left(\frac{p}{p+1/a} \right) {}_2F_1 \left[\begin{matrix} p+1/a, & 1/a; \; z \\ p+1/a+1; & \end{matrix} \right] \right\}$$

$$T_{Dagum} = \left(\frac{1}{a} \right) [\psi(p+1/a) - \psi(1-1/a)] + \ell n(b/\mu).$$

3.2 Estimation of inequality measures

Income data are either reported with individual observations or in a grouped format. For samples with individual observations there are two approaches to estimating measures of income inequality. The first might be thought of as being nonparametric with the inequality measures being estimated as natural extensions of their definitions:

$$\hat{G} = \left(\frac{1}{2\bar{Y}} \right) \sum_{i,\,j} |y_i - y_j| \left(\frac{1}{N^2} \right) \qquad (8.22\ a-c)$$

$$\hat{P} = \left(\frac{1}{2\bar{Y}} \right) \sum_{i} |y_i - \bar{Y}| \left(\frac{1}{N} \right)$$

$$\hat{T} = \sum_{i=1}^{N} \left(\frac{y_i}{\bar{Y}} \ln \left(\frac{y_i}{\bar{Y}} \right) \right) \left(\frac{1}{N} \right).$$

The second approach is parametric and involves estimating the distributional parameters $\left(\tilde{\theta}\right)$and then substituting these estimates into the equations (8.18 $a-d$), (8.19 $a-d$), and (8.20 $a-d$) to yield:

$$\tilde{G}_d = G_d\left(\tilde{\theta}\right) \qquad (8.23\ a-c)$$

$$\tilde{P}_d = P_d\left(\tilde{\theta}\right)$$

$$\tilde{T}_d = T_d\left(\tilde{\theta}\right)$$

where d denotes the distributional type, e.g., for d = Dagum

$$G_{Dagum} = \frac{B(\hat{p},\hat{p})}{B(\hat{p},\hat{p}+1/\hat{a})} - 1.$$

For data in a grouped format, the same two approaches (nonparametric and parametric) are available with the y_i in the "nonparametric" estimating equations for G, P, or T usually being interpreted as the group mean or group midpoint. Both of these interpretations neglect intra-group variation, with the midpoint approach also ignoring problems where the last income interval is of the form $[Y_{g-1}, \infty)$. McDonald and Ransom (1979a,b) explore a few of these issues and find that estimates of population characteristics depend on the assumed functional form and estimation technique with estimators allowing for intra-group variation performing better than those which do not.

4 Application

In order to illustrate the concepts in the previous sections we will fit the distributions discussed to a sample of U.S. Family income for the year 2000. The data used consist of 1,010,418 positive observations. The corresponding nonparametric estimates of G, P, and T, based on individual observations (equations (8.22 $a-c$)), are 0.4624, .3298, and .3867. In the example below, for computational tractability, we have grouped the individual observations into 50 equal probability bins. These groupings are reported in Appendix III. The grouped methods outlined in Section 2.2 are used to estimate the unknown distribution parameters. Simplex search algorithms in Matlab were used to maximize the log-likelihood function (equation (8.12 b)) with a convergence criterion of 10 (-7). The estimated parameters and standard errors[19], along with four goodness of fit measures, for the Weibull, lognormal, gamma, generalized gamma, Dagum, Singh-Maddala, GB1, GB2, and GB distributions are reported in Table 8.1.

Based on the SAE and SSE, the gamma is the best fitting two-parameter distribution; however, the LN has a smaller chi-square value. While the log-likelihood value does not facilitate a likelihood ratio test for non-nested models, model selection criteria such as Akaike's Information Criterion (AIC= $-\ell\left(\hat{\theta}\right)$ + r) will select, when comparing models with the same number of parameters (r), the model with the maximum likelihood value. Introducing a third parameter produces statistically significant improvements when comparing the generalized gamma with either of its special cases (Weibull or gamma) or limiting cases (lognormal) using a likelihood ratio test. However, two other three parameter distributions (Dagum and Singh-Maddala) appear to yield a better fit to the data than does the generalized gamma with the Dagum being the best three –parameter distribution of those

[19] The standard errors are obtained from the square root of the Hessian matrix obtained after the Simplex optimizations are completed.

Table 8.1: Parameter Estimates and Goodness of Fit Measures for 2000 Family Income Data*

Distribution	a(μ)	b(σ)	c	p	q	SAE	SSE	Chi2	logL
2-parameter distributions									
Weibull	1.1806	56568.1				0.18656	0.00140	107839.8	-40838.0
	[0.00091]	[50.568]							
Lognormal	10.5046	0.9439				0.19021	0.00117	69027.3	-31188.9
	[0.00094]	[0.00069]							
Gamma		37604.5		1.4174		0.14935	0.00088	83887.0	-30103.8
		[59.636]		[0.0020]					
3-parameter distributions									
GG	0.4835	1434.8		5.2566		0.13875	0.00065	35074.0	-17113.8
	[0.0046]	[103.33]		[0.0951]					
Dagum	2.4126	56971.6		0.5660		0.10176	0.00036	19488.3	-9717.3
	[0.0045]	[130.31]		[0.0019]					
Singh-Maddala	1.5807	67530.9			1.97101	0.10885	0.00041	21114.7	-10619.8
	[0.0022]	[334.17]			[0.0106]				
4-parameter distributions									
GB1	0.483052	1.10E+11		5.2623	6438.1	0.13875	0.00065	35088.6	-17118.9
	[0.0011]	[397.18]		[0.022]	[76.03]				
GB2	2.2474	58441.5		0.6186	1.118	0.10148	0.00036	19380.5	-9673.9
	[0.0185]	[226.82]		[0.0066]	[0.0142]				
5-parameter distribution									
GB	2.2529	58380.2	1.0	0.6167	1.113	0.10148	0.00036	19380.5	-9673.9
	[0.0184]	[224.67]	[1.22e-19]	[0.0066]	[0.0142]				

*Data comes from Integrated Public Use Microdata Series (IPUMS) from US Census Bureau (http://usa.ipums.org/usa/) See Appendix III for additional details.

considered. The generalized gamma appears to be observationally equivalent to the four-parameter GB1 whether compared using the SAE, SSE, Chi square, or log-likelihood values[20]. The five-parameter GB is observationally equivalent to the four-parameter GB2 which gives a statistically significant improvement relative to any of its special cases (Singh-Maddala, Dagum, generalized gamma, gamma, lognormal, or Weibull).

Table 8.2: Income Inequality Indices – Family Income Data year 2000

	Gini Coefficient		Pietra Index		Theil Index	
Distribution	**analytic**	**numeric**	**analytic**	**numeric**	**analytic**	**numeric**
2-parameter distributions						
Weibull	0.4441	0.4441	0.3239	0.3239	0.3272	0.3271
Lognormal	0.4955	0.4955	0.3630	0.3630	0.4454	0.4454
Gamma	0.4346	0.4346	0.3162	0.3162	0.3130	0.3130
3-parameter distributions						
GG	0.4544	0.4544	0.3310	0.3310	0.3501	0.3501
Dagum	0.4708	0.4708	0.3358	0.3358	0.4284	0.4284
Singh-Maddala	0.4624	0.4624	0.3328	0.3328	0.3872	0.3871
4-parameter distributions						
GB1	0.4544	0.4544	0.3311	0.3311	0.3500	0.3500
GB2	0.4682	0.4683	0.3345	0.3345	0.4172	0.4172
5-parameter distribution						
GB	N/A	0.4683	N/A	0.3345	N/A	0.4172

Table 8.2 reports *analytic* and *numeric* estimates of the Gini, Pietra, and Theil indices corresponding to the estimated distributional parameters $\left(\tilde{\theta}\right)$ reported in Table 8.1. The *analytic* estimates are obtained by substituting $\tilde{\theta}$ into equations (8.23 $a-c$). The *numeric* estimates are obtained by numerically integrating the "expectational" definitions of G, P, and T (equations (8.17 $a-c$)) corresponding to the estimated pdf's and cdf's. The standard Matlab package includes the gamma, beta, and psi functions and can evaluate the Gini coefficients for all distributions except for the generalized gamma, GB1, GB2, and GB. Theil indices can be estimated for all distributions with the psi function and a knowledge of parameter estimates. Indices using various forms of the hypergeometric series require the Symbolic Math Toolbox in Matlab or programming the necessary functions into the basic Matlab package. An alternative approach to evaluating inequality indices when the required functions are not available is to use numerical integration. The Dorfman form for the Gini coefficient was used for numerical evaluations using the Matlab adaptive Simpson quadrature integration algorithm. The same integration algorithm was used to estimate the Pietra and Theil indices.

[20] While the numerical value of the log-likelihood value for the GG dominates that for the GB1, the two fitted distributions are observationally equivalent.

An analytic expression for the GB has not been derived; hence, the method of numerical integration was used to estimate corresponding values for G, P, and T. Given the remarkably close agreement between the analytic and numeric methods, when both are available, we feel that this approach is viable in the absence of analytical expressions when there are no numerical problems. One such potential numerical problem is when vertical asymptotes exist such as when $ap < 1$.

5 Summary and Discussion

There is an extensive literature that examines statistical distributions as models of the size distribution of income, going back at least to Pareto (1897). Contributions to this literature have generally taken the form of increasing flexibility by incorporating additional parameters in the distribution and have included the lognormal, gamma, Weibull, Dagum, Singh-Maddala, generalized gamma, and generalized beta types 1 and 2 distributions. The generalized beta is a very flexible distribution that includes as special or limiting cases virtually all previous distributions used as models. In this paper we have examined some theoretical and practical issues related to the use of the generalized beta to model the income distribution.

For all of the special and limiting cases, we provide formulas that express popular measure of income inequality, the Gini, Pietra and Theil indexes, as functions of the distributional parameters. Although formulas for the Gini index have been previously derived (McDonald, 1984), we have here added new results for the Pietra and Theil indexes. Thus, we provide an almost complete catalog of formulas for the inequality measures as they relate to parameters of the income distribution. The results are missing only for the generalized beta. For the generalized beta, we suggest a convenient method by which the inequality indexes may be calculated using numerical integration.

While there are nonparametric methods for estimating inequality indexes, there may be some advantages to using the approach of first fitting the income distribution and then using the related formula. First, in many cases data on income distribution are only available in grouped form. The grouping of data eliminates some information about how incomes vary within groups. In such cases, calculation of inequality indexes requires assumptions about the distribution of income within the groups, which this approach provides. (See Gastwirth (1972) for some discussion of this issue.)

Another reason for the parametric approach is to model how the income distribution changes with changes in exogenous variables. Thus, a researcher might relate certain parameters of the income distribution to certain exogenous variables, as in Thurow (1970). For example, one might be interested in studying how the rise of computerization has influenced income inequality. In this approach, we can see how such a variable, x, has affected income inequality by examining its effect on the parameters of the fitted income distribution, and through those parameters, the index of income inequality:

$$\varepsilon_{G:x} = \sum_i \varepsilon_{G:\theta_i} \varepsilon_{\theta_i:x}$$

where $\varepsilon_{G:x}$ denotes the elasticity of G with respect to x , $\varepsilon_{G:\theta_i}$ denotes the elasticity of G with respect to the ith distributional parameter , and $\varepsilon_{\theta_i:x}$ denotes the elasticity of the ith distributional parameter with respect to x. Hence, equations (8.23 $a-c$) and regression analysis can be used to determine the sensitivity of measures of income inequality with respect to various explanatory variables.

We have also examined some numerical issues related to fitting the distribution of income or the calculation of indexes of income inequality. In the case of the generalized beta, the indexes of income inequality as functions of the parameters of the underlying distribution have not yet been derived. However, it is possible to calculate the indexes through numerical integration methods, as we show. In other cases, the Gini index for the GG, the GB1 and the GB2, require the evaluation of generalized hypergeometric series. Some researches may find it easier to evaluate the numerical integral forms than to evaluate the generalized hypergeometric series, although they are available in Symbolic Math Toolbox in Matlab.

We have also provided an example of these techniques using the 1 percent public use micro sample from the 2000 United States Census.

Appendix I

The Gini and Pietra indices

The Gini and Pietra indices can be defined in terms of the integrals

$$I_d(x,h;\theta) = \int_0^x y^h f_d(y;\theta)\,dy$$

$$I_d^*(i,j;\theta) = \int_0^\infty x^i f_d(x;\theta) \int_0^x y^j f_d(y;\theta)\,dydx$$

as follows:

$$G_d(\theta) = E(|y-x|)/2\mu = (I_d^*(1,0;\theta) - I_d^*(0,1;\theta))/\mu$$

$$P_d(\theta) = E(|y-\mu|)/2\mu = I_d(\mu,0;\theta) - I_d(\mu,1;\theta)/\mu$$

where d denotes the distribution type and θ and μ, respectively, represent the distributional parameters and the mean.

McDonald (1984) reports the equations expressing the Gini coefficients in terms of the distributional parameters for the GB1, GB2, GG, B1, B2, Singh-Maddala, LN, GA, and Weibull distributions. The text of this paper reports the Gini coefficients for the LN, GG, GB1, and GB2 distributions from which the others can be obtained by appropriate substitutions. The G, P, and T indices associated with the lognormal can be obtained from results found in Aitchison and Brown (1969).

The Pietra indices for the GG, GB1, and GB2 can be obtained from the corresponding expressions for $I_d(x,h;\theta)$ given in McDonald (1984):

$$I_{GG}(x,h;\theta) = \frac{b^h e^{-(x/b)^a}(x/b)^{ap+h}}{\Gamma(p)(p+h/a)} \cdot {}_1F_1\begin{bmatrix} 1;\ (x/b)^a \\ p+\frac{h}{a}+1; \end{bmatrix}$$

$$I_{GB1}(x,h;\theta) = \frac{b^h (x/b)^{ap+h}}{B(p,q)(p+h/a)} \cdot {}_2F_1\begin{bmatrix} p+\frac{h}{a},\ 1-q;\ (x/b)^a \\ p+\frac{h}{a}+1; \end{bmatrix}$$

$$I_{GB2}(x,h;\theta) = \frac{b^h (z)^{ap+h}}{B(p,q)(p+h/a)} \cdot {}_2F_1\begin{bmatrix} p+\frac{h}{a},\ 1+\frac{h}{a}-q;\ z \\ p+\frac{h}{a}+1; \end{bmatrix}$$

where $z = \frac{(x/b)^a}{1+(x/b)^a}$.

Appendix II

The Theil index

The Theil Index, defined by

$$T_d(\theta) = E\left[\left(\frac{y}{\mu}\right)\ell n\left(\frac{y}{\mu}\right)\right]$$

can be written as

$$T = \left(\frac{1}{\mu}\right)\int_0^\infty y\ell n(y) f_d(y;\theta)\,dy - \ell n(\mu).$$

For the generalized gamma (d=GG), the Theil index is equal to

$$T_{GG} = \left(\frac{1}{\mu}\right)\int_0^\infty y\ell n(y)\left\{\frac{|a|\,y^{ap-1}e^{-(y/\beta)}}{\beta^{ap}\Gamma(p)}\right\}dy - \ell n(\mu)$$

where $\mu = \frac{\beta\,\Gamma(p+1/a)}{\Gamma(p)}$ is the mean of the generalized gamma. Making the change of variable for $s = (y/\beta)^a$ in T_{GG} and simplifying yields

$$T_{GG} = \left(\frac{\beta}{\mu\Gamma(p)}\right)\int_0^\infty (\ell n(\beta) + (1/a)\,\ell n(s))\,s^{p+1/a-1}e^{-s}ds - \ell n(\mu).$$

Using the integral representation for a gamma function, $\int_0^\infty s^{h-1}e^{-s}ds = \Gamma(h)$, and $\int_0^\infty \ell n(s)\,s^{h-1}e^{-s}ds = \Gamma(h)\,\psi(h)$, we can write

$$T_{GG} = \left(\frac{1}{a}\right)\psi(p+1/a) + \ell n(\beta/\mu).$$

The derivations for T_{GB1} and T_{GB2} follow the same format as for T_{GG} except that the final step in the derivation uses the integral definition of the beta function $B(h,q) = \int_0^1 s^{h-1}(1-s)^{q-1}\,ds$ and the two integrals

$$\int_0^1 \ell n(s)\, s^{h-1}(1-s)^{q-1}\,ds = B(h,q)\left[\psi(h) - \psi(h+q)\right]$$

$$\int_0^\infty \left(\frac{\ell n(s)\, s^{h-1}}{(1+s)^{h+q}}\right) ds = B(h,q)\left[\psi(h) - \psi(q)\right],$$

respectively.

Appendix III

The data

The 2000 Family Income data comes from the Integrated Public Use Microdata Series (IPUMS) from US Census Bureau (http://usa.ipums.org/usa/). The family income data includes negative, zero, and positive incomes. The example reported in this paper used the 1,010,418 positive observations. The total data set included 1,054,797 observations; hence, 44,379 non-positive observations (4%) were omitted in the estimation.

The upper bounds on the 50 bins are listed below.

3,600	6,000	7,390	9,000	10,200
11,800	13,000	14,600	16,000	17,500
19,000	20,030	21,760	23,100	24,600
26,000	27,400	29,000	30,020	31,900
33,300	35,000	36,500	38,200	40,000
41,600	43,400	45,200	47,300	49,500
51,100	53,420	55,900	58,220	60,700
63,500	66,400	70,000	73,000	77,000
81,000	86,000	92,000	99,100	107,200
119,300	135,820	160,300	238,300	greater than 238,300

References

Aitchison, J. and J. A. C. Brown (1969) *The Lognormal Distribution with Special References to Its Uses in Economics*, Cambridge University Press, Cambridge.

Ammon, O. (1895) *Die Gesellschaftsordnung und ihre Natürlichen Grundlagen*, Jena.

Amoroso, L. (1924-1925) Ricerche Intorno alla Curva dei Redditi Annali di Matematica, *Pura ed Applicata, Series 4-21*, **2**, 123–157.

Bandourian, R., J. B. McDonald and R. S. Turley (2003) A Comparison of Parametric Models of Income Distribution across Countries and over Time, *Estadística*, **55**, 135–152.

Bartels, C. P. A. and H. van Metelen (1975) Alternative Probability Density Functions of Income, Vrije University Amsterdam: Research Memorandum 29.

Bordley, R. F., J. B. McDonald and A. Mantrala (1996) Something New, Something Old: Parametric Models for the Size Distribution of Income, *Journal of Income Distribution*, **6**, 91–103.

Cox, D. R. and D. V. Hinkley (1974) *Theoretical Statistics*, Chapman and Hall, London.

Dagum, C. (1977) A New Model for Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, **30**, 413–437.

Dagum, C. (1980) The Generation and Distribution of Income, the Lorenz Curve and the Gini Ratio, *Economie Appliquée*, **33**, 327–367.

Dastrup, S. R., R. Hartshorn and J. B. McDonald (2007) The Impact of Taxes and Transfer Payments on the Distribution of Income: A Parametric Comparison, *Journal of Economic Inequality*, **5**, 353–369.

Dorfman, R. (1979) A Formula for the Gini Coefficient, *Review of Economics and Statistics*, **61**, 146–149.

Gastwirth, J. L. (1971) A General Definition of the Lorenz Curve, *Econometrica*, **39**, 1037–1039.

Gastwirth, J. L. (1972) The Estimation of the Lorenz Curve and Gini Index, *Review of Economics and Statistics*, **54**, 306–316.

Gibrat, R. (1931) *Les Inégalités Économiques*, Librairie du Recueil Sirey, Paris.

Gini, C. (1912) Variabilita' e Mutabilita, Studio Economicogiuridici, Universita di Cagliari. 3:2a. Reprinted in C. Gini (1955), 211-382.

Israelsen, D. and J. B. McDonald (2003) Measurement Error and the Distribution of Income, *Journal of Income Distribution*, **12**, 21–31.

March, L. (1898) Quelques Exemples de Distribution des Salaires, *Journal de la Société Statistique de Paris*, pp. 193–206 and 241–248.

McDonald, J. B. (1981) Some Issues Associated with the Measurement of Income Inequality, in C. Taillie, G. P. Patil and B. Balderssari (eds.) *Statistical Distributions in Scientific Work*, vol. 6, Reidel, Boston.

McDonald, J. B. (1984) Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52**, 647–663.

McDonald, J. B. and R. J. Butler (1990) Regression Models for Positive Random Variables, *Journal of Econometrics*, **43**, 227–251.

McDonald, J. B. and M. Ransom (1979a) Alternative Parameter Estimators Based on Grouped Data, *Communications in Statistics: Theory and Methods*, **A8**, 899–917.

McDonald, J. B. and M. Ransom (1979b) Functional Forms, Estimation Techniques and the Distribution of Income, *Econometrica*, **47**, 1513–1525.

McDonald, J. B. and Y. J. Xu (1995) A Generalization of the Beta Distribution with Applications, *Journal of Econometrics*, **66**, 133–152, Erratum: Journal of Econometrics, **69**: 427-428.

Pareto, V. (1895) La Legge della Domanda, *Giornale degli Economisti*, 10, 59-68. English Translation in Rivista di Politica Economica, 87 (1997), 691-700 .

Pareto, V. (1897) *Cours d'Economie Politique*, Rouge, Lausanne.

Parker, S. C. (1999) The Generalized Beta as a Model for the Distribution of Earnings, *Economics Letters*, **62**, 197–200.

Rainville, E. D. (1960) *Special Functions*, MacMillan, New York.

Salem, A. B. Z. and T. D. Mount (1974) A Convenient Descriptive Model of Income Distribution: The Gamma Density, *Econometrica*, **42**, 1115–1127.

Sarabia, J. M., E. Castillo and D. J. Slottje (2002) Lorenz Ordering between Mcdonalds Generalized Functions of the Income Size Distribution, *Economics Letters*, **75**, 265–270.

Singh, S. K. and G. S. Maddala (1976) A Function for the Size Distribution of Incomes, *Econometrica*, **44**, 963–970.

Taillie, C. (1981) Lorenz ordering within the Generalized Gamma Family of Income Distributions, in C. Taillie, G. P. Patil and B. Balderssari (eds.) *Statistical Distributions in Scientific Work*, vol. 6, pp. 181–192, Reidel, Boston.

Thurow, L. C. (1970) Analyzing the American Income Distribution, *American Economic Review*, **48**, 261–269.

CHAPTER 9

Parametric Lorenz Curves: Models and Applications

José María Sarabia†

Abstract

The Lorenz curve (LC) is an important instrument for analyzing the size of distribution of income or wealth and inequality. Finding an appropriate functional form is an important practical and theoretical problem. In this chapter we study parametric models for the LC and some important applications.

The basic properties that a function should satisfy in order to be a genuine LC are discussed. Next, we study the different ways for generating parametric families of LCs, as well as some of their basic properties, including their relationship with the underlying income distribution function. The basic parametric models proposed in the literature are studied, including the Pareto, lognormal and other important families of LCs.

Some general strategies to obtain extensions and generalizations of the basic parametric models are presented. One of the main applications of LCs is the study of inequality. We begin studying different measures of inequality together with their expressions in terms of the LC. These measures include the Gini index and some of their generalizations proposed by Kakwani (1980) and Yitzhaki (1983). Their corresponding expressions for the proposed parametric families of LCs will be obtained. The Lorenz ordering is also studied. The Lorenz ordering is a partial order that allows the comparison of two distributions when its corresponding LCs do not intersect. Some basic properties of this order are studied, including the effect of transformations, its relations with other partial orderings and their application to important parametric income distributions. The recent proposal of multivariate versions of the LC are studied. Finally, some applications of the Lorenz curve are presented.

---

† Department of Economics, University of Cantabria, Avda de los Castros s/n, 39005 Santander, Spain. E-mail: sarabiaj@unican.es

1 Introduction

The merits of parametric methods as opposed to non-parametric methods for the construction of indices and inequality measures for income distributions have been pointed out by Slottje (1990) and Ryu and Slottje (1996, 1999). These authors conclude, among other things, that the indices should be constructed using the parametric method and then the results checked using a non-parametric method. In this regard, the Lorenz curve (LC) is an essential instrument for analyzing the size distribution of income, wealth and inequality and the problem of finding an appropriate functional form is both an important practical and theoretical problem.

Some recent advances have contributed to the current development of this research instrument. New ways to specify the Lorenz curve have been developed and studied (see Section 4). On the other hand, the Lorenz ordering has been characterized in important families of income distributions (see Kleiber and Kotz (2003) and Section 6). The interest in and development of multivariate inequality measures as well as the multivariate versions of the Lorenz curve (see Section 8) have led to an increase in the amount of research devoted to this area. In this chapter we study parametric models for the LC and some important applications.

The contents of this chapter are as follows. In Section 2 we study basic properties of the LC, including their relationship with the underlying income distribution function. Section 3 reviews the LC of some important income models, including the following distributions: classical Pareto, lognormal, Singh-Maddala and Dagum type I. There exists a variety of approaches for the construction of parametric families of LC's. In Section 4 we study the different ways of generating parametric families of LC and some general strategies to obtain extensions and generalizations of the basic parametric models. One of the main applications of the LC's is the study of inequality. Inequality indices derived from the Lorenz curve and other classical inequality measures are studied in Section 5. The Lorenz ordering is a partial order that allows us to compare two distributions when their corresponding LC's do not intersect. Properties of this order, and their application to important parametric income distributions are studied in Section 6. Section 7 presents some variations of the LC. The recent proposal of multivariate versions of the LC are studied in Section 8. Finally, some applications of the Lorenz curve are presented in Section 9.

2 The Lorenz Curve. Basic Properties

The Lorenz curve is defined by points $(p, L(p))$, where p represents the cumulative proportion of income-receiving units, and $L(p)$ the cumulative proportion of incomes, when the incomes are arranged in ascending order of magnitude.

In the empirical case, if we denote the ordered individual incomes in the population by $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$, then for $i = 1, 2, \ldots, n$

$$L(\frac{i}{n}) = \frac{\sum_{j=1}^{i} x_{j:n}}{\sum_{j=1}^{n} x_{j:n}}. \tag{9.1}$$

The points $(\frac{i}{n}, L(\frac{i}{n}))$ are then linearly interpolated to complete the corresponding Lorenz curve.

Now, our next step is to extend (9.1) to the continuous case. If n is large the distribution of incomes within the population can be approximated by a continuous distribution function $F(x)$, with density $f(x)$ related by $F(x) = \int_0^x f(y)dy$. The interpretation here is similar to the previous one: for each positive x, $F(x)$ represents or approximates the proportion of individuals in the population whose income is less than or equal to x. Now, we consider the k-moment distribution of the population $F_{(k)}(x)$ defined by

$$F_{(k)}(x) = \frac{\int_0^x y^k dF(y)}{\int_0^\infty y dF(y)}, \ \ k = 1, 2, \ldots \tag{9.2}$$

where the denominator is assumed to be finite. If we set $k = 1$ in (9.2), then for each x, $F_{(1)}(x)$ represents the proportion of the total incomes which accrues to individuals with incomes less than or equal to x. The Lorenz curve corresponding to the distribution F can be described as the set of points,

$$(F(x), F_{(1)}(x)) \tag{9.3}$$

defined in the unit square, where x ranges from 0 to ∞ completed if necessary by linear interpolation.

An expression for the Lorenz curve can be constructed using the parametric representation (9.3). We may write

$$L(p) = F_{(1)}(F^{-1}(p)). \tag{9.4}$$

To use formula (9.4) we obviously need closed form expressions for $F_{(1)}$ and F^{-1}.

Let $\mathscr{L}$ be the class of all non-negative random variables with positive finite expectations. For a random variable X in $\mathscr{L}$ with cumulative distribution function F_X we define its inverse distribution function by

$$F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\} \tag{9.5}$$

Note that the mathematical expectation of X is $\mu_X = \int_0^1 F_X^{-1}(y)dy$. According to Gastwirth (1971) we have the following definition.

Definition 9.1. Let $X \in \mathscr{L}$ with cumulative distribution function F_X and inverse distribution function F_X^{-1}. The Lorenz curve L_X corresponding to X is defined by

$$L_X(p) = \frac{1}{\mu_X} \int_0^p F_X^{-1}(y)dy, \ \ 0 \leq p \leq 1. \tag{9.6}$$

This definition contains the definition provided by (9.1) in the case of a finite population and (9.2) in the continuous case. In formula (9.6) F_X^{-1} is piecewise continuous and the integrals can be assumed to be ordinary Riemann integrals.

From definition (9.6) we can show that a Lorenz curve will be a continuous, non-decreasing convex function that is differentiable almost everywhere in $[0,1]$ and $L(0)=0$ and $L(1)=1$. These are properties that we expect to characterize an LC. A formal characterization of a Lorenz curve attributed to Gaffney and Anstin by Pakes (1981) is the following.

Theorem 9.1. *Suppose $L(p)$ is defined and continuous on $[0,1]$ with second derivative $L''(p)$. The function $L(p)$ is a Lorenz curve if and only if*

$$L(0)=0,\ \ L(1)=1,\ \ L'(0^+)\geq 0,\ \ L''(p)\geq 0\ \ in\ \ (0,1). \tag{9.7}$$

The Lorenz curve determines the distribution of X up to a scale factor transformation. This is true since $F_X^{-1}(x)=\mu_X L'(x)$ almost everywhere and F_X^{-1} will determine F_X. Concerning the probability density function $f_X(x)$ associated with a Lorenz curve $L(p)$, we have the following result (Arnold, 1987).

Theorem 9.2. *If $L''(p)$ exists and is positive everywhere in an interval (x_1,x_2), then F_X has a finite positive density in the interval $(\mu L'(x_1^+),\mu L'(x_2^-))$ which is given by*

$$f_X(x)=\frac{1}{\mu L''(F_X(x))}. \tag{9.8}$$

As an illustration of these results, we consider Chotikapanich's LC defined in (9.28). The cumulative distribution function corresponding to this LC model is

$$F(x;k,\mu)=\frac{1}{k}\log\left(\frac{x}{c_k\mu}\right),\ \ c_k\mu\leq x\leq c_k\mu e^k,$$

where $c_k=k/(e^k-1)$ and $F(x;k,\mu)=0$ if $x\leq c_k\mu$ and $F(x;k,\mu)=1$ if $x\geq c_k\mu e^k$. Note that the cdf depends on k and a new scale parameter μ which represent the population mean.

From a geometric viewpoint it is natural to enquire whether an LC exhibits symmetry. A Lorenz curve is symmetric if

$$L[1-L(p)]=1-p,\ \ 0\leq p\leq 1. \tag{9.9}$$

If a random variable X has mean μ and density $f_X(x)$, its LC is symmetric if and only if

$$\frac{f_X(\mu^2/x)}{f_X(x)}=\left(\frac{x}{\mu}\right)^3,$$

for every x with $f_X(x)>0$.

3 Lorenz Curves of Some Classical Income Distributions

In this section we review the LC corresponding to some important models of income distributions. We begin with models corresponding to income distributions with closed expressions for the inverse cdf which can be integrated so that and then Gastwirth's formula can be used. The first example corresponds to the classical Pareto distribution (see Arnold (1983)) with cumulative distribution function

$$F_X(x) = 1 - \left(\frac{x}{\sigma}\right)^{-\alpha}, \; x \geq \sigma \tag{9.10}$$

where $\sigma > 0$ is a scale parameter and $\alpha > 0$ a shape parameter. For the Pareto distribution the quantile function is,

$$F_X^{-1}(y) = \sigma(1-y)^{-1/\alpha}, \; 0 < y < 1$$

and the mean $\mu_X = \alpha\sigma/(\alpha-1)$ if $\alpha > 1$. Using (9.6) we obtain

$$L_X(p) = \frac{\alpha-1}{\alpha\sigma}\int_0^p \sigma(1-y)^{-1/\alpha}dy = 1-(1-p)^{1-1/\alpha}, \; 0 < p < 1 \tag{9.11}$$

provided $\alpha > 1$.

The Singh-Maddala distribution is one of the most popular distributions used in practice to fit income and wealth data (Kleiber and Kotz, 2003). This distribution was obtained by Singh and Maddala (1976) by considering the hazard rate of income. Let X be a random variable with Singh-Maddala distribution with cdf,

$$F_X(x) = 1 - \frac{1}{[1+(x/\sigma)^a]^q}, \; x > 0 \tag{9.12}$$

where $a, q, \sigma > 0$. Definition (9.12) corresponds to the Pareto IV distribution, in the Arnold (1983) Pareto hierarchy. If $q > 1/a$ then using expression (9.6) the Lorenz curve of (9.12) is,

$$\begin{aligned} L_X(p) &= \frac{1}{\mu_X}\int_0^p \sigma[(1-y)^{-1/q}-1]^{1/a}dy \\ &= \frac{\sigma q}{\mu_X}\int_0^z t^{1/a}(1-t)^{q-1/a-1}dt \\ &= I_z(1+1/a, q-1/a), \; 0 \leq p \leq 1 \end{aligned}$$

where $z = 1-(1-p)^{1/q}$ and $I_x(a,b)$ denotes the incomplete beta function ratio defined as $(0 < x < 1)$

$$I_x(a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1}dt}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}. \tag{9.13}$$

Another important income distribution is the Dagum type I distribution (Dagum, 1977) with cdf

$$F(x) = [1+(x/\sigma)^{-a}]^{-q}, \ \ x>0 \tag{9.14}$$

where $a, q, \sigma > 0$. This distribution is related with the Singh-Maddala distribution by the inverse transformation $1/X$. Since the quantile function is available in closed form, the LC can be written as (Dagum, 1977),

$$L(p) = I_z(q+1/a, 1-1/a), \ \ 0 \le p \le 1 \tag{9.15}$$

where $z = p^{1/q}$, $a > 1$ and $I_x(a,b)$ is defined in (9.13). The Gini index is given by,

$$G = \frac{\Gamma(q)\Gamma(2q+1/a)}{\Gamma(2q)\Gamma(q+1/a)} - 1.$$

Another group of income distributions corresponds to families where all of the kth moment distributions and the original distribution belong to the same family so that formulas (9.3) or (9.4) can be applied. Consider a lognormal distribution, for which the cumulative distribution function is given by

$$F(x) = \Phi(\frac{\log x - \mu}{\sigma}), \ \ x>0 \tag{9.16}$$

where Φ denotes the cdf of the standard normal distribution. This distribution will be denoted by $X \sim \mathscr{LN}(\mu, \sigma^2)$. The inverse of the cdf is $F^{-1}(x) = \exp[\mu + \sigma\Phi^{-1}(x)]$ and the cdf of the kth moment distributions is again lognormal and is given by Aitchison and Brown (1957),

$$X_{(k)} \sim \mathscr{LN}(\mu + k\sigma^2, \sigma^2), \ \ k = 1, 2, \ldots \tag{9.17}$$

In particular $X_{(1)} \sim \mathscr{LN}(\mu + \sigma^2, \sigma^2)$. Now, by introducing F^{-1} and $F_{(1)}(x)$ in formula (9.4) we obtain the LC corresponding to the lognormal distribution which is given by

$$L(p) = \Phi(\Phi^{-1}(p) - \sigma), \ \ 0 < p < 1. \tag{9.18}$$

The gamma distribution is another popular distribution used in analysis of income and wealth data. The pdf of a gamma distribution is

$$f(x) = \frac{x^{\alpha-1}e^{-x/\sigma}}{\sigma^\alpha \Gamma(\alpha)}, \ \ x>0 \tag{9.19}$$

where $\sigma > 0$ is a scale and $\alpha > 0$ a shape parameter. A random variable with pdf (9.19) will be denoted as $X \sim \mathscr{G}(\alpha, \sigma)$. The gamma distribution includes as particular cases the exponential ($\alpha = 1$) and the chi-square distribution ($\alpha = n/2$, $n = 1, 2, \ldots$). The cdf of the gamma distribution can be written as

$$F(x) = \frac{\gamma(\alpha, x/\sigma)}{\Gamma(\alpha)}, \ \ x>0 \tag{9.20}$$

where $\gamma(a,x)$ denotes the incomplete gamma function defined as,

$$\gamma(a,x) = \int_0^x t^{a-1}e^{-t}dt \tag{9.21}$$

with $a,x > 0$. The kth moment distribution is distributed again as a gamma distribution, that is, $X_{(k)} \sim \mathscr{G}(\alpha + k, \sigma)$ and thus the Lorenz curve can be expressed in a parametric fashion using (9.3) and (9.20). We thus have that

$$(p, L(p)) = (\frac{\gamma(\alpha, x/\sigma)}{\Gamma(\alpha)}, \frac{\gamma(\alpha+1, x/\sigma)}{\Gamma(\alpha+1)}), \;\; x > 0. \tag{9.22}$$

Sarabia and Castillo (2005) have obtained expressions for the LC and the Gini index for a general class of max-stable income distributions.

In order to complete this section we include the LC corresponding to a discrete random variable. Let X be a geometric distribution with probability mass function $\Pr(X = k) = pq^{k-1}$, $k = 1,2,\ldots$ with $0 < p < 1$ and $q = 1-p$. Using formula (9.6), the LC is (Gastwirth, 1971)

$$L(u) = 1 - kq^{k-1} + (k-1)q^k + kp[u - (1-q^{k-1})],$$

if $1 - q^{k-1} \le u \le 1 - q^k$, $k = 1,2,\ldots$ The Gini index is given by (Dorfman, 1979): $G = (1-p)/(2-p)$.

Table 9.1 summarized the Lorenz curves and the Gini index of some important income distributions.

Table 9.1: Lorenz curves and Gini indices of Classical Income Distributions.

Distribution	Lorenz Curve	Gini Index
Uniform $\mathscr{U}[a,b]$	$L(p) = \frac{2ap+(b-a)p^2}{a+b}$	$G = \frac{b-a}{3(a+b)}$
Exponential[1]	$L(p) = p + (1+\frac{\mu}{\sigma})^{-1}(1-p)\log(1-p)$	$G = \frac{\sigma}{2(\mu+\sigma)}$
Classical Pareto	$L(p) = 1-(1-p)^{1-1/\alpha}$	$G = \frac{1}{2\alpha-1}$
Singh-Maddala	$L(p) = I_z(1+1/a, q-1/a)$ where $z = 1-(1-p)^{1/q}$	$G = 1 - \frac{\Gamma(q)\Gamma(2q-1/a)}{\Gamma(q-1/a)\Gamma(2q)}$
Dagum	$L(p) = I_z(q+1/a, 1-1/a)$ where $z = p^{1/q}$	$G = \frac{\Gamma(q)\Gamma(2q+1/a)}{\Gamma(2q)\Gamma(q+1/a)} - 1$
Lognormal	$L(p) = \Phi(\Phi^{-1}(p) - \sigma)$	$G = 2\Phi(\frac{\sigma}{\sqrt{2}}) - 1$
Classical Gamma	$(p, L(p)) = (\frac{\gamma(\alpha,x/\sigma)}{\Gamma(\alpha)}, \frac{\gamma(\alpha+1,x/\sigma)}{\Gamma(\alpha+1)})$	$G = \frac{\Gamma(\alpha+1/2)}{\sqrt{\pi}\Gamma(\alpha+1)}$

[1]Exponential distribution with cdf $F(x) = 1 - e^{-(x-\mu)/\sigma}$ if $x > \mu$, with $\mu, \sigma > 0$.

4 Models of Parametric Lorenz Curves

There exists a variety of approaches for the construction of parametric families of LC's. The first obvious approach consists of starting from an appropriate parametric family of income distribution functions and obtaining the corresponding LC by analytically using representations (9.4) or (9.6), as we have seen in the previous Section. A second approach consists of selecting parametric families of simple curves satisfying the required conditions for Lorenz curves given in Theorem 9.1. This method usually leads to complicated distribution functions, but may be flexible enough for fitting empirical Lorenz curves.

Several parametric models have been proposed in using the second approach. The pioneer model was established by Kakwani and Podder (1973), who proposed the functional form,

$$L(p) = p^{\alpha} e^{-\beta(1-p)}, \ \ 0 \le p \le 1, \tag{9.23}$$

with $\beta > 0$ and $\alpha \ge 1$ (see also Rao and Tam (1987)). An alternative parameterization of this model was provided by Gupta (1984). Kakwani and Podder (1976) also proposed a new parametric model based on a geometric motivation. This model expresses a point of the LC as (x, y), where y is the length of the ordinate from LC on the egalitarian line and x is the distance of the ordinate from the origin along the egalitarian line. This model was completed by Rasche *et al.* (1980) who proposed the family of curves

$$L(p) = [1-(1-p)^{\alpha}]^{\beta}, \ \ 0 \le p \le 1 \tag{9.24}$$

where $0 < \alpha \le 1$ and $\beta \ge 1$. If $\beta = 1$ we obtain the LC (9.11) corresponding to the classical Pareto distribution, and if $\alpha = 1/\beta$ a symmetric LC is obtained according to definition (9.9).

Using several well-known sets of data Villaseñor and Arnold (1989) observed that segments of ellipses frequently fit data surprisingly well. The class of elliptical LC is given by

$$L(p;\alpha,\beta,\delta) = \frac{1}{2}\left[(a-\beta p) - \sqrt{a^2+bp+cp^2}\right] \tag{9.25}$$

where $a = \alpha+\beta+\delta+1 > 0$, $b = -2a\beta - 4\delta$, $c = \beta^2 - 4\alpha$, $\alpha+\delta \le 1$, and $\delta \ge 0$. Equation (9.25) implies that any point (p_i, q_i) must satisfy $y_i = \alpha x_i + \beta z_i + \delta w_i$, $i = 1, 2, \dots, n$, where $y_i = q_i(1-q_i)$, $x_i = p_i^2 - q_i$, $z_i = q_i(p_i - 1)$, and $w_i = p_i - q_i$. This is a linear function of α, β and δ and the least square estimation method can be applied. Using this fact, robust estimation methods have been proposed by Castillo *et al.* (1998). This functional form provides excellent fit and the associated distribution and density functions are available in closed form. In a similar geometric context and from a proposal by Aggarwal (1984) and Aggarwal and Singh (1984), Arnold (1986) considered a hyperbolic functional form for the LC given by

$$L(p;\alpha,\beta) = \frac{p[1+(\alpha-1)p]}{1+(\alpha-1)+\beta(1-p)},\ \ 0 \leq p \leq 1 \tag{9.26}$$

where $\alpha,\beta > 0$ and $\alpha - \beta < 1$. Models (9.25) and (9.26) can be considered to be within the class of general quadratic Lorenz curves (Villaseñor and Arnold, 1989). The circular LC was considered by Ogwang and Rao (1996).

Arnold *et al.* (1987) proposed a class of LC of the form,

$$L(p;\sigma) = F(F^{-1}(p) - \sigma),\ \ \sigma \geq 0, \tag{9.27}$$

where $F(\cdot)$ is any strongly unimodal cdf. For instance, if $F = \Phi$, we obtain the LC (9.18), corresponding to a classical lognormal distribution.

Chotikapanich (1993) proposed the uniparametric model,

$$L(p;k) = \frac{e^{kp}-1}{e^{k}-1},\ \ 0 \leq p \leq 1 \tag{9.28}$$

where $k > 0$ and where the limit case $k \to 0$ corresponds to the egalitarian line. With several data sets the model outperforms those of Kakwani and Podder (1976) and Rasche *et al.* (1980) in terms of the Gini coefficient estimation but is not as good for predicting expenditures shares.

Sarabia (1997) considered an alternative method for the construction of LC specifying an appropriate quantile function, and using it to generate the LC. Using the generalized Tukey's Lambda distribution, this author obtained a family of nested models, which, in the most general case, is

$$L(p) = \pi_1 p + \pi_2 p^{\alpha_1} + (1-\pi_1-\pi_2)[1-(1-p)^{\alpha_2}],\ \ 0 \leq p \leq 1,$$

where $0 \leq \pi_1, \pi_2 \leq 1$, $\alpha_1 \geq 1$ and $0 < \alpha_2 \leq 1$. This model is a mixture of the egalitarian line, the power LC and the classical Pareto LC.

Another important model was considered by Basmann *et al.* (1990), which extend Kakwani and Podder's model (9.23). Ryu and Slottje (1996) introduced two flexible functional form approaches to approximate Lorenz curves, an exponential polynomial and a Bernstein polynomial expansion. Holm (1993) has based his model on the principle of maximum entropy and Sarabia and Pascual (2002) on linear exponential loss functions.

Recent research on the Lorenz curve (Basmann *et al.*, 1990; Ryu and Slottje, 1996; Ogwang and Rao, 2000) has shown that some families of LCs approximate some segments of the income distributions well but not others segments. In the next subsection we propose some general strategies to obtain extensions and generalizations of the basic parametric models.

4.1 A Hierarchical Family

Recently, Sarabia *et al.* (1999) have suggested a general method for obtaining a hierarchical family of LC that unifies and synthesizes some of the previous proposals, as well as providing good fit in the whole the range of the data. If we begin with any Lorenz curve L_0 the following curves are also Lorenz curves that generalize the initial model L_0:

$$L_1(p;\alpha) = p^{\alpha}L_0(p),\ \ \alpha \geq 1 \text{ or } 0 \leq \alpha < 1 \text{ and } L_0'''(p) \geq 0, \tag{9.29}$$
$$L_2(p;\gamma) = [L_0(p)]^{\gamma},\ \ \gamma \geq 1, \tag{9.30}$$
$$L_3(p;\alpha,\gamma) = p^{\alpha}[L_0(p)]^{\gamma},\ \ \alpha \geq 1 \text{ or } 0 \leq \alpha < 1 \text{ and } L_0'''(p) \geq 0, \tag{9.31}$$

An advantage of this method is that Lorenz ordering results are obtained. Equations (9.29) and (9.30) are ordered with respect to their parameters α and γ and a combination of these cases yield ordering results for (9.31).

This method allows for the generation of a hierarchy of Lorenz curves starting from an initial curve L_0. A relevant family is generated from

$$L_0(p) = L_0(p;k) = 1-(1-p)^k,\ \ 0 < k \leq 1,$$

which is the LC (9.11) associated to the classical Pareto distribution. Since $L_0'''(p;k) > 0$ we can apply results in a general way. We can consider the parametric family of Lorenz curves,

$$L_1(p;k,\alpha) = p^{\alpha}[1-(1-p)^k],\ \ \alpha \geq 0 \tag{9.32}$$
$$L_2(p;k,\gamma) = [1-(1-p)^k]^{\gamma},\ \ \gamma \geq 1, \tag{9.33}$$
$$L_3(p;k,\alpha,\gamma) = p^{\alpha}[1-(1-p)^k]^{\gamma},\ \ \alpha \geq 0,\ \gamma \geq 1, \tag{9.34}$$

which is called the Pareto hierarchy of Lorenz curves, since they originate from the Pareto distribution. Family (9.32) coincides with the family proposed by Ortega *et al.* (1991) and (9.33) with the family proposed by Rasche *et al.* (1980). A detailed study of the family (9.34) can be found in Sarabia *et al.* (1999). The method has been used to generate other families of Lorenz curves beginning with different choices for L_0. If we begin with the Chotikapanich LC given in (9.28), we obtain a new family of LC, called the exponential family of LC by Sarabia *et al.* (2001). This approach was also used by Sarabia and Pascual (2002). Table 9.2 summarized the Pareto LC family.

4.2 Mixture Lorenz Curve

A possible solution for obtaining better fit consists in building more complex models combining some of the classical models using convex linear combinations of LCs. The proposals of Sarabia (1997) and Ogwang and Rao (2000) respond to this idea.

Table 9.2: The Pareto Lorenz curve family.

Lorenz Curve	Gini Index
$L_0(p;k) = 1-(1-p)^k$	$G = \frac{1-k}{1+k}$
$L_1(p;k,\alpha) = p^\alpha[1-(1-p)^k]$	$G = 1-2[B(\alpha+1,1)-B(\alpha+1,k+1))]$
$L_2(p;k,\gamma) = [1-(1-p)^k]^\gamma$	$G = 1-\frac{2}{k}[B(1/k,\gamma+1)]$
$L_3(p;k,\alpha,\gamma) = p^\alpha[1-(1-p)^k]^\gamma$	$G = 1-2\sum_{i=0}^{\infty}\frac{\Gamma(i-\gamma)}{\Gamma(i+1)\Gamma(-\gamma)}B(\alpha+1,ki+1)$

In this sense, one of the reasons that can explain the lack of fit in some LC's is the existence of some factor of heterogeneity in the population (for example, age, gender or education), so the LC varies from some individuals to others. If we compose the initial LC with the heterogeneity (described in terms of a known pdf) we obtain a new LC called a mixture LC (Sarabia *et al.*, 2005). If $L(p;\theta)$ denotes a LC, and we assume that θ varies according to an absolutely continuous density function $\pi(\theta)$ with support on a set $\Theta \subset R$, the expression

$$\tilde{L}(p) = \int_\Theta L(p;\theta)\pi(\theta)d\theta$$

defines a genuine LC. Several mixture LC models have been proposed by Sarabia *et al.* (2005). For example, if a power LC is composed with a gamma distribution, we obtain the LC,

$$L(p;\alpha,\sigma) = \frac{p}{(1-\sigma\log p)^\alpha}.$$

5 Inequality Measures Derived from the Lorenz Curve

The two best known measures of inequality which are directly related to the Lorenz curve are the Gini and Pietra indices. Both indices can be viewed as alternative forms of measuring the distance between the Lorenz curve and the egalitarian line. The Gini index is defined as twice the area between the egalitarian line and the Lorenz curve

$$G_X = 2\int_0^1 [p-L_X(p)]dp = 1-2\int_0^1 L_X(p)dp. \tag{9.35}$$

There are several alternative expressions of the Gini index equivalent to (9.35). One of the most important is

$$G_X = 1 - \frac{E(X_{1:2})}{\mu} = 1 - \frac{1}{\mu}\int_0^{\infty}[1-F_X(x)]^2 dx \qquad (9.36)$$

where $X_{1:2}$ is the smaller of a sample of size 2 coming from the cdf F_X. This expression is useful when we have a closed form for the cumulative distribution function (see, for example, Cronin (1979)).

A second important inequality measure is the Pietra index, which is defined as the maximal vertical deviation between the Lorenz curve and the egalitarian line

$$P_X = \max_{0\le p\le 1}\{p - L_X(p)\}. \qquad (9.37)$$

If we assume that F is strictly increasing on its support, the function $p - L_X(p)$ will be differentiable everywhere on $(0,1)$ and its maximum will be reached when $1 - F^{-1}(x)/\mu$ is zero, that is, when $x = F(\mu)$. The value of $p - L_X(p)$ in this point is given by

$$P_X = F(\mu) - \frac{1}{\mu}\int_0^{F(\mu)}[\mu - F^{-1}(y)]dy = \frac{1}{2\mu}\int_0^{\infty}|z-\mu|dF(z),$$

in consequence

$$P_X = \frac{E|X-\mu|}{2\mu},$$

which is an alternative formula for the Pietra index. For the classical Pareto distribution (9.10) the mean is $\mu = \alpha/(\alpha-1)$ if $\alpha > 1$ and $F(\mu) = 1 - (\alpha/(\alpha-1))^{-\alpha}$ and thus the Pietra index is

$$P_X = F(\mu) - L(F(\mu)) = \frac{(\alpha-1)^{\alpha-1}}{\alpha^{\alpha}}.$$

There are several generalizations of the Gini index proposed in the literature. Mehran (1976)considered the general class of linear measures of the form

$$I(w) = \int_0^1 [p - L_X(p)]dw(p), \qquad (9.38)$$

where $w(p)$ is some increasing function which allows value judgments about inequality to be incorporated. Note that $I(w)$ is always compatible with the Lorenz order. If we take $w(p) = 2p$, $0 \le p \le 1$, we obtain the Gini index.

Another important generalization of the Gini index was proposed by Yitzhaki (1983). This author proposed the generalized Gini index defined as

$$G_\nu = 1 - \nu(\nu-1)\int_0^1 (1-p)^{\nu-2}L_X(p)dp, \qquad (9.39)$$

where $\nu > 1$. If $\nu = 2$ we obtain the Gini index. When ν increases, higher weights are attached to small incomes. The limit case when ν goes to infinity depends on the lowest income, expressing the judgement introduced by Rawls, that social welfare depends only on the poorest society member. On the other hand, it can be proved that (Muliere and Scarsini, 1989)

$$G_\nu = 1 - \frac{E(X_{1:\nu})}{\mu_X},$$

which can also be seen as a generalization of (9.36). For the classical Pareto LC (9.11) Yitzhaki's index (9.39) is,

$$G_\nu = \frac{\nu - 1}{\alpha\nu - 1}, \;\; \alpha, \nu > 1.$$

Arnold (1983, p. 109) has proposed next generalization of the Gini index,

$$\tilde{G}_n = 1 - \frac{E(X_{1:n+1})}{E(X_{1:n})}.$$

The Gini index corresponds to the case $n = 1$. The set of all such indices $\{\tilde{G}_n : n = 1, 2, \ldots\}$ determines the parent distribution up to a scale factor.

Another two important inequality measures deserve our attention: the Atkinson (1970) inequality measures and the generalized entropy indices. The Atkinson inequality indices are defined as

$$A(\varepsilon) = 1 - \left[\int_0^\infty (x/\mu)^{1-\varepsilon} dF(x)\right]^{1/(1-\varepsilon)}, \;\; \varepsilon > 0, \tag{9.40}$$

where ε is a parameter that controls the inequality aversion. The limit cases $\varepsilon \to 1$ and $\varepsilon \to \infty$ are

$$\begin{aligned} A(1) &= 1 - \frac{1}{\mu} \exp\left\{\int_0^\infty \log(x) dF(x)\right\}, \\ A(\infty) &= 1 - \frac{F^{-1}(0)}{\mu}. \end{aligned}$$

The generalized entropy indices are

$$G(\theta) = \frac{1}{\theta(\theta - 1)} \int_0^\infty \left[(x/\mu)^\theta - 1\right] dF(x), \;\; \theta \neq 0, 1 \tag{9.41}$$

and

$$\begin{aligned} G(0) &= \int_0^\infty \log(\mu/x) dF(x), \\ G(1) &= \int_0^\infty (x/\mu) \log(x/\mu) dF(x). \end{aligned}$$

These two latter indices are known as the Theil coefficients. Indices (9.40) and (9.41) can be written in terms of the LC using the formulas,

$$A(\varepsilon) = 1 - \left\{ \int_0^1 [L'_X(p)]^{1-\varepsilon} dp \right\}^{1/(1-\varepsilon)}, \ \ \varepsilon > 0 \tag{9.42}$$

$$G(\theta) = \frac{1}{\theta(\theta-1)} \int_0^1 \left\{ [L'_X(p)]^{\theta} - 1 \right\} dp, \ \ \theta \neq 0,1 \tag{9.43}$$

These formulas allow these indices to be obtained directly from the Lorenz curve without the necessity of knowing the underlying cumulative distribution function. For the classical Pareto distribution with LC (9.11), using (9.43) the generalized entropy index is given by ($\theta \neq 0,1$),

$$G(\theta) = \frac{1}{\theta(\theta-1)} \left[\left(1 - \frac{1}{\alpha}\right)^{\theta} \frac{\alpha}{\alpha - \theta} - 1 \right]$$

where $\alpha > \max\{1, \theta\}$.

6 Lorenz Order

In this section we study the Lorenz ordering and its applications to the most important income distributions, including the members of the family proposed by McDonald (1984). Lorenz curves can be used to define an ordering in the space of the $\mathscr{L}$ distributions. If two distribution functions have associated Lorenz curves which do not intersect, they can be ordered without ambiguity in terms of welfare functions which are symmetric, increasing and quasiconcave (Atkinson, 1970); (Dasgupta *et al.*, 1973; Shorrocks, 1983).

Definition 9.2. Let X and Y be random variables belonging to $\mathscr{L}$ class. The Lorenz order $\leq_L$ on $\mathscr{L}$ is defined by,

$$X \leq_L Y \iff L_X(p) \geq L_Y(p), \ \ \forall p \in [0,1]. \tag{9.44}$$

If $X \leq_L Y$, then X exhibits less inequality than Y in the Lorenz sense. Note that the Lorenz order is a partial order and is invariant with respect to scale transformation. We present two relevant examples of the Lorenz order:

- Let $X_i \sim \mathscr{P}a(\alpha_i, \sigma_i)$, $i = 1,2$ be Pareto distributions with cdf (9.10). Then:

$$X_1 \leq_L X_2 \iff \alpha_1 \geq \alpha_2.$$

- Let $X_i \sim \mathscr{L}\mathscr{N}(\mu_i, \sigma_i)$, $i = 1,2$ be lognormal distributions with cdf (9.16). Then:

$$X_1 \leq_L X_2 \iff \sigma_1 \leq \sigma_2.$$

The proof of these results is direct by checking the Lorenz curve. Other stronger definitions of stochastic orderings are useful in this context. Let X and Y be random variables in $\mathscr{L}$ with distribution functions F_X and F_Y. Star-shaped ordering is defined as follows (Arnold, 1987).

Definition 9.3. We say that X is star-shaped with respect to Y, and we write $X \leq_* Y$ if $F_X^{-1}(x)/F_Y^{-1}(x)$ is a non-increasing function of x.

This definition is specially useful when the quantile function is available in a closed form. The star-shaped ordering implies the Lorenz ordering.

Theorem 9.3. *Suppose that $X,Y \in \mathscr{L}$. If $X \leq_* Y$, then $X \leq_L Y$.*

The proof of this result is as follows. Without loss of generality we may assume that $E(X) = E(Y) = 1$, since both orders are scale invariant. Then,

$$L_X(p) - L_Y(p) = \int_0^p [F_X^{-1}(y) - F_Y^{-1}(y)]dy.$$

Now, since $F_X^{-1}(y)/F_Y^{-1}(y)$ is a non-increasing function, the integrand is first positive and then negative as y ranges from 0 to 1. In consequence the integral assumes its smallest value when $p = 1$. Thus, $L_X(p) - L_Y(p) \geq L_X(1) - L_Y(1) = 1 - 1 = 0$, and $X \leq_L Y$.

This result was used by Wilfling (1996) for proving the Lorenz ordering in the Singh-Maddala family (see below).

The next theorem established by Fellman (1976), examines the Lorenz order between a random variable X and a transformation $g(X)$.

Theorem 9.4. *Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be a continuous function satisfying*

1. *$g(x) > 0$ for all $x > 0$,*
2. *$g(x)$ is non-decreasing on $[0,\infty)$ and $g(x)/x$ is non-decreasing on $(0,\infty)$.*

If $g(X) \in \mathscr{L}$ then $g(X) \leq_L X$.

Let us now focus our attention on three important income distributions proposed in the literature. The generalized gamma (GG) and generalized beta of the first and second kind (GB1 and GB2) (see McDonald (1984)) are defined in terms of their probability density functions ($a,p,q,\sigma > 0$):

$$f_{GG}(x;a,p,\sigma) = \frac{ax^{ap-1}e^{-(x/\sigma)^a}}{\sigma^{ap}\Gamma(p)}, \quad x \geq 0, \tag{9.45}$$

$$f_{GB1}(x;a,p,q,\sigma) = \frac{ax^{ap-1}[1-(x/\sigma)^a]^{q-1}}{\sigma^{ap}B(p,q)}, \quad 0 \leq x \leq b \tag{9.46}$$

$$f_{GB2}(x;a,p,q,\sigma) = \frac{ax^{ap-1}}{\sigma^{ap}B(p,q)[1+(x/\sigma)^a]^{p+q}}, \quad x \geq 0 \tag{9.47}$$

and 0 otherwise. The parameter σ in (9.45), (9.46) and (9.47) is a scale parameter and, due to the fact that the Lorenz ordering is invariant with respect to scale changes, it can be assumed without loss of generality that it is equal to 1. Thus we will represent them as $X \sim GG(a,p)$, $X \sim GB1(a,p,q)$ and $X \sim GB2(a,p,q)$.

These models include an important number of income distributions proposed in the literature. The generalized gamma includes the usual gamma distribution ($GG(1,p) \equiv G(p)$), the Weibull distribution ($GG(a,1) \equiv W(a)$) and the exponential distribution ($GG(1,1) \equiv E(1)$). The GB2 includes the usual beta distribution of the second kind ($GB2(1,p,q) \equiv B2(p,q)$), the Singh-Maddala distribution ($GB2(a,1,q) \equiv SM(a,q)$), the Dagum (1977) distribution ($GB2(a,p,1) \equiv D(a,p)$), the Lomax distribution ($GB2(1,1,q) \equiv L(q)$) and the Fisk distribution ($GB2(a,1,1) \equiv F(a)$). Both of the generalized beta distributions include the generalized gamma as a limiting case.

The next result provides the Lorenz order within the family of generalized gamma distributions defined in (9.45) (Taillie, 1981; Wilfling, 1996).

Theorem 9.5. *Let $X_i \sim GG(a_i, p_i)$, $i = 1,2$ be generalized gamma distributions. Then,*

$$X_1 \geq_L X_2 \iff a_1 \leq a_2 \text{ and } a_1 p_1 \leq a_2 p_2.$$

For the GB2 family, the Lorenz ordering can be verified for certain parametric configurations (Kleiber, 1999).

Theorem 9.6. *Let $X_i \sim GB2(a_i, p_i, q_i)$, $i = 1,2$ be GB2 distributions with finite means. Then*

1. *If $a_1 \leq a_2$, $a_1 p_1 \leq a_2 p_2$ and $a_1 q_1 \leq a_2 q_2$ then $X_1 \geq_L X_2$.*
2. *If $X_1 \geq_L X_2$ then $a_1 p_1 \leq a_2 p_2$ and $a_1 q_1 \leq a_2 q_2$.*

This theorem leaves open some parameter configurations of the kind $a_1 \leq a_2$, $p_1 \geq p_2$ and $q_1 \geq q_2$, with $a_1 p_1 \geq a_2 p_2$ and $a_1 q_1 \geq a_2 q_2$. In spite of these holes, this result allows a complete characterization of many subfamilies coming from GB2 distribution. Some important cases are the following:

- Let $X_i \sim SM(a_i, q_i)$, $i = 1,2$ be Singh-Maddala distributions with cdf given in (9.12). Then (Wilfling and Krämer, 1993; Wilfling, 1996):

$$X_1 \geq_L X_2 \iff a_1 q_1 \leq a_2 q_2, \text{ and } a_1 \leq a_2.$$

- Let $X_i \sim B2(p_i, q_i, \sigma_i)$, $i = 1,2$ be beta distributions of the second kind. Then:

$$X_1 \geq_L X_2 \iff p_1 \leq p_2, \text{ and } q_1 \leq q_2.$$

- Let $X_i \sim D(a_i, q_i)$, $i = 1,2$ be Dagum distributions with cdf (9.14). Then (Kleiber, 1996, 1999)

$$X_1 \geq_L X_2 \iff a_1 q_1 \leq a_2 q_2, \text{ and } a_1 \leq a_2.$$

The following results (Sarabia *et al.*, 2002) establish some additional Lorenz orderings involving the three families of distributions (9.45)-(9.47).

Theorem 9.7. *Assume that one of the following conditions holds:*

1. *Let $X \sim GG(\tilde{a}, \tilde{p})$ and $Y \sim GB2(a,p,q)$, with $aq > 1$, $\tilde{a} \geq a$ and $\tilde{a}\tilde{p} \geq a$.*

2. *Let $X \sim GB1(a,p,q)$ and $Y \sim GB2(\tilde{a},\tilde{p},\tilde{q})$, with $\tilde{a}\tilde{q} > 1$, $a \geq \tilde{a}$, $ap \geq \tilde{a}\tilde{p}$ and $aq \geq \tilde{a}\tilde{q}$.*
3. *Let $X \sim GB1(a,p,q)$ and $Y \sim GG(\tilde{a},\tilde{p})$, with $a \geq \tilde{a}$, $ap \geq \tilde{a}\tilde{p}$.*

Then: $X \leq_L Y$

A whole range of literature is available for studying sampling theory of Lorenz curves (Beach and Davidson (1983) and Bishop *et al.* (1989) among others). The problem of making inequality comparison when Lorenz curves intersect has been studied by Shorrocks and Foster (1987) and Davies and Hoy (1995).

7 Variations of the Lorenz Curve

The generalized Lorenz curve (GLC) introduced by Shorrocks (1983) is the most important variation of the LC. The LC is scale invariant and is thus only an indicator of relative inequality. However, it does not provide a complete basis for making social welfare comparisons. The Shorrocks proposal is the generalized Lorenz curve defined as

$$GL_X(p) = \mu_X \cdot L_X(p) = \int_0^p F_X^{-1}(y)dy, \ \ 0 \leq p \leq 1. \tag{9.48}$$

Note that $GL_X(0) = 0$ and $GL_X(1) = \mu_X$. A distribution with a dominating GLC provides greater welfare according to all concave increasing social welfare functions defined on individual incomes (Kakwani (1984) and Davies *et al.* (1998)). On the other hand, the GLC is no longer scale-free and in consequence it determines any distribution with finite mean. The order induced by (9.48) is the second-order stochastic dominance

$$X_1 \leq_{GL} X_2 \iff \int_0^x F_1(y)dy \leq \int_0^x F_2(y)dy, \ \ x \geq 0,$$

which has been studied by Thistle (1989). This order is a new partial ordering, and sometimes it allows a bigger percentage of curves to be ordered than in the Lorenz ordering case. The normative interpretations for the restrictions required on the class of social welfare function to satisfy a GLC dominance have been studied by Shorrocks and Foster (1987) and Davies and Hoy (1994) among others.

Other variations of the LC have been proposed. The absolute Lorenz curve introduced by Moyes (1987) is defined by,

$$AL_X(p) = \mu_X \cdot [L_X(p) - p] = \int_0^p [F_X^{-1}(u) - \mu_X]du, \ \ 0 < p < 1.$$

Note that the new definition changes scale invariance with location invariance. Zenga (1984) defined next concentration curve,

$$ZC(p) = 1 - \frac{F^{-1}(p)}{F_{(1)}^{-1}(p)}, \ \ 0 < p < 1,$$

which is scale free and belongs to the unit square.

8 Multivariate Lorenz Curves

We finish this chapter about Lorenz curves with their extensions to higher dimensions. Although the use of multivariate income data is becoming increasingly more habitual, the proposals of multivariate Lorenz curves are very recent. The pioneer work in this field is due to Taguchi (1972a,b) and Arnold (1987). A recent multivariate version of the LC is based on the concept of Lorenz zonoid of the population introduced by Koshevoy (1995) and Koshevoy and Mosler (1996). Their idea is based on a vision of the usual LC as a convex region bordered by $L(p)$ and $\tilde{L}(p)$, where $\tilde{L}(p) = 1 - L(1-p)$ is the dual Lorenz curve. With this idea, the area between these two curves is the classical Gini index.

The multivariate Lorenz curve is a generalization of this concept to $d+1$ space. Consider the set $\mathscr{L}^d$ of probability distribution functions on R_+^d that have finite and strictly positive expectations $\mu_j = \int_{\mathbb{R}_+} x_j dF(x)$, $j = 1,2,\dots,d$ and set
$\tilde{\mathbf{x}} = (\tilde{x_1},\dots,\tilde{x_j})^\top$, $\tilde{x_j} = \frac{x_j}{\mu_j}$, $j = 1,2,\dots,d$.

Then, $\tilde{\mathbf{X}}$ is the normalization of $\mathbf{X}$ with expectation $\mathbf{1}_d = (1,\dots,1)^\top$. For $F \in \mathscr{L}^d$, the set

$$LZ(F) = \{\mathbf{z} \in R^{d+1} : \mathbf{z} = (z_0, z_1, \dots, z_d) = \zeta(h)\}$$

where

$$\zeta(h) = \left(\int_{R_+^d} h(\mathbf{x}) dF(\mathbf{x}), \int_{R_+^d} h(\mathbf{x}) \tilde{\mathbf{x}} dF(\mathbf{x})\right)$$

for every measurable function $h : R_+^d \to [0,1]$, is called the Lorenz zonoid. The Lorenz zonoid is a convex compact subset of the unit hypercube in R_+^{d+1} containing the origin and the point $\mathbf{1}_{d+1}$ in R^{d+1}. Now, we define a generalization of the LC. For $F \in \mathscr{L}^d$, let us consider the set

$$Z(F) = \{\mathbf{y} \in R_+^d : \mathbf{y} = \int_{R_+^d} h(\mathbf{x}) \tilde{\mathbf{x}} dF(\mathbf{x}), h : R_+^d \to [0,1], \text{measurable}\},$$

which is called the F zonoid.

Note that if $(z_0, z_1, \dots, z_d) \in LZ(F)$, then $(z_1, \dots, z_d) \in Z(F)$. The F zonoid is contained in the unit cube on R_+^d and consists of all total portion vectors held by subpopulations. If $d = 1$, $Z(F)$ is the unit interval. For a given $(z_1, \dots, z_d) \in Z(F)$, we have $(z_0, z_1, \dots, z_d) \in LZ(F)$ if and only if z_0 is in the closed interval between the smallest and the largest percentage of the population by which the portion vector $(z_1, \dots, z_d)$ is held. This leads us to the definition of an inverse Lorenz function. The function $l_F : Z(F) \to R_+$ defined as

$$l_F(\mathbf{y}) = \max\{t \in R_+ : (t, \mathbf{y}) \in LZ(F)\},$$

is called the inverse Lorenz function of F. Its graph is the Lorenz surface of F. In terms of a distribution of commodities, the function $l_F(\mathbf{y})$ is equal to the maximum percentage of the population whose total portion amounts to $\mathbf{y}$. The multivariate order is defined as the set inclusion ordering of Lorenz zonoids

$$F \geq_{LZ} G \iff LZ(F) \supseteq LZ(G),$$

and implies the usual Lorenz ordering of all marginal distributions. Finally, the multivariate Gini index is defined as the volume of their Lorenz zonoid $LZ(F)$

$$\mathbf{G} = \mathrm{vol}[LZ(F)] = \frac{E(|\det \mathbf{Q}_F|)}{(d+1)!\prod_{j=1}^{d} E(X_j)},$$

where $\mathbf{Q}_F$ is the $(d+1)\times(d+1)$ matrix with rows $(1,\mathbf{X}_i)$, $i = 1,2,\ldots,d+1$, and $\mathbf{X}_1,\ldots,\mathbf{X}_{d+1}$ are i.i.d. with cdf F.

The Lorenz zonoid order and the multivariate Gini index appear to be good choices as suitable d-dimensional analogs of the Lorenz order and the Gini index. However, there are some problems. Sometimes, the zonoid can have zero volume for some non-degenerate distributions. In response to this, Mosler (2002) has provided a modified definition to rectify this problem. Several alternative definitions for a Lorenz order among d-dimensional non-negative random vectors have been proposed by Arnold (2007).

9 Applications of the Lorenz Curves

Application of Lorenz curves and associated concentration measures is encountered in a broad spectrum of modern scientific fields. Many authors in very different areas of investigation have realized the usefulness of these instruments. Atkinson (1970), in his seminal and influential paper showed that the rules for ordering risky prospects can be written in terms of Lorenz curves (Hadar and Russell, 1969; Hanoch and Levy, 1969; Rothschild and Stiglitz, 1970). Perhaps the greatest number of applications can be found in the usual field of income distributions and poverty (Sen, 1976) but also in the field of finance. In this last field, rules for ordering risky prospects using the Gini index and for the evaluation of risky assets have been studied and developed (Yitzhaki, 1982; Shalit and Yitzhaki, 1984).

Other applications include the use of the Lorenz/Leimkuhler concentration curves in informetric contexts (Burrell, 2005), Lorenz curves of cumulative electricity consumption (Jacobson and Kammen, 2005), LC and Gini index to assess yield inequality within paddocks (Sadras and Bongiovanni, 2004) or characterization of the early growth inequality of ninety crosses of Chinese fir (Ma *et al.*, 2006), to mention but a few examples.

Acknowledgements

I would like to thank the Ministerio de Educación y Ciencia (project SEJ2004-02810) for partial support of this work. The author thank the referees and the editor for constructive suggestions that have improved the content and presentation of this chapter.

References

Aggarwal, V. (1984) On Optimal Aggregation of Income Distribution Data, *Sankhya B*, **46**, 343–35.

Aggarwal, V. and R. Singh (1984) On Optimum Stratification with Proportional Allocation for a Class of Pareto Distributions, *Communications in Statistics: Theory and Methods*, **13**, 3107–3116.

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

Arnold, B. C. (1983) *Pareto Distributions*, International Cooperative Publishing House, Fairland, Maryland USA.

Arnold, B. C. (1986) A Class of Hyperbolic Lorenz Curves, *Sankhya B*, **48**, 427–436.

Arnold, B. C. (1987) *Majorization and the Lorenz Curve: A Brief Introduction*, Lecture Notes in Statistics, 43, Springer-Verlag, Berlin.

Arnold, B. C. (2007) *The Lorenz Curve: Evergreen after 100 Years*, Advances in Income Inequality and Concentration Measures, Routledge, New York.

Arnold, B. C., C. A. Robertson, P. L. Brockett and B. Y. Shu (1987) Generating Ordered Families of Lorenz Curves by Strongly Unimodal Distributions, *Journal of Business and Economic Statistics*, **5**, 305–308.

Atkinson, A. B. (1970) On the Measurement of Inequality, *Journal of Economic Theory*, **2**, 244–263.

Basmann, R. L., K. L. Hayes, D. J. Slottje and J. D. Johnson (1990) A General Functional Form for Approximating the Lorenz Curve, *Journal of Econometrics*, **43**, 77–90.

Beach, C. M. and R. Davidson (1983) Distribution-Free Statistical Inference with Lorenz Curves and Income Shares, *Review of Economics and Statistics*, **50**, 723–735.

Bishop, J. A., S. Chakravorty and P. D. Thistle (1989) Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves, *Review of Economics and Statistics*, **71**, 725–727.

Burrell, Q. L. (2005) Symmetry and Other Transformation of Lorenz/Leimkuhler Representations of Informetric Data, *Information Processing and Management*, **41**, 1317–1329.

Castillo, E., A. S. Hadi and J. M. Sarabia (1998) A Method for Estimating Lorenz Curves, *Communications in Statistics: Theory and Methods*, **27**, 2037–2063.

Chotikapanich, D. (1993) A Comparison of Alternative Functional Forms for the Lorenz Curve, *Economic Letters*, **41**, 129–138.

Cronin, D. C. (1979) A Function for the Size Distribution of Income: A Further Comment, *Econometrica*, **47**, 773–774.

Dagum, C. (1977) A New Model for Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, **30**, 413–437.

Dasgupta, P., A. K. Sen and D. Starret (1973) Notes on the Measurement of Inequality, *Journal of Economic Theory*, **6**, 180–187.

Davies, J. and M. Hoy (1994) The Normative Significance of Using Third-Degree Stochastic Dominance in Comparing Income Distributions, *Journal of Economic Theory*, **64**, 520–530.

Davies, J. and M. Hoy (1995) Making Inequality Comparisons when Lorenz Curves Intersect, *American Economic Review*, **85**, 980–986.

Davies, J. B., D. A. Green and H. J. Paarsch (1998) *Economic Statistics and Social Welfare Comparisons. A Review.*, Handbook of Applied Economic Statistics, 1-38, Marcel Dekker, New York.

Dorfman, R. (1979) A Formula for the Gini Coefficient, *Review of Economics and Statistics*, **61**, 146–149.

Fellman, J. (1976) The Effect of Transformations on the Lorenz Curve, *Econometrica*, **44**, 823–824.

Gastwirth, J. L. (1971) A General Definition of the Lorenz Curve, *Econometrica*, **39**, 1037–1039.

Gupta, M. R. (1984) Functional Form for Estimating the Lorenz Curve, *Econometrica*, **52**, 1313–1314.

Hadar, J. and W. R. Russell (1969) Rules for Ordering Uncertain Prospects, *American Economic Review*, **59**, 25–34.

Hanoch, G. and H. Levy (1969) The Efficiency Analysis of Choices Involving Risk, *Review of Economic Studies*, **36**, 335–346.

Holm, J. (1993) Maximum Entropy Lorenz Curves, *Journal of Econometrics*, **44**, 377–389.

Jacobson, A. A. D. M. and D. M. Kammen (2005) Letting the (Energy) Gini out of the Bottle: Lorenz Curves of Cumulative Electricity Consumption and Gini Coefficients as Metrics of Energy Distribution and Equity, *Energy Policy*, **33**, 1825–1832.

Kakwani, N. (1980) On a class of poverty measures, *Econometrica*, **48**, 437–446.

Kakwani, N. and N. Podder (1973) On estimation of lorenz curves from grouped observations, *International Economic Review*, **14**, 278–292.

Kakwani, N. C. (1984) *Welfare Ranking in Income Distribution, Innequality Measurement and Policy*, vol. 3 of *Advances in Econometrics, 191-215*, JAI Press, Gleenwitch, Conn.

Kakwani, T. and N. Podder (1976) Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations, *Econometrica*, **44-1**, 137–149.

Kleiber, C. (1996) Dagum vs. Singh-Maddala Income Distributions, *Economics Letters*, **53**, 265–268.

Kleiber, C. (1999) On the Lorenz Order within Parametric Families of Income Distributions, *Sankhyā B*, **61**, 514–517.

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

Koshevoy, G. (1995) Multivariate Lorenz Majorization, *Social Choice and Welfare*, **12**, 93–102.

Koshevoy, G. and K. Mosler (1996) The Lorenz Zonoid of a Multivariate Distribution, *Journal of American Statistical Association*, **91**, 873–882.

Ma, Z., J. Shi, G. Wang and Z. He (2006) Temporal Changes in the Inequality of Early Growth of Gunninghamia Lanceolata (lamb.) Hook: A Novel Application of the Gini Coefficient and Lorenz Asymmetry., *Genetica*, **126**, 343–663.

McDonald, J. B. (1984) Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52**, 647–663.

Mehran, F. (1976) Linear Measures of Income Inequality, *Econometrica*, **44**, 805–809.

Mosler, K. (2002) *Multivariate Dispersion, Ccentral Regions and Depth. the Lift Zonoid Approach*, Lecture Notes in Statistics, 165, Springer-Verlag, Berlin.

Moyes, P. (1987) A New Concept of Lorenz Domination, *Economic Letters*, **23**, 203–207.

Muliere, P. and M. Scarsini (1989) A Note on Stochastic Dominance and Inequality Measures, *Journal of Economic Theory*, **49**, 314–323.

Ogwang, T. and U. L. G. Rao (1996) A New Functional Form for Approximating the Lorenz Curve, *Economic Letters*, **52**, 21–29.

Ogwang, T. and U. L. G. Rao (2000) Hybrid Models of the Lorenz Curve, *Economic Letters*, **69**, 39–44.

Ortega, P., A. Martín, A. Fernández, M. Ladoux and A. Garcá (1991) A New Functional Form for Estimating Lorenz Curves, *Review of Income and Wealth*, **37**, 447–452.

Pakes, A. G. (1981) On Income Distributions and Their Lorenz Curves, Tech. rep., Department of Mathematics, University of Western Australia.

Rao, U. L. G. and A. Y. P. Tam (1987) An Empirical Study of Selection and Estimation of Alternative Models for the Lorenz Curve, *Journal of Applied Statistics*, **14**, 275–280.

Rasche, R. H., J. Gaffney, A. Y. C. Koo and N. Obst (1980) Functional Forms for Estimating the Lorenz Curve, *Econometrica*, **48**, 1061–1062.

Rothschild, M. and J. E. Stiglitz (1970) Increasing Risk: I. A Definition, *Journal of Economic Theory*, **2**, 225–253.

Ryu, H. K. and D. J. Slottje (1996) Two Flexible Functional Form Approaches for Approximating the Lorenz Curve., *Journal of Econometrics*, **72**, 251–274.

Ryu, H. K. and D. J. Slottje (1999) *Handbook on Income Inequality Measurement*, chap. Parametric Approximations of the Lorenz Curve, pp. 291–314, Kluwer, Boston.

Sadras, V. and R. Bongiovanni (2004) Use of Lorenz Curves and Gini Coefficients to Asses Yield Inequality within Paddocks, *Field Crops Research*, **90**, 303–310.

Sarabia, J. M. (1997) A Hierarchy of Lorenz Curves Based on the Generalized Tukey's Lambda Distribution, *Econometric Reviews*, **16**, 305–320.

Sarabia, J. M. and E. Castillo (2005) About a Class of Max-Stable Families with Applications to Income Distributions, *Metron*, **63**, 505–527.

Sarabia, J. M., E. Castillo, M. Pascual and M. Sarabia (2005) Mixture Lorenz Curves, *Economics Letters*, **89**, 89–94.

Sarabia, J. M., E. Castillo and D. Slottje (2001) An Exponential Family of Lorenz Curves, *Southern Economic Journal*, **67**, 748–756.

Sarabia, J. M., E. Castillo and D. J. Slottje (1999) An Ordered Family of Lorenz Curves, *Journal of Econometrics*, **91**, 43–60.

Sarabia, J. M., E. Castillo and D. J. Slottje (2002) Lorenz Ordering between Mcdonalds Generalized Functions of the Income Size Distribution, *Economics Letters*, **75**, 265–270.

Sarabia, J. M. and M. Pascual (2002) A Class of Lorenz Curves Based on Linear Exponential Loss Functions, *Communications in Statistics: Theory and Methods*, **31**, 925–942.

Sen, A. K. (1976) Poverty: An Ordinal Approach to Measurement, *Econometrica*, **44**, 219–231.

Shalit, H. and S. Yitzhaki (1984) Mean-Gini, Portfolio Theory and the Pricing of Risky Assets, *Journal of Finance*, **39**, 1449–1468.

Shorrocks, A. F. (1983) Ranking Income Distributions, *Economica*, **50**, 3–17.

Shorrocks, A. F. and J. E. Foster (1987) Transfer Sensitive inequality Measures, *Review of Economic Studies*, **54**, 485–497.

Singh, S. K. and G. S. Maddala (1976) A Function for the Size Distribution of Incomes, *Econometrica*, **44**, 963–970.

Slottje, D. J. (1990) Using Grouped Data for Constructing Inequality Indices: Parametric vs. Non-parametric Methods, *Economic Letters*, **32**, 193–197.

Taguchi, T. (1972a) On the Two-Dimensional Concentration Surface and Extensions of Concentration Coefficient and Pareto Distribution to the Two-Dimensional Case-I, *Annals of the Institute of Statistical Mathematics*, **24**, 355–382.

Taguchi, T. (1972b) On the Two-Dimensional Concentration Surface and Extensions of Concentration Coefficient and Pareto Distribution to the Two-Dimensional Case-II, *Annals of the Institute of Statistical Mathematics*, **24**, 599–619.

Taillie, C. (1981) Lorenz ordering within the Generalized Gamma Family of Income Distributions, in C. Taillie, G. P. Patil and B. Balderssari (eds.) *Statistical Distributions in Scientific Work*, vol. 6, pp. 181–192, Reidel, Boston.

Thistle, P. D. (1989) Ranking Distributions with Generalized Lorenz Curves, *Southern Economic Journal*, **56**, 1–12.

Villaseñor, J. A. and B. C. Arnold (1989) Elliptical Lorenz Curves, *Journal of Econometrics*, **40**, 327–338.

Wilfling, B. (1996) Lorenz Ordering of Generalized Beta-II Income Distributions, *Journal of Econometrics*, **71**, 381–388.

Wilfling, B. and W. Krämer (1993) Lorenz Ordering of Singh-Maddala Income Distributions, *Economic Letters*, **43**, 53–57.

Yitzhaki, S. (1982) Stochastic Dominance, Mean Variance and Gini's Mean Difference, *American Economic Review*, **72**, 178–185.

Yitzhaki, S. (1983) On an Extension of the Gini Inequality Index, *International Economic Review*, **24**, 617–628.

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*, **48**, 301–326.

Part III
Current research

Chapter 10

Maximum Entropy Estimation of Income Distributions from Bonferroni Indices†

Hang Keun Ryu‡

Abstract

This paper presents an information efficient technique to determine the functional forms of income distributions subject to the given side conditions such as the Bonferroni index (BI) and the Gini coefficient (GINI). The original GINI is insensitive to the income share changes of the lower income groups and greater weight is attached to those group shares when the BI was defined. To compare the performance of the BI with those of the GINI and the Theil entropy measure (THEIL) the income deciles of 113 countries were introduced using The UNU/WIDER World Income Inequality Database WIID (2005). The information efficient technique provided guidelines on which income inequality measure performs better for certain countries. The BI performed better for the Czech Republic (GINI=0.26) and the U.S.A. (GINI=0.40), but poorly for Brazil (GINI=0.63). The GINI coefficient performed better for Brazil, but not for the Czech Republic and the U.S.A. The BI is a better index in describing the relative income changes of very evenly distributed country like the Czech Republic or for moderately distributed country like the U.S.A. The GINI is good for an extremely uneven society like Brazil. Based on regression R squared values of 113 countries, the THEIL showed ability in describing the income share changes of upper income groups, the GINI was productive in describing the middle income group shares, and the BI was capable in describing the lower income group shares.

---

† The author thanks the two anonymous referees for excellent comments and useful suggestions

‡ Department of Economics, Chung-Ang University, Seoul 156-756, South Korea. Email: hangryu@cau.ac.kr

1 Introduction

Summary measures are used in describing the degree of income inequalities. The original GINI is best known, but other measures such as the Kuznets measure, the THEIL, and Atkinson's measure are used for various purposes. A review of these measures are found and demonstrated by Ryu and Slottje (1998). Public policy makers rely on these measures to enact public policy and to predict the impact of potential policy actions on income distribution as shown by Sawyer (1976), Dagum (1977, 1996), Dagum and Slottje (2000).

To lessen the weakness of the original GINI as insensitive to the income shares of lower income groups, Yitzhaki (1983, 1998) introduced a generalized Gini measure attaching different weight to the lower and upper ends of the distributions. Yitzhaki introduced two numbers for each country to describe income inequalities. One number is given to control the weighting and another for inequalities. A single value of the control parameter can be chosen by the researcher and used for comparisons across countries (or across time). An inequality table can be constructed where inequalities can be tabulated across a range of weight control values and the countries; this is the strength of the contribution by Yitzhaki. The Bonferroni index involves an implicit weighting of population segments and does not avoid the normative restriction.

This paper utilizes the Bonferroni (1930) index. Unlike the GINI, BI is more sensitive at lower levels of income distribution as it gives more weights to transfer among the poor: see Giorgi and Mondani (1995) and Nygard and Sandstorm (1981) for the definition and properties of BI. The original GINI is defined using the area between the 45 degree line and the Lorenz curve, but the BI is defined using the ratios of the areas between the Lorenz curve and horizontal axis to the area between the 45 degree line and the horizontal axis for each income group. Based on this definition, more weight is given to the lower income groups and less weight to the upper income groups. The BI will be very sensitive to the changes in income shares of the lower income groups as later shown. A good review of the Bonferroni curve and the properties of the BI can be found in Pundir *et al.* (2005) and Giorgi and Crescenzi (2001).

The overview of the paper is as follows. In section 2, the definition of the BI is reviewed. In section 3, income distributions are derived from the BI using the maximum entropy method. This method is replicated for the GINI. In section 4, the income deciles of 113 countries are used to compare the performance of the GINI, BI, and THEIL. Summary and concluding remarks are provided in section 5.

2 The Bonferroni Index (BI)

Let X be a non-negative and absolutely continuous random variable. Its cumulative distribution function $F(x) = \int_0^x f(t)dt$ is continuous and differentiable at least twice and its first moment $\mu = \int_0^\infty x f(x)dx$ is finite. The first incomplete moment is

$$F_1(x) = \frac{1}{\mu}\int_0^x t f(t)dt \tag{10.1}$$

The partial mean of the probability distribution is

$$\mu_x = \frac{\int_0^x t f(t)dt}{\int_0^x f(t)dt} = \frac{\mu F_1(x)}{F(x)} \tag{10.2}$$

The Bonferroni curve (BC) is defined in the orthogonal plane $[F(x), B_F(F)]$ within a unit square (Giorgi and Crescenzi, 2001) where

$$B_F[F(x)] = \frac{\mu_x}{\mu} = \frac{F_1(x)}{F(x)} \tag{10.3}$$

For a cumulative distribution function $z = F(x)$ with inverse, the parametric expression of BC is

$$B_F(z) = \frac{F_1[x(z)]}{z} = \frac{\int_0^z x(z')dz'}{z\mu}, \qquad z \in (0,1], \tag{10.4}$$

where $x(z') = \inf\{x : z(x) = F(x) \geq z'\}$.

When $z \to 0$, $B_F(z)$ takes the form 0/0. The BC does not always start from the origin of the orthogonal plane, as it depends on the definition of X.

The Bonferroni index (BI) is defined as

$$\mathrm{BI} = 1 - \int_0^1 B_F(z)dz \tag{10.5}$$

The expressions for Bonferroni curve and Bonferroni index for some common distributions have been derived and are summarized in Pundir *et al.* (2005).

For a probability density function $f(x)$ with the cumulative distribution function $z(x)$, the Lorenz curve is given by:

$$L[z(x)] = \frac{\int_0^x t f(t)dt}{\int_0^\infty t f(t)dt} \tag{10.6}$$

For a cumulative distribution function $z(x)$ with inverse $x(z)$, the Lorenz curve $L(z)$ is

$$L(z) = \frac{\int_0^z x(z')dz'}{\int_0^1 x(z')dz'} \tag{10.7}$$

The BC can be written as a ratio of two statistics.

$$B_F(z) = \frac{L(z)}{z} \tag{10.8}$$

$$\mathrm{BI} = 1 - \int_0^1 \frac{L(z)}{z}dz \tag{10.9}$$

The BI becomes zero when there is an identical level of income with $L(z) = z$ and the BI becomes one when the income level is extremely unequal with $L(z) = 0$ for $z \in [0,1)$ and L(z=1)=1.

Since the GINI is defined as

$$\text{GINI} = 1 - 2\int_0^1 L(z)dz \tag{10.10}$$

The BI puts more weight to the lower income groups and less weight to the upper income groups compared to the GINI.

The BI can be shown to be a special case of the well known Mehran's (1976) index. Mehran considered a linear averaging method and defined the class of linear measures of income inequality by

$$\text{MF} = \int_0^1 [z - L(z)]W(z)dz$$

where $W(z)$ is a score function. Taking $W(z) = 1/z$, MF becomes BI. Similarly, the GINI and the generalized Gini coefficient (Yitzhaki, 1983) is shown as members of this MF class with an appropriate choice of the weighting function $W(z)$ as explained by Lambert (1993). When Bonferroni curves are used to estimate income inequality from the sampled data, they derived the sample estimates of the Bonferroni curve characterized by k ordinates. For a sample of size n, the sample estimate of $k-1$ element vector of the Bonferroni curve has a limiting multivariate normal distribution when multiplied by $\sqrt{n}$.

The BI can also be defined for discrete distributions. Tarsitano (1990) assumed the population to consist of n income receiving units labeled in a non-descending order of income, $x_1, x_2, \ldots, x_n$. Let μ be the mean income, P_i and Q_i be the cumulative population share and the cumulative income share corresponding to the first i income receiving units. Then

$$P_i = \frac{i}{n}; \qquad Q_i = \frac{\sum_{j=1}^{i} x_j}{n\mu} \quad (i = 1, 2, \ldots n) \tag{10.11}$$

The mean of incomes less than or equal to x_i is

$$M_i = \frac{1}{i}\sum_{j=1}^{i} x_j \tag{10.12}$$

Therefore, for a discrete distribution, the BI is

$$\text{BI}_n = \frac{1}{n-1}\sum_{i=1}^{n-1}\left[\frac{\mu - M_i}{\mu}\right] = \frac{1}{n-1}\sum_{i=1}^{n-1}\left[\frac{P_i - Q_i}{P_i}\right] \tag{10.13}$$

The statistical inference for Bonferroni curve is explained in Pundir *et al.* (2005).

3 Maximum Entropy Estimation of Income Distributions

Information is lost because different income distributions can produce the same inequality measure when the degree of income inequality of a society is described with an inequality measure. For example, if Lorenz curves of two countries cross, the GINIs can be the same if the areas under the Lorenz curves are the same. This effect is called as the Lorenz Dominance Effect as shown by Choo and Ryu (1994).

To compare performance of the BI and the Gini, the income distributions are derived from the BI and from the GINI respectively, and the approximated values are compared with the empirical values.

3.1 Derivation of an income distribution from the given BI

A review of the definition of a share function supposes that the observed income of the i^{th} individual is x_i for $i = 1, 2, \cdots, n$, then the share is

$$s_i = \frac{x_i}{\sum_{i=1}^{n} x_i} \tag{10.14}$$

The share function is interpreted as a probability density function because the share s_i is the probability associated with the probability density that each dollar of total measured income will end up with the i^{th} person. Since each individual has different attributes and a different location position, some will collect more money than others. Each dollar bill will end up in the hand of the i^{th} individual with probability s_i so that the continuous share function $s(z)$ can be considered as a probability density function. This share function is assumed as strictly positive for $z \in [0, 1]$. The coordinate z is the population income coordinate with $z = 0.01$ for the lowest 1% group and $z = 1$ for the highest 1% group.

Once the share function is considered as a density function, the maximum entropy method is applied to determine the functional form of the share function. What follows the share function is derived in the most conservative way by maximizing entropy subject to given conditions. To understand the notion of entropy, Jaynes (1979) noted that it is possible to define a kind of measure on the space of probability distributions such that distributions of higher entropy represent greater 'disorder', or are 'smoother', or are 'more probable', or 'assume less'. Shannon (1948) defines entropy as

$$W \equiv -\int f(x) \log f(x)\, dx$$

Zellner and Highfield (1988) and Ryu (1993) solve the following problem.

$$Max_f\, W \equiv -\int f(x) \log f(x)\, dx \tag{10.15}$$

satisfying

$$\int P_m(x) f(x) dx = \mu_m, \tag{10.16}$$

for $m = 0, 1, \cdots, N$ with the μ_m having known values and $P_m(x)$ is any known function. Since entropy is a measure of a lack of information, the distribution of maximum entropy (ME) is the least informative and most conservative distribution, while distributions of lower entropy are more informative. The given side conditions allows for the determining of the functional forms of a density in a most conservative way by choosing a ME density function. Through maximizing entropy a probability density function is as flat as possible. If no information is given, the ME probability density function is flat, and if the first and second moments are given, then the ME probability density function becomes a normal function.

It is shown that knowledge of the BI is equivalent to knowledge of

$$\int_0^1 \log\, z\, dL \tag{10.17}$$

because

$$\int_0^1 \log\, z\, dL = [(\log\, z)\, L]_{z=0}^{z=1} - \int_0^1 \frac{L(z)}{z} dz \tag{10.18}$$

The first RHS term becomes zero using the L'Hopital's rule

$$\begin{aligned} \lim_{z \to 0} \frac{\log\, z}{\frac{1}{L(z)}} &= -\lim_{z \to 0} \frac{1/z}{\frac{s(z)}{L^2(z)}} = -\lim_{z \to 0} \frac{L^2(z)}{z\, s(z)} \\ &= -\lim_{z \to 0} \frac{2L(z)\, s(z)}{s(z) + z\, s'(z)} = -\lim_{z \to 0} \frac{2s^2(z) + 2L(z)\, s'(z)}{2s'(z) + z\, s''(z)} = 0 \end{aligned} \tag{10.19}$$

when $s(0) = 0, s'(0) \neq 0, s''(0) \neq 0, L(0) = 0.$
Since

$$dL(z) = s(z) dz \tag{10.20}$$

(10.17) can be rewritten

$$\int_0^1 \log\, z\, dL = \int_0^1 \log\, z\, s(z) dz = -\int_0^1 \frac{L(z)}{z} dz = \mathrm{BI} - 1 \tag{10.21}$$

Solving an entropy maximization problem as stated in Ryu (1993) has

$$Max_s W \equiv -\int s(z) \log s(z) dz \tag{10.22}$$

satisfying,

$$\int_0^1 s(z) dz = 1 \qquad \text{and} \qquad \int_0^1 \log\, z s(z) dz = \mathrm{BI} - 1 \tag{10.23}$$

The Lagrangian method

$$s(z) = \exp\left[a + b\ \log z\right]\ = A\, z^b\ = (1+b)\, z^b \tag{10.24}$$

The constant A is removed using the normalization condition of the share function. Applying the integral formula of Dwight (1961),

$$\int_0^1 \log z\, s(z)dz = (1+b)\int_0^1 (\log z)\, z^b\, dz = -\frac{1}{(1+b)} = \text{ BI} - 1 \tag{10.25}$$

If BI=0 with identical level of income, $b = 0$ produces a flat share function. If BI=1 with completely uneven income distribution, b becomes a large number. The share function will sharply increase. In the following section income distributions are derived from a GINI value rather than the BI value. More details of this method are found in Ryu and Slottje (2003).

3.2 Derivation of an income distribution from the GINI

The definition of the Lorenz curve shows

$$g \equiv \int_0^1 L(z)dz = \frac{1 - GINI}{2} \tag{10.26}$$

It means $g = 0.5$ if GINI=0 and $g = 0$ if GINI=1. Now recalling the definition of the Lorenz curve,

$$L \equiv \int_0^z s(z')dz'$$

Consider the partial integration of

$$\int_0^1 zdL = zL(z)\Big|_0^1 - \int_0^1 L(z)dz = 1 - g \tag{10.27}$$

Since

$$dL(z) = s(z)dz$$

the mean of the share function is

$$\mu_1 = \int_0^1 zs(z)dz = 1 - g = \frac{1 + GINI}{2} \tag{10.28}$$

Knowledge of the GINI is equivalent to knowledge of the first moment of the true share function. This result is also reported in Yitzhaki (1998). Solving an entropy maximization problem as stated in Ryu (1993) has

$$Max_s W \equiv -\int s(z)\log s(z)dz \tag{10.29}$$

satisfying

$$\int zs(z)dz = \mu_1, \tag{10.30}$$

The Lagrangian method

$$s(z) = \exp[a+bz] = \left[\frac{b}{e^b-1}\right]\cdot\exp[bz] \tag{10.31}$$

where the normalization condition of the share function is used to remove a. Now the first moment condition of (10.26) produces,

$$\mu_1 = \left[\frac{b}{e^b-1}\right]\int_0^1 z\exp[bz]dz = \frac{1+GINI}{2} \tag{10.32}$$

Since the integration is a function of b, $h(b)$ is used to label μ_1.

$$h(b) \equiv -\frac{1}{b} + \frac{e^b}{e^b-1} = \frac{1+GINI}{2} \tag{10.33}$$

If the GINI=0, then b approaches zero while if GINI=1, b approaches infinity.

4 Applications

In this section performance of the GINI, BI, and the THEIL are compared. Since entropy describes the degree of concentration, Theil (1967) defined the entropy measure of income distribution is as follows

$$\text{THEIL} = -\sum_{i=1}^{n} s_i \log s_i \tag{10.34}$$

where s_i is income share of the i^{th}person.

The World Income Inequality Database V2.0a June 2005 is applied to compare the performance of BI, GINI, and THEIL. The UNU/WIDER World Income Inequality Database (WIID) collects and stores information on income inequality for developed, developing, and transition countries (database and documentation are available on the website). For the income deciles of 113 countries, the performance of each inequality measures is evaluated and compared for ten income share groups.

Figure 1 shows the scatter plot of the BI and THEIL values against the GINI value. Since the plots of BI and THEIL are not monotonic functions; it is possible that when the GINI value increases, the BI and THEIL values decrease. This means the order of inequality can be reversed depending on the choice of inequality measures.

Fig. 1 Compariosn of Gini, BI, and THEIL

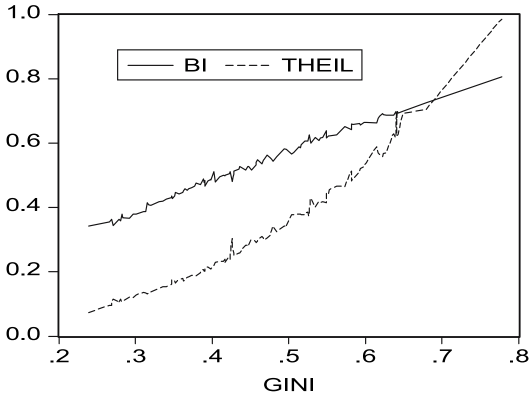
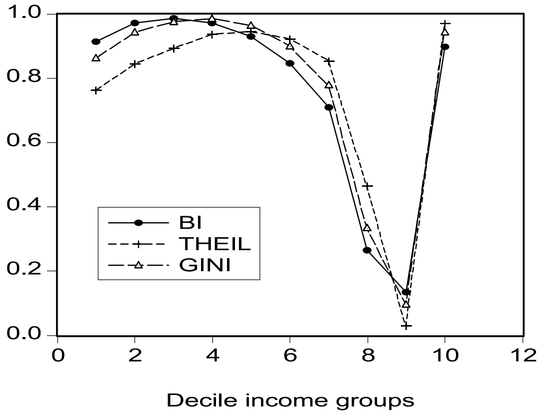


Figure 2 reports the results of regression analysis. For 113 countries, income shares of the lowest income group denoted by s(1) can be regressed against the BI values. If the lowest income group income share s(1) is very small, the corresponding BI value should be large to show an extreme inequality. If income share s(1) is not very small, then the corresponding BI value should not be large enough to reflect a certain degree of income equality. If the income share of the richest group

Fig.2 Regression R squared values



denoted by s(10) is very large, then the corresponding BI value should be large to reflect extreme inequality. Similar logic applies for the GINI and THEIL values.

$$\begin{aligned} s(1)_i &= a(1)_i + b(1)\,\mathrm{BI}_i + u(1)_i \\ &\vdots \qquad\quad \vdots \\ s(10)_i &= a(10)_i + b(10)\,\mathrm{BI}_i + u(10)_i \end{aligned} \tag{10.35}$$

$$\begin{aligned} s(1)_i &= A(1)_i + B(1)\,\mathrm{THEIL}_i + v(1)_i \\ &\vdots \qquad\quad \vdots \\ s(10)_i &= A(10)_i + B(10)\mathrm{THEIL}_i + v(10)_i \end{aligned} \tag{10.36}$$

$$\begin{aligned} s(1)_i &= \alpha(1)_i + \beta(1)\,\mathrm{GINI}_i + e(1)_i \\ &\vdots \qquad\quad \vdots \\ s(10)_i &= \alpha(10)_i + \beta(10)\,\mathrm{GINI}_i + e(10)_i \end{aligned} \tag{10.37}$$

Figure 2 shows the R^2 values produced by each regression. Using the BI values as explanatory variables produced higher R^2 values for the lower income groups compared to those values produced by the GINI and THEIL values. This explains the details of income inequality of lower income groups are better shown with the BI values. Similarly, the THEIL values explain the changes of the upper income groups. The GINI values describe the changes of middle income group shares. All three measures are quite insensitive to changes of the eighth and ninth income group shares. These groups correspond to pivot points of a seesaw and the shares are indifferent whether they belong to an evenly distributed society or to an unevenly distributed society. The seesaw balances are parallel to the surface and fixed to a big angle even though the pivot points do not move.

All 113 countries can be sorted in the order of increasing BI values. Figure 3 shows twenty countries with large BI values. The income distributions are unequal. The GINI and BI values are near 0.6 and 0.7 for those countries. The THEIL values fluctuate widely and increase rapidly for the last country. The description of inequality in a country depends on the choice of inequality measure. The THEIL value approaches infinity for an extremely unequal distribution. One to one comparison with the BI and GINI values are difficult as the BI and Gini coefficients are bound within [0,1].

Figure 4 shows similar comparison to those of Figure 3 for the 20 countries with small BI values. Both the GINI and the THEIL values are not monotonic increasing functions. Though estimated inequality measures are plotted on Figure 3 and Figure 4, they do not mean the inequality measure is better or worse than others.

Figure 5 shows scatter plot of the BI, GINI, and THEIL values against the income shares of the poorest deciles of 113 countries. As the first deciles values increase inequality measures decrease. Figures 6-14 show similar scatter plots of the BI, GINI, and THEIL values against the second, third, and tenth deciles of 113 countries. For the lower 60 percent, up to sixth deciles, the inequality measures decrease as the

Fig.3 Comparison of various inequality measures
for the most unevenly distributed 20 countries

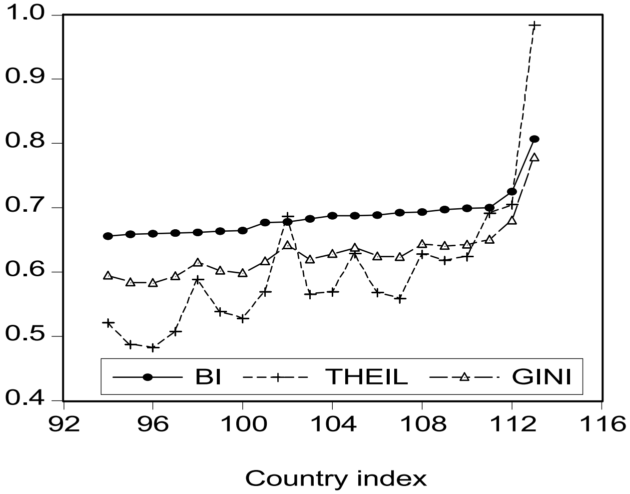


Fig.4 Comparison of various ineuqality measures
for the most evenly distributed 20 countries

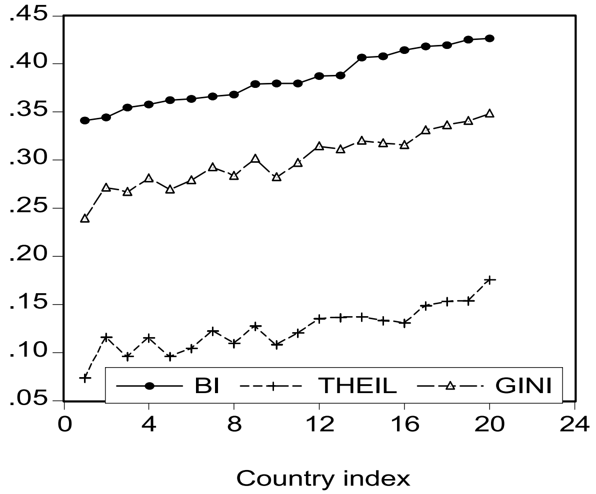


Fig.5 First group income shares and various inequality measures

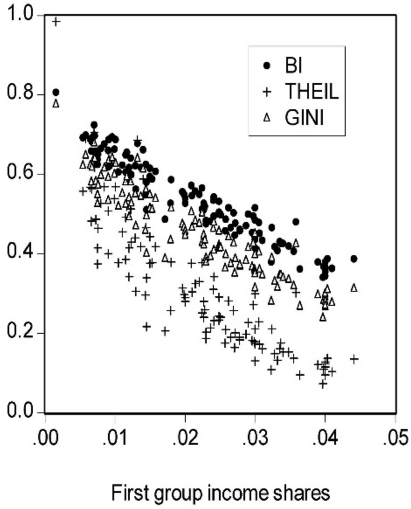


Fig. 6 Second group income shares and various ineuqality measures

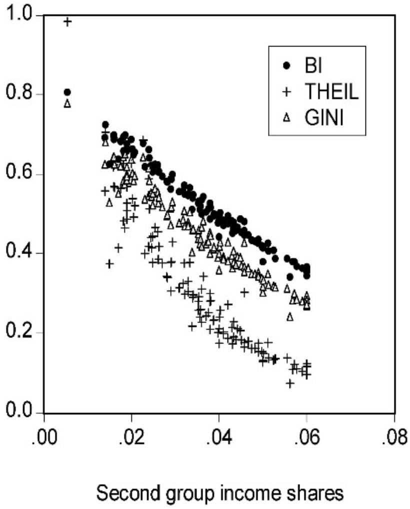


Fig.7 Third group income shares and various inequality measures

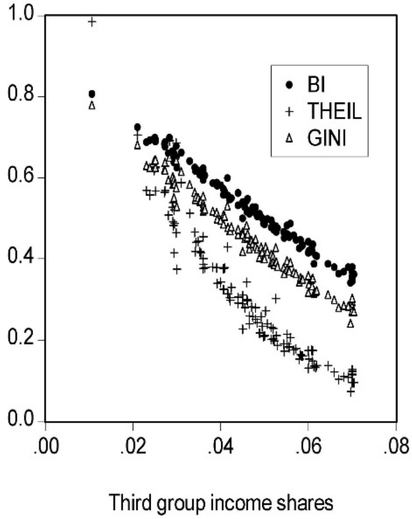


Fig.8 Fourth group income shares and various inequality measures

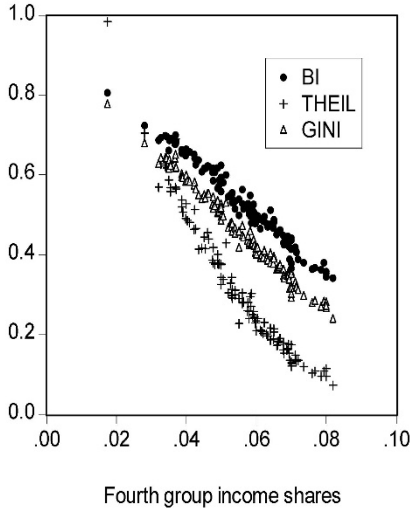


Fig.9 Fifth group income shares and various inequality measures

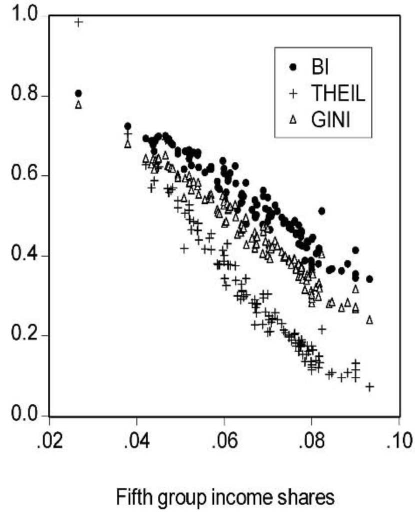


Fig.10 Sixth group income shares and various inequality measures

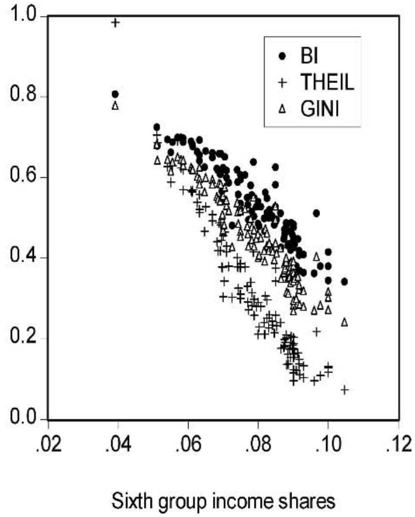


Fig.11 Seventh group income shares and various inequality measures

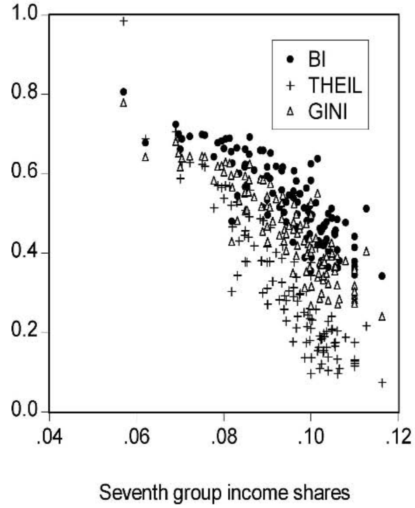


Fig.12 Eighth group incomes shares and various inequality measures

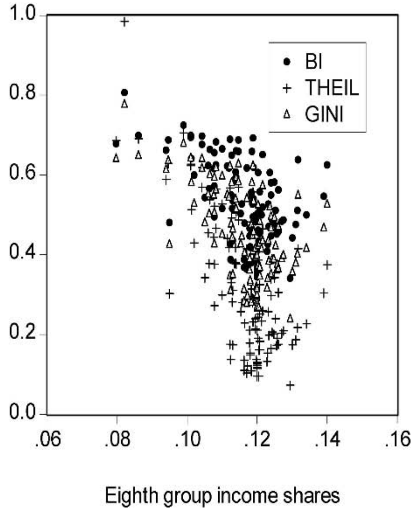


Fig.13 Ninth group income shares and various inequality measures

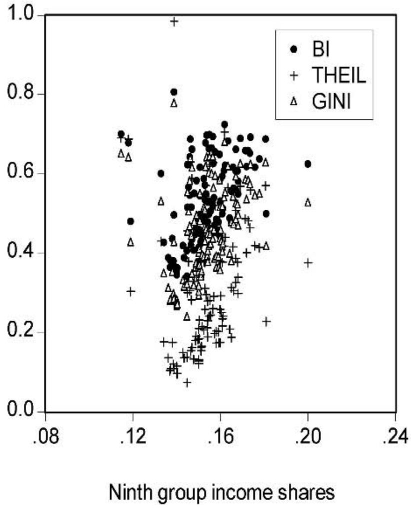
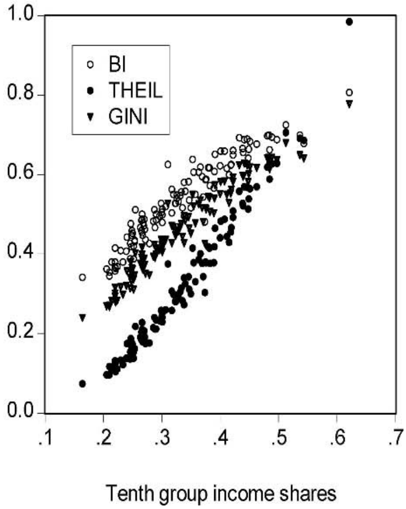


Fig.14 Tenth group income shares and various inequality measures



portions of income share increases. Figures 11-13 show the inequality measures have no definite patterns. The groups belonging to the 60-90% income share groups act as pivot points. Figure 14 shows that inequality measures increase as inequality worsens (as the tenth deciles shares increase). The extreme inequality case is found in Zimbabwe where the estimated GINI value was 0.778 and the income share of the richest 10% was 62%.

Figures 15-17 show the observed log share values and estimated log share values for Czech Republic (GINI=0.26), the U.S.A. (GINI=0.40), and the Brazil (GINI=0.63). For the given GINI and BI values, log shares are derived by the maximum entropy method. For countries with smaller or moderate GINI values (such as the Czech Republic and the U.S.A.) the BI produced better approximation for the log share functions. For countries with large GINI values, the GINI produced better approximation for the log share functions.

The whole purpose of introducing the BI was to describe the changes of income shares for the poorest group as well. This does not necessarily mean that the BI is effective in describing the inequality of an extremely uneven society like Brazil, but it means that the BI is effective in describing the relative share changes of the poor groups of certain countries. The BI was good for the Czech Republic and the U.S.A., but was bad for an extremely uneven society like Brazil. The income share of the poorest group in Brazil was 0.7% and the next poor group income share was 1.4%. Those small numbers gave negligible effects to the BI and the GINI. In comparison the income share of the poorest group in the Czech Republic was 3.98% and the next

Fig. 15 Derived log shares for CZECH

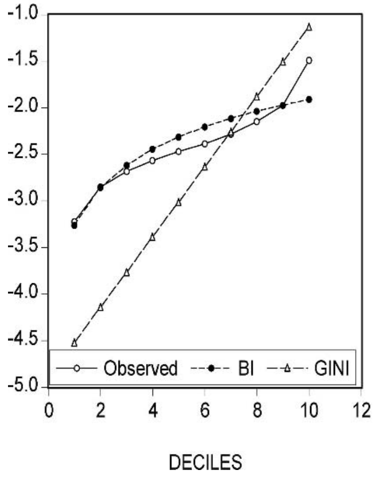


Fig. 16 Derived log shares for U.S.A.

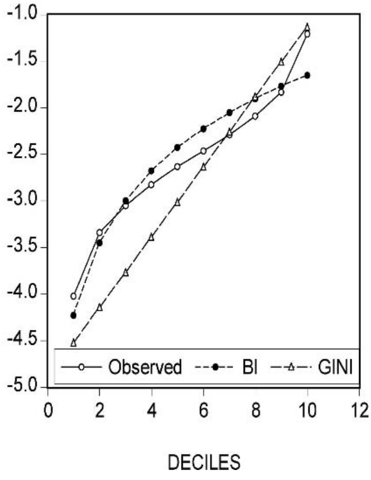
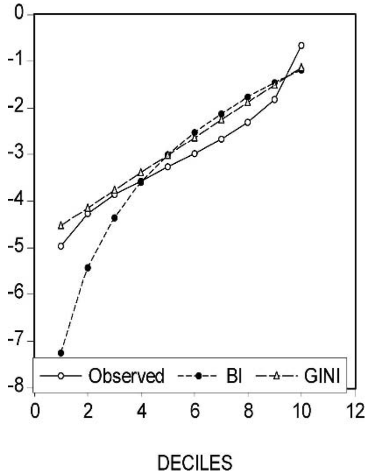


Fig. 17 Derived log shares for Brazil



poor group income share was 5.77%. Such big numbers gave considerable effects to the BI and the GINI. Thus the use of the BI produced limited results for the changes of income shares for the poorest group in Brazil, but good results for the Czech Republic and the U.S.A.

The reason why the GINI produced a better result for the relative changes of the poor groups of Brazil is the following. From the maximum entropy estimation method, the BI produced the logarithm of the share function to have a functional form of $\log(z)$ and the GINI produced a linear function of z. See (10.24) and (10.31) for the functional forms and figures 15-17 for graphical shapes. If the observed log share function is a concave function as in the cases of the Czech Republic and the U.S.A., then the log share function increases but not too rapidly and the BI produced better approximations for those countries. If the observed log share function can be approximated by a linear function, then the share function is a rapid increasing function for such uneven society and the GINI method is productive.

5 Summary and Concluding Remarks

To test the performance and usefulness of an inequality measure, the underlying income distributions are derived from the given inequality measures. Ryu and Slottje (2003) showed how to derive income distributions from the GINI and the generalized Gini coefficient proposed by Yitzhaki (1983). In this paper (as an extension of the above paper) the BI measure is reviewed and the corresponding income distributions are derived from the BI using the maximum entropy method. The BI measure gives more weight to the lower income groups as in the case of the generalized Gini. The difference is the BI uses one summary number, but the generalized Gini coefficients use a control variable to change the weight to the lower income groups.

The objective of introducing the maximum entropy estimation of income distributions is as follows. It provides some guideline which income inequality measure performs better for certain countries. Though the distinction is not clear cut, figures 15, 16, and 17 report that the BI performs better for the Czech Republic (GINI=0.26) and the U.S.A. (GINI=0.40), but poorly for Brazil (GINI=0.63). The GINI coefficient performs better for Brazil, but not for the Czech Republic and the U.S.A. Thus the GINI coefficient is an appropriate measure to describe income inequality of the very uneven distribution like Brazil and the BI is better one for very evenly distributed country like the Czech Republic and for moderately distributed country like the U.S.A. The point of the maximum entropy estimation of the income distribution is that not much information about inequality is lost if the BI is used for the Czech Republic and the U.S.A. and if the GINI is used for Brazil. To study the year to year income inequality changes of the U.S.A., the BI is better than the commonly used GINI.

Derivations of the underlying distributions were possible for the GINI, BI, and the generalized Gini, but it was not possible for other measures. The future aim is to derive income distributions from other measures such as Theil's entropy measure or Atkinson's measure to compare the performance of various inequality measures.

References

Bonferroni, C. (1930) *Elemente di Statistica Generale*, Libereria Seber, Firenze.

Choo, H. and H. Ryu (1994) Gini Cefficient, Lorenz Curves, and Lorenz Dominance Effect: An Application to Korean Income Distribution Data, *Journal of Economic Development*, **19**(2), 47–65.

Dagum, C. (1977) A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, **30**, 413–436.

Dagum, C. (1996) A Systematic Approach to the Generation of Income Distribution Models, *Journal of Income Distributions*, **6**, 105–126.

Dagum, C. and D. J. Slottje (2000) A New Method to Estimate the Level and Distribution of Household Human Capital with Application, *Structural Change and Economic Dynamics*, **11**, 67–94.

Dwight, H. (1961) *Tables of Integrals and Other Mathematical Data*, the Macmillan Company, NY, New York.

Giorgi, G. and M. Crescenzi (2001) A Look at the Bonferroni Inequality Measure in a Reliability Framework, *Statistica*, **anno LXL 4**, 571–583.

Giorgi, G. and R. Mondani (1995) Sampling Distribution of the Bonferroni Inequality Index from Exponential Population, *Sankhyã*, **57, Series B**, 10–18.

Jaynes, E. (1979) Concentration of Distributions, in R. Rosenkrantz (ed.) *Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, The Netherlands.

Lambert, P. (1993) *The Distribution and Redistribution of Income*, 2nd ed., Manchester University Press, Manchester.

Mehran, F. (1976) Linear Measures of Income Inequality, *Econometrica*, **44**(4), 805–809.

Nygard, F. and A. Sandstorm (1981) *Measuring Income Inequality*, Almqvist and Wicksell International, Stockholm.

Pundir, S., A. Sangeeta and K. Jain (2005) Bonfeeroni Curve and the Related Statistical Inference, *Statistics & Probability Letters*, **75**, 140–150.

Ryu, H. (1993) Maximum Entropy Estimation of Density and Regression Functions, *Journal of Econometrics*, **56**, 337–440.

Ryu, H. and D. J. Slottje (1998) *Measuring Trends in U.S. Income Inequality, Theory and Applicaitons*, Springer, New York.

Ryu, H. and D. J. Slottje (2003) *Estadistica*, vol. 55, Instituto Interamericano de Estadistica.

Sawyer, M. (1976) Income Distribution in OECD Countries, OECD Economic Outlook, OECD, Paris.

Shannon, C. (1948) The Mathematical Theory of Communication, *Bell System Technical Journal*, July-Oct, reprinted in: C. Shannon and W. Weaver, *The Mathematical Theory of Communicaiton*, 3-91, University of Illinois Press, Urbana, IL.

Tarsitano, A. (1990) The Bonfferoni Index of Income Inequality, in C. Dagum and M. Zenga (eds.) *Income and Wealth Distribution, Inequality, and Poverty*, pp. 228–242, Springer-Verlag, Berlin.

The UNU/WIDER World Income Inequality Database WIID (2005) Wolrd Income Inequality Database V 2.0a June 2005, World Institute for Development Economic Research, `www.wider.unu.edu/wiid/wiid.htm`.

Theil, H. (1967) *Economics and Information Theory*, North-Holland, Amsterdam, Holland.

Yitzhaki, S. (1983) On an Extension of the Gini Index, *International Economic Review*, **24**, 617–628.

Yitzhaki, S. (1998) More Than A Dozen Alternative Ways of Spelling Gini, *Research on Economic Inequality*, **8**, 13–30.

Zellner, A. and R. Highfield (1988) Calculation of Maximum Entropy Distribution and Approximation of Marginal Posterior Distributions, *Journal of Econometrics*, **37**, 195–209. An earlier verison of this paper was presented to the meeting of the NBER–NSF seminar on Bayesian inference in econometrics, October, 1982.

CHAPTER 11

New Four- and Five-Parameter Models for Income Distributions

William J. Reed[†] and Fan Wu[‡]

Abstract

Two parametric models for income distributions are introduced. The models fitted to log(income) are the 4-parameter *normal-Laplace* (NL) and the 5-parameter *generalized normal-Laplace* (GNL) distributions. The NL model for log(income) is equivalent to the double-Pareto lognormal (dPlN) distribution applied to income itself. Definitions and properties are presented along with methods for maximum likelihood estimation of parameters. Both models along with 4- and 5-parameter beta distributions, are fitted to nine empirical distributions of family income. In all cases the 4-parameter NL distribution fits better than the 5-parameter generalized beta distribution. The 5-parameter GNL distribution provides an even better fit. However fitting of the GNL is numerically slow, since there are no closed-form expressions for its density or cumulative distribution functions. Given that a fairly recent study involving 83 empirical income distributions (including the nine used in this paper) found the 5-parameter beta distribution to be the best fitting, the results would suggest that the NL be seriously considered as a parametric model for income distributions.

[†] Department of Mathematics and Statistics, University of Victoria, PO Box 3045, Victoria, B.C., Canada V8W 3P4, e-mail: reed@math.uvic.ca

[‡] Department of Mathematics and Statistics, University of Victoria, PO Box 3045, Victoria, B.C., Canada V8W 3P4, e-mail: fwu@uvic.ca.

1 Introduction

One area of interest of Camilo Dagum was the development of parametric models for income distributions (IDs). Indeed two such distributions bear his name (Dagum Types I and II). It is a measure of his authority in the field that he was chosen to write the entry entitled Income Distribution Models in the first edition of the *Encyclopedia of Statistical Sciences* (Dagum, 1983).

The interest in finding parametric models for IDs goes back over 100 years to the work of Pareto (1897), who demonstrated the power-law behaviour of empirical IDs, at least in the upper tail. The eponymous *Pareto distribution* is a two-parameter distribution with a power-law density on a support (x_0, ∞). Gibrat (1931) proposed the (two-parameter) *lognormal distribution* as model for IDs, based on a multiplicative stochastic model for the growth of individual incomes. This idea was further explored by Aitchison and Brown (1969). Other two-parameter distributions which have been proposed are the *gamma distribution* (Ammon, 1895) and the *Weibull distribution* (Bartels and van Metelen, 1975).

Three-parameter models which have been proposed include generalized forms of the gamma (Taillie, 1981) and beta (Thurow, 1970) distributions, and the (Singh and Maddala, 1976) and *Dagum Type I* (Dagum, 1977) distributions, which are both particular cases of what is known in the statistics literature as the *Burr family* (see *e.g.* (Johnson *et al.*, 1994, p. 54) of distributions (Types XII and III respectively).

Two four-parameter generalizations of the beta distribution were introduced by McDonald (1984) and named GB1 and GB2. McDonald showed that all of the previously mentioned two- and three-parameter models occurred as special or limiting cases of one or other of these two distributions. Parker (1999) presented a theoretical model of income generation, which resulted in a GB2 distribution of earnings.

Other four-parameter models which have been considered include the *Dagum Type II* distribution (Dagum, 1996) and the *double Pareto-lognormal*(dPlN) distribution (Reed and Jorgensen, 2004). This latter model emerges from a stochastic model for the way in which income distributions are generated and which offers an explanation for why Pareto's law (power-law behaviour in the upper tail) should be expected to hold (Reed, 2003). A discrete time formulation of the model is given in Reed (2004). To date no broad study of the performance of this model, with respect to its fit to IDs has been conducted. One of the purposes of this article is to remedy this.

McDonald and Xu (1995) developed the five-parameter *generalized beta* (GB) distribution family. This family includes the GB1 and GB2 as special cases (of course along with all of the two- and three-parameter distributions nested within them). Bandourian *et al.* (2002) made a comparison of the distributions contained in the generalized beta family by fitting them to 83 datasets covering 23 countries.

If incomes follow a dPlN distribution then the logarithm of incomes follow a *normal-Laplace* (NL) distribution, which is the convolution of normal and Laplace (double exponential) components (Reed and Jorgensen, 2004). Just as GB1 and GB2 distributions can be generalized to the 5-parameter generalized beta family, so the NL distribution can be generalized to a five-parameter family of distributions, which includes as special cases many other distributions, including the NL, the normal and the Laplace and the generalized Laplace (Kotz *et al.*, 2001) distributions. It has been called the *generalized normal Lapace* (GNL) distribution (Reed, 2007). In this paper we consider fitting this GNL distribution to the logarithm of incomes[21], along with the NL. Unlike the 5-parameter generalized beta family, which has a finite support, the GNL model for incomes has support on the positive reals.

A description of the NL distribution and some of its properties are presented in the next section and in Section 3 the GNL distribution is introduced. Section 4 deals with maximum likelihood estimation from grouped data and from percentile data. In Section 5 comparisons of the fit of the four-parameter NL and the five-parameter GNL distributions with the four- and five-parameter generalized beta distributions are presented. For the nine datasets considered the NL and GNL perform better (in terms of their goodness of fit as assessed by four different statistics) than their beta counterparts. Indeed the four-parameter NL distribution provides a better fit than the five-parameter GB distribution.

2 The Normal-Laplace (NL) and Double-Pareto Lognormal (dPlN) Distributions

The *normal-Laplace* distribution arises as the convolution

$$Y \stackrel{d}{=} W + V$$

where W is normally distributed, $W \sim \mathrm{N}(\mu, \sigma^2)$, and V follows a *Laplace distribution*[22] with probability density function (pdf)

$$f(v) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta}\, e^{\beta v}, & \text{for } v \leq 0 \\ \\ \frac{\alpha\beta}{\alpha+\beta}\, e^{-\alpha v}, & \text{for } v > 0 \end{cases} \tag{11.1}$$

with $\alpha, \beta > 0$.

The pdf of such a NL distribution can be shown to be (Reed and Jorgensen, 2004)

$$f_{NL}(y) = \frac{\alpha\beta}{\alpha+\beta}\, \phi\left(\frac{y-\mu}{\sigma}\right) \left[R\left(\alpha\sigma - (y-\mu)/\sigma\right) + R\left(\beta\sigma + (y-\mu)/\sigma\right)\right], \tag{11.2}$$

[21] If log(income) follows a GNL distribution, one could say that income itself followed a *generalized double Pareto-lognormal* distribution on $(0, \infty)$. However this terminology seems rather clumsy and we will refer to fitting the GNL.

[22] Often the name Laplace distribution is confined to use with the symmetric version of this distribution ($\alpha = \beta$), with the name *skew-Laplace* used for the asymmetric case ($\alpha \neq \beta$). Here he refer to both symmetric and asymmetric versions as Laplace distributions.

where R is *Mills' ratio* (of the complementary cumulative distribution function (cdf) to the pdf of a standard normal variate):

$$R(z) = \frac{\Phi^c(z)}{\phi(z)} = \frac{1-\Phi(z)}{\phi(z)}.$$

The cumulative distribution function (cdf) of the NL distribution is

$$\begin{aligned} F_{NL}(y) =& \Phi\left(\frac{y-\mu}{\sigma}\right) \\ & - \frac{1}{\alpha+\beta}\,\phi\left(\frac{y-\mu}{\sigma}\right)\left[\beta R(\alpha\sigma-(y-\mu)/\sigma) - \alpha R(\beta\sigma+(y-\mu)/\sigma)\right]. \end{aligned} \tag{11.3}$$

We shall write $Y \sim \mathrm{NL}(\mu,\sigma^2,\alpha,\beta)$ to indicate that the random variable Y has the above *normal-Laplace* distribution.

If Y follows the above NL distribution then $X = e^Y$ follows a *double Pareto-lognormal* (dPlN) distribution. Thus the dPlN distribution bears the same relationship to the normal distribution as the lognormal distribution bears to the normal. It is the dPlN distribution which has been proposed as a model for income distributions (Reed, 2003) and it can be thought of as a generalization of the lognormal (which is indeed nested within it). The dPlN distribution arises as the state of a geometric Brownian motion (GBM) after an exponentially distributed time, if the initial state is lognormally distributed. Its pdf is simply

$$f_{dPlN}(x) = \frac{1}{x}\, f_{NL}(\log x) \tag{11.4}$$

while its cdf is

$$F_{dPlN}(x) = F_{NL}(\log x) \tag{11.5}$$

The dPlN model for incomes arises from a simple stochastic model first presented in Reed (2003). In this respect it satisfies the preferred property (as specified by Camilo Dagum in 1983 *Encyclopedia of Statistical Sciences* article on Income Distributions) of having a *stochastic foundation*. The model is based on the assumption that individual incomes evolve follwing GBM, with starting incomes lognormally distributed. This is the continuous-time version of a model with a long pedigree *viz.* that individual incomes evolve in a random multiplicative way (see, e.g., Gibrat, 1931; Champernowne, 1953). Gibrat derived the lognormal distribution as that of an individual after a fixed time whereas Champernowne looked at a population in equilibrium. The innovation in the paper of Reed (2003) was to assume that the population of income earners grows at a fixed rate, so that the time that any individual has been earning follows a (truncated) exponential distribution. Ignoring the truncation (due to the fact that no person lives for ever) the model implies that the distribution of incomes over the population should be given by the state of a GBM (with lognormally distributed starting state) after an exponentially distributed time *i.e.* it should follow a dPlN distribution.

A number of properties of the NL and dPlN distributions are given in Reed and Jorgensen (2004). Among these are the moment generating function (mgf) of the NL$(\mu, \sigma^2, \alpha, \beta)$ distribution

$$M_{NL}(s) = \frac{\alpha\beta \exp(\mu s + \sigma^2 s^2/2)}{(\alpha - s)(\beta + s)}. \tag{11.6}$$

from which the mean and variance and higher-order moments can be determined:

$$E(Y) = \mu + 1/\alpha - 1/\beta; \quad \text{var}(Y) = \sigma^2 + 1/\alpha^2 + 1/\beta^2 \tag{11.7}$$

The third and fourth order cumulants are

$$\kappa_3 = 2/\alpha^3 - 2/\beta^3; \quad \kappa_4 = 6/\alpha^4 + 6/\beta^4. \tag{11.8}$$

The shape of the NL distribution combines elements of both the normal and Laplace components. Like both of those distributions it is unimodal with support on $(-\infty, \infty)$. If $\alpha = \beta$, the distribution is symmetric about μ but if $\alpha > \beta$ the distribution is skewed to the left and *vice-versa*. The parameters μ and σ^2 affect the central location and spread of the distribution. Besides affecting the skewness of the distribution the parameters α and β affect the tails of the distribution (and hence its kurtosis). In fact

$$f_{NL}(y) \sim k_1 e^{-\alpha y} \ \ (\text{as } y \to \infty); \qquad f_{NL}(y) \sim k_2 e^{-\beta y} \ \ (\text{as } y \to -\infty). \tag{11.9}$$

where k_1 and k_2 are constants. Thus the tails of the NL distribution are fatter than those of the normal distribution, behaving like those of a Laplace distribution.

Figure 11.1 (top row) shows some example of the NL distribution. In all three panels $\mu = 1$ and $\sigma^2 = 1$; in the left and centre panels $\alpha = \beta$, so the distributions are symmetric, assuming common values of 2 (left-hand panel) and 0.5 (center panel); in the right-hand panel $\alpha = 0.5$, $\beta = 2$, so the distribution is skewed to the right. The dot-dash curve is the pdf of N$(1,1)$ (corresponding to $\alpha = \beta = \infty$), drawn for comparison purposes.

Turning now to the dPlN distribution it follows from (11.4) and (11.9) that the dPlN follows power-law behaviour in both of its tails *i.e.*

$$f(x) \sim c_1\, x^{-\alpha-1} \ \ (x \to \infty); \qquad f(x) \sim c_2\, x^{\beta-1} \ \ (x \to 0)$$

for constants c_1 and c_2. The upper-tail power-law behaviour is simply Pareto's law (so the simple model on which the dPlN distribution is based offers an explanation of Pareto's law). The power-law behaviour at zero is a prediction from the model which sems to be born out in actual data (see Reed (2003)) and was indeed identified many years ago by Champernowne (1953).

The shape of the pdf of the dPlN distribution is somewhat like that of the lognormal distribution in that it is skewed to the right. Indeed if $\beta > 1$ it is unimodal (like the lognormal pdf), but if $0 < \beta < 1$ it is monotonically decreasing. However unlike the lognormal, the tails of the dPlN follow a power-law form. In the upper

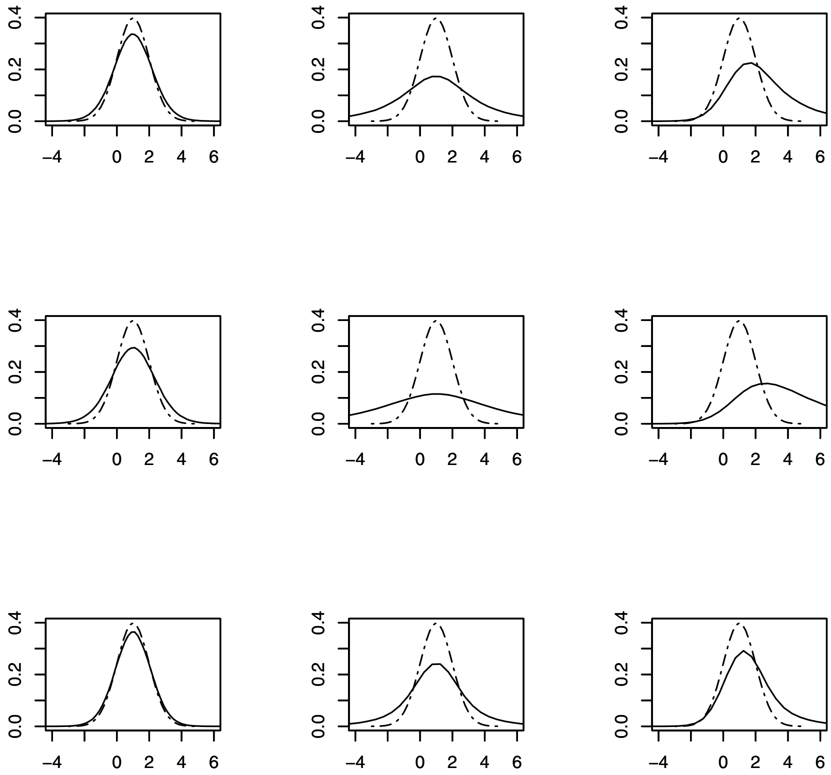


Fig. 11.1: Plots of the NL and GNL probability density functions. The top row is for the NL, and the second and third rows are for the GNL with $\rho = 2$ and 0.5, respectively. In all plots $\mu = 1, \sigma^2 = 1$. The columns correspond to $\alpha = \beta = 2$ (left-hand column); $\alpha = \beta = 0.5$ (centre column); and $\alpha = 0.5, \beta = 2$ (right-hand column). Also shown (dot-dash curve) is the normal pdf with mean and variance parameters set equal to ρ. This is for comparison purposes, to show how the NL and GNL distributions differ from the corresponding normal distribution.

tail, the smaller the value of α, so the longer the tail. In the limiting case $\alpha,\ \beta \to \infty$ the dPlN distribution tends to a lognormal distribution and for large values of $\alpha,\ \beta$ the dPlN pdf is close to that of a lognormal.

Figure 11.2 (top row) shows some examples of the dPlN pdf. The cases correspond to those of Figure 11.1 (top row) and also shown (dot-dash) is the lognormal pdf with $\mu = 1$, $\sigma^2 = 1$. Note how in the centre panel (in which $\beta = 0.5$) the dPlN pdf is decreasing over $(0, \infty)$.

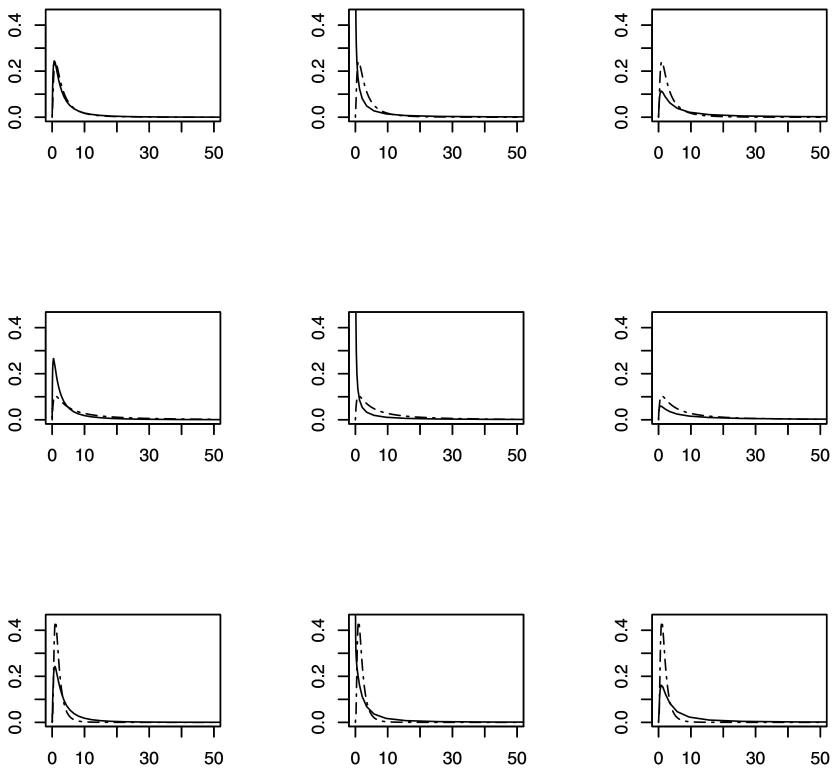


Fig. 11.2: Plots of the dPlN and generalized dPlN probability density functions. The top row is for the dPlN, and the second and third rows are for the generalized dPlN with $\rho = 2$ and 0.5, respectively. The panels correspond to those in Figure 11.1 *i.e.* in all plots $\mu = 1, \sigma^2 = 1$, and the columns correspond to $\alpha = \beta = 2$ (left-hand column); $\alpha = \beta = 0.5$ (centre column); and $\alpha = 0.5, \beta = 2$ (right-hand column). Note how in the middle column (which has $\beta = 0.5 < 1$) the pdfs are monotone decreasing. Also shown (dot-dash curve) is the lognormal pdf with mean and variance parameters set equal to ρ. This is for comparison purposes, to show how the dPlN and generalized dPlN distributions differ from the corresponding lognormal distribution.

Like the pdf of the log-hyperbolic distribution (Barndorff-Nielsen, 1977), when plotted on logarithmic axes, the dPlN pdf has a shape similar to a hyperbola, with asymptotes of slope $-(\alpha + 1)$ and $\beta - 1$. In the case $0 < \beta < 1$, both arms have negative slope. How the tail parameters α and β depend on the parameters of the stochastic model from which the dPlN distribution is derived is discussed in Reed (2003); and for a related discrete-time formulation in Reed (2004).

The moment generating function of the dPlN distribution does not exist. However lower-order moments about zero are easy to obtain. They are

$$\mu_r' = \mathrm{E}(X^r) = \frac{\alpha\beta}{(\alpha - r)(\beta + r)} \exp\left(r\nu + r^2\sigma^2/2\right) \tag{11.10}$$

for $r < \alpha$. As with the Pareto distribution μ_r' does not exist for $r \geq \alpha$. The mean (for $\alpha > 1$) is

$$\mathrm{E}(X) = \frac{\alpha\beta}{(\alpha-1)(\beta+1)} e^{\mu+\sigma^2/2} \tag{11.11}$$

while the variance and coefficient of variation (for $\alpha > 2$) are

$$\mathrm{var}(X) = \frac{\alpha\beta e^{2\mu+\sigma^2}}{(\alpha-1)^2(\beta+1)^2}\left[\frac{(\alpha-1)^2(\beta+1)^2}{(\alpha-2)(\beta+2)} e^{\sigma^2} - \alpha\beta\right] \tag{11.12}$$

and

$$\mathrm{CV} = \left[\frac{(\alpha-1)^2(\beta+1)^2)}{\alpha\beta(\alpha-2)(\beta+2)} e^{\sigma^2} - 1\right]^{1/2}$$

Clearly the CV is independent of μ, increases with σ^2 and decreases with α and β. Closed-form expressions exist neither for the Gini coefficient nor for the Lorenz curve.

3 The Generalized Normal-Laplace (GNL) and Generalized Double-Pareto Lognormal Distributions

The *generalized normal-Laplace* (GNL) distribution is defined in terms of its characteristic function. The characteristic function of the ordinary NL distribution is

$$\phi_{NL}(s) = \mathrm{E}(e^{isY}) = \frac{\alpha\beta \exp(\mu is - \sigma^2 s^2/2)}{(\alpha - is)(\beta + is)}, \tag{11.13}$$

where i is the imaginary square root of -1 (*i.e.* $i^2 = -1$). The *generalized normal-Laplace* (GNL) distribution is obtained by considering the distribution with characteristic function

$$\phi_{GNL}(s) = \left[\frac{\alpha\beta \exp(\mu is - \sigma^2 s^2/2)}{(\alpha - is)(\beta + is)}\right]^{\rho} \tag{11.14}$$

where $\rho > 0$, and we write $Y \sim \mathrm{GNL}(\mu, \sigma^2, \alpha, \beta, \rho)$ to indicate that the random variable Y follows this distribution. There are no closed-form expressions for the pdf and cdf of the GNL distribution. However it is simple to determine the shape of the $\mathrm{GNL}(\mu, \sigma^2, \alpha, \beta, \rho)$ pdf for given parameter values by simulating many replications of the random variable, using the representation

$$Y \stackrel{d}{=} \rho\mu + \sigma\sqrt{\rho}Z + \frac{1}{\alpha}G_1 - \frac{1}{\beta}G_2 \tag{11.15}$$

where Z, G_1 and G_2 are independent with $Z \sim$ N(0,1) and G_1, G_2 gamma random variables with scale parameter 1 and shape parameter ρ, *i.e.* with probability density function (pdf)

$$g(x) = \frac{1}{\Gamma(\rho)} x^{\rho-1} e^{-x}.$$

(see Reed (2003)).

In general for $\rho < 1$, the GNL pdf is longer in the tails, narrower in the flanks and more peaked than the corresponding NL distribution. The opposite holds for $\rho > 1$.

The mean and variance of the GNL$(\mu, \sigma^2, \alpha, \beta, \rho)$ distribution are

$$\mathrm{E}(Y) = \rho\left(\mu + \frac{1}{\alpha} - \frac{1}{\beta}\right); \qquad \mathrm{var}(Y) = \rho\left(\sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2}\right)$$

while the higher order cumulants are (for $r > 2$)

$$\kappa_r = \rho(r-1)!\left(\frac{1}{\alpha^r} + (-1)^r \frac{1}{\beta^r}\right).$$

As with the ordinary NL distribution, the parameters μ and σ^2 influence the central location and spread of the distribution, while α and β affect the symmetry. If $\alpha > \beta$ the distribution is skewed to the left, and vice versa. The parameter ρ affects the lengths of the tails. The tails are respectively longer (shorter) than those of the NL$(\mu, \sigma^2, \alpha, \beta)$ distribution depending on whether $\rho < 1$ (or $\rho > 1$). Precisely $f(y) \sim c_1 y^{\rho-1} e^{-\alpha y}$ $(y \to \infty)$ and $f(y) \sim c_2(-y)^{\rho-1} e^{\beta y}$ $(y \to -\infty)$, (where c_1 and c_2 are constants).

The parameter ρ affects all moments. However the coefficients of skewness $(\gamma_1 = \kappa_3/\kappa_2^{3/2})$ and of excess kurtosis $(\gamma_2 = \kappa_4/\kappa_2^2)$ both decrease with increasing ρ (and converge to zero as $\rho \to \infty$) with the shape of the distribution becoming more normal with increasing ρ, (exemplifying the central limit effect since the sum of n iid GNL$(\mu, \sigma^2, \alpha, \beta, \rho)$ random variables has a GNL$(\mu, \sigma^2, \alpha, \beta, n\rho)$ distribution).

Figure 11.1 (second and third rows) display some plots of the pdf of the GNL distribution. In the centre and left panels the distributions are symmetric with $\alpha = \beta$ and equal to 2 and 0.5 respectively. In the right-hand panels $\alpha = 0.5$ and $\beta = 2$, with the distributions being skewed to the right. A comparison of the three rows shows the effect of changing ρ, corresponding to $\rho = 1$ (top row, ordinary NL); $\rho = 2$ (middle row) and $\rho = 0.5$ bottom row. Notice how decreasing ρ results in a thinning of the flanks of the distribution with the probability mass near zero becoming larger (more peaked distribution). At the same time the tails become longer, although this does show up in the figures.

The generalized dPlN distribution bears the same relationship to the GNL, as does the dPlN to the NL (or the lognormal to the normal) *i.e.* the log of a generalized dPlN random variable follows a GNL distribution. The shape of the generalized dPlN distribution is somewhat similar to that of the dPlN. This can be seen in

Figure 11.2, where the parameter values in the panels are the same as those in Figure 11.1 - in particular the top row corresponds to $\rho = 1$ (ordinary NL); the middle row to $\rho = 2$ and the bottom row to $\rho = 0.5$. Notice how the pdf is monotone decreasing in the three panels in the middle column, because of the fact that $\beta = 0.5 < 1$.

4 Fitting the Models to Income Data by Maximum Likelihood

To fit the dPlN to grouped income data (with cell boundaries $0 = x_1 < x_2 < x_3 \dots$) by maximum likelihood, one needs to maximize numerically the log likelihood

$$\sum_j f_j \log \theta_j \tag{11.16}$$

where the f_j are the frequencies in the classes and $\theta_j = \theta_j(\mu, \sigma^2, \alpha, \beta)$ is the probability of an observation in class j, or

$$\theta_j = F_{NL}(\log x_{j+1}) - F_{NL}(\log x_j)$$

where F_{NL} is the cdf of the NL (11.3).

If the data are in the form of percentiles of the distribution, then the likelihood is proportional to the joint distribution of the order statistics corresponding to the empirical percentiles. For example if $x_{(1)}, x_{(2)}, \dots, x_{(20)}$ are the 5th, 10th, ..., 95th percentiles of a sample of size N, then the log-likelihood is of the form

$$\ell = c + \sum_{i=1}^{19} \log f_{NL}(\log x_{(i)}) + \frac{N}{20} \sum_{i=0}^{19} \log[F_{NL}(\log x_{(i+1)}) - F_{NL}(\log x_{(i)})]. \tag{11.17}$$

Typically in income distribution studies the sample size N will be very large, so that the second summation term in the log-likelihood will dominate over the first. If one ignores the first term one arrives at precisely the multinomial log likelihood, (11.16), above. In either case one needs to maximize the log-likelihood (11.16) or (11.17) numerically over the four parameters, μ, σ^2, α and β. Method of moments estimates, based on the first four moments (for log(income) – (11.7), (11.8)) can be used as starting values.

To fit the 5-parameter GNL (or generalized dPlN) by maximum likelihood one can follow a similar procedure, except in this case there is no closed form for the cdf, corresponding to (11.3). However for given parameter values the cdf at a point can be computed numerically by inverting the characteristic function (11.14). This enables computation of the likelihood function and thence its maximization. Of course the numerical inversion of the characteristic function adds considerably to computation time.

Table 11.1: Goodness of fit of two four-parameter distributions (GB2 and NL); and two five-parameter distributions (GB and GNL) to various empirical IDs. The goodness-of-fit statistics are the sum of squared errors (SSE) and the sum of absolute errors (SAE) between observed and fitted frequencies; the Pearson goodness-of-fit statistic (χ^2) and the maximized log-likelihood plus an additive constant, included for convenience of presentation (max ℓ).

GoF Stat.	Distribution	ID								
		AU94	BE97	CA87	CA97	IT00	MX00	TW00	UK99	US97
SSE$*10^3$	GB2	2.379	0.974	1.038	0.456	4.569	0.679	2.396	0.467	0.504
	NL	1.820	0.849	0.684	0.234	4.480	0.530	2.203	0.226	0.322
	GB	2.345	0.990	1.031	0.438	4.582	0.657	2.279	0.452	0.499
	GNL	0.702	0.494	0.360	0.165	4.238	0.349	0.204	0.139	0.276
SAE$*10$	GB2	1.437	1.078	1.037	0.791	2.418	0.892	1.408	0.696	0.738
	NL	1.323	1.027	0.854	0.604	2.345	0.778	1.270	0.491	0.566
	GB	1.353	1.096	1.027	0.757	2.393	0.837	1.321	0.686	0.719
	GNL	0.982	0.787	0.612	0.465	2.233	0.630	0.484	0.406	0.548
χ^2	GB2	195.2	54.1	191.2	228.5	628.7	127.1	495.6	133.1	418.6
	NL	152.7	47.9	129.5	118.7	586.2	102.5	457.7	64.3	266.6
	GB	192.1	53.0	189.6	219.5	624.7	123.7	470.4	128.9	414.6
	GNL	66.4	29.1	66.8	84.3	557.0	68.2	50.0	42.1	218.5
max ℓ	GB2	-136.2	-32.8	-156.6	-330.8	-77.12	-110.1	-204.3	-99.8	-480.1
	NL	-114.7	-29.5	-126.4	-275.9	-63.36	-97.7	-184.8	-64.9	-406.2
	GB	-134.8	-32.2	-155.9	-326.3	-76.45	-108.3	-191.4	-97.6	-478.2
	GNL	-69.9	-20.0	-96.5	-258.7	-51.21	-80.7	-75.5	-53.3	-384.2

5 Results of Fitting to Empirical Distributions

Nine empirical income distributions were used. Data was obtained from the Luxembourg Income Study (2004) website for a representative sample of cases considered by Bandourian *et al.* (2002). The IDs used were household income in each of the following: Australia, 1994; Belgium, 1997; Canada, 1987, 1997; Italy, 2000; Mexico, 2000; Taiwan, 2000; UK, 1999; USA, 1997. In each case twenty intervals were used, with equal frequencies in each cell - the lower cell boundaries were thus 0, and the 5^{th} through 95^{th} percentiles, and parameter estimates were obtained by maximizing the appropriate log likelihood for the best four- and five-parameter models found by Bandourian *et al.* (2002)(*viz.* the GB2 and the GB) along with the four-parameter NL model and the five-parameter GNL model. Goodness of fit was assessed using: the sum of squared errors (SSE); the sum of absolute errors (SAE); the Pearson chi-squared statistic (χ^2) and the maximized log likelihood (max ℓ) all for percentage frequencies as in Bandourian *et al.* (2002). The results are given in Table 11.1. Note that, to simplify presentation, the same constant has been added to each value of max ℓ in a given column, in Table 11.1. This is of no importance since the log-likelihood is only defined up to an additive constant.

While the numerical results for the GB2 and GB are similar to those of Bandourian *et al.* (2002), there are some small differences. Some possible explanations for these differences are: (i) end-point differences - we used 0 and ∞ for the lower and upper boundaries of the smallest and largest class; (ii) differences in retrieval of data from LIS - we used SAS 9.13; (iii) possibility of multiple local maxima of the likelihood function.

As can be seen from the table, in all cases the four-parameter NL is performing better than the four-parameter GB2; and indeed better than the five-parameter GB. The GNL performs considerably better than all of the other distributions.

We emphasize that the nine cases reported are the only ones to which model fitting was done. They were not selected because the NL and GNL fitted well. We anticipate that similar results would hold for most if not all of the 83 datasets considered by Bandourian *et al.* (2002), and indeed for most empirical IDs.

6 Conclusions

The results of fitting theoretical distributions to empirical income data, presented in the last section, show overwhelmingly how the dPlN fits better than the GB2 distribution. Since Bandourian *et al.* (2002) claimed that the GB2 was the best fitting 4-parameter model to date, it would seem that the title "best-fitting 4-parameter model" can now be fairly applied to the dPlN. Bearing in mind that the dPlN satisfies all of the demands of a good income-distribution model as outlined by Dagum (1983) in his *Encyclopedia of Statistical Sciences* entry on Income Distributions (*viz.* (i) it is based on a plausible stochastic model; (ii) it satisfies Pareto's law in the upper tail; and (iii) it fits well to data) serious consideration should be given to its widespread use in income modelling.

The 5-parameter GNL (or generalized dPlN) model provides an even better fit than the dPlN, as of course it should since the dPlN is nested within it. However there are no closed-form expressions for its pdf or cdf, and one might question whether one needs 5 parameters to model an income distribution (especially if data is reduced to frequencies in twenty classes). Also, unlike the dPlN, it is not based on a stochastic model foundation; and furthermore, and of considerable importance, fitting the GNL to data is computationally difficult and slow.

References

Aitchison, J. and J. A. C. Brown (1969) *The Lognormal Distribution with Special References to Its Uses in Economics*, Cambridge University Press, Cambridge.

Ammon, O. (1895) *Die Gesellschaftsordnung und ihre Natürlichen Grundlagen*, Jena.

Bandourian, R., J. B. McDonald and R. S. Turley (2002) A Comparison of Parametric Models of Income Distribution across Countries and over Time, Luxembourg Income Study Working Paper, No. 305.

Barndorff-Nielsen, O. (1977) Exponential Decreasing Distributions for the Logarithm of Particle Size, *Proc. R. Soc. Lond.*, **353**, 401–419.

Bartels, C. P. A. and H. van Metelen (1975) Alternative Probability Density Functions of Income, Vrije University Amsterdam: Research Memorandum 29.

Champernowne, D. G. (1953) A Model of Income Distribution, *Economic Journal*, **63**, 318–351.

Dagum, C. (1977) A New Model for Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, **30**, 413–437.

Dagum, C. (1983) Income Distribution Models, in S. Kotz, N. L. Johnson and C. Read (eds.) *Encyclopedia of Statistical Sciences*, vol. 4, John Wiley, New York.

Dagum, C. (1996) A Systemic Approach to the Generation of Income Distribution Models, *Journal of Income Distribution*, **6**, 105–126.

Gibrat, R. (1931) *Les Inégalités Économiques*, Librairie du Recueil Sirey, Paris.

Johnson, N. L., S. Kotz and N. Balakrishnan (1994) *Continuous Univariate Distributions*, vol. 1, 2nd ed., Wiley, New York.

Kotz, S., T. J. Kozubowski and K. Podgórski (2001) *The Lapace Distribution and Generalizations*, Birkhauser, Boston.

Luxembourg Income Study (2004) www.lis.ceps.lu.

McDonald, J. B. (1984) Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52**, 647–663.

McDonald, J. B. and Y. J. Xu (1995) A Generalization of the Beta Distribution with Applications, *Journal of Econometrics*, **66**, 133–152, Erratum: Journal of Econometrics, **69**: 427-428.

Pareto, V. (1897) *Cours d'Economie Politique*, Rouge, Lausanne.

Parker, S. C. (1999) The Generalized Beta as a Model for the Distribution of Earnings, *Economics Letters*, **62**, 197–200.

Reed, W. J. (2003) The Pareto Low of Incomes - An Explanation and An Extension, *Physica A*, **319**, 579–597.

Reed, W. J. (2004) On Pareto's Law and the Determinants of Pareto Exponents, *Journal of Income Distribution*, **13**, 7–17.

Reed, W. J. (2007) Brownian-Laplace Motion and Its Use in Financial Modelling, *Communications in Statistics Theory and Methods*, **36**, 473–484.

Reed, W. J. and M. Jorgensen (2004) The Double Pareto-Lognormal Distribution - A New Parametric Model for Size Distribution, *Communications in Statistics Theory and Methods*, **33**, 1733–1753.

Singh, S. K. and G. S. Maddala (1976) A Function for the Size Distribution of Incomes, *Econometrica*, **44**, 963–970.

Taillie, C. (1981) Lorenz ordering within the Generalized Gamma Family of Income Distributions, in C. Taillie, G. P. Patil and B. Balderssari (eds.) *Statistical Distributions in Scientific Work*, vol. 6, pp. 181–192, Reidel, Boston.

Thurow, L. C. (1970) Analyzing the American Income Distribution, *American Economic Review*, **48**, 261–269.

CHAPTER 12

Fuzzy Monetary Poverty Measures under a Dagum Income Distributive Hypothesis

Gianni Betti*, Antonella D'Agostino† and Achille Lemmi‡

Abstract

This chapter explores the potential of introducing the Dagum distribution into the IFR *(Integrated Fuzzy Relative)* poverty measure. This implies using the Dagum model for fitting the empirical cumulative distribution that forms one of the components of the membership function to the set of poor in the IFR methodology. Moreover, we propose a heterogeneous Dagum model in order to allow the form of income distribution to vary with personal characteristics. In this way, we are able to make comparisons across sub-groups of the population between the traditional and the IFR measures of poverty.

1 Introduction

Poverty estimates are mainly used in policy making in order to design a plan of action to help the public authorities (international, national, regional) reduce

* Research Centre on Income Distribution "C. Dagum", University of Siena, Italy. Email: betti2@unisi.it

† Department of Statistics and Mathematics for the Economic Research, University of Naples "Parthenope", Naples, Italy. Email: antonella.dagostino@uniparthenope.it

‡ Research Centre on Income Distribution "C. Dagum", University of Siena, Italy. Email: lemmi@unisi.it

unfavourable living conditions in defined segments of a population. Such segments normally refer to demographic or geographical (or both) variables, while the hardship characteristics can be of monetary and/or of non-monetary nature; moreover they can be chronic, persistent or cyclical. Therefore policy makers should be in a position to undertake specific and efficient policy action.

A possible method for evaluating ex-ante the effects of anti-poverty policies is micro-simulation, where referring to a set of sufficient and reliable variables of living conditions (poverty indicators) and to adequate models of diffusion and distribution of such variables (or of, at least, the most relevant of them), different parametric hypotheses can be identified and statistically tested. A model representing personal or household income distribution which has the two properties of great adequacy and of deep economic meaning of its parameters can be very useful for analysing plausible outcomes associated with alternative ways of decreasing poverty or improving equality in income distribution. An income distributive model with similar characteristics is the so-called 3-parameter Dagum model, the socio-economic and mathematical foundations of which are well-known and the performances of which have been widely experimented in several different economic realities. In particular, the model parameters have a specific economic meaning and for this reason each model has an immediate advantage over its empirical equivalent. Model parameters reflect substantially different economic changes that can influence the income distribution.

This means that deterioration or improvement in the income distribution, as a result of a number of factors, can be related to them. In particular, deterioration can occur when the income distribution moves as a whole to the left and consequently the total mean income and the mean income of each decile decrease; an improvement can be the result of a tax cut that benefits a wider range of individuals or of localized subsidies directed at the poorest individuals and so forth.

The aim of this Chapter is to explore the potentiality of the Dagum distribution into the IFR *(Integrated Fuzzy Relative)* poverty measure. This implies using the Dagum model for fitting the empirical cumulative distribution that is one of the components of the membership function to the set of the poor in the IFR methodology.

In this way, our proposal bridges two research fields to which Camilo Dagum greatly contributed during his long authoritative career: income distribution and the fuzzy approach to poverty measurement.

In the first half of the 1990s, scholars working on poverty began applying the concept of total fuzzy and relative (TFR) measures (Cheli and Lemmi, 1995). During this period Dagum spent long periods of time at the University of Siena, participating in the activities of the Research Centre on Income Distribution. It turns out that his contribution played a crucial role in this research program and in fact a paper bearing his name (Dagum *et al.*, 1992) was presented at an international conference in which the TFR poverty measures were proposed for the first time to the international scientific community (Warsaw, 1992). In subsequent years Dagum further developed an approach similar to TFR leading to the so-called Dagum decomposition (Dagum and Costa, 2004).[23]

The Chapter is made up of five Sections. After this introduction, Section 2 describes the Integrated Fuzzy and Relative (IFR) approach to the measurement of poverty. Section 3 presents some theory behind the Dagum type I income distribution, and its use in the IFR approach. Section 4 briefly describes the data set of the Italian survey of EU-SILC (European Union - Statistics on Income and Living Conditions) for the year 2004, which is the basis of the application. Section 5 concludes the Chapter.

2 The Integrated Fuzzy and Relative Approach

2.1 Fuzzy set approach to poverty measurement

One of the main limitations of the traditional approach to poverty measurement is the rigid poor/non-poor dichotomisation; it is undisputable that such a clear-cut division causes a loss of information and removes the nuances that exist between the two extremes of substantial welfare on the one hand and distinct material hardship on the other. In other words, poverty should be considered as a matter of degree rather than as an attribute that is simply present or absent for individuals in the population.

An early attempt to incorporate this concept at the methodological level (and in a multidimensional framework) was made in Cerioli and Zani (1990) which, in proposing the Totally Fuzzy and Absolute (TFA) approach, drew inspiration from the theory of Fuzzy Sets initiated in Zadeh (1965) and developed in Dubois and Prade (1980). Given an X set of elements $x \in X$, any fuzzy subset A of X is defined as follows: $A = \{x, \mu_A(x)\}$, where $\mu_A(x)$: $X \rightarrow [0,1]$ is called the *membership function (m.f.)* in the fuzzy subset A.

The value $\mu_A(x)$ indicates the degree of membership of x in A. Thus $\mu_A(x)$= 0 means that x does not belong to A, whereas $\mu_A(x)$= 1 means that x belongs to A completely. With $\mu_A(x) \in (0-1)$, x belongs to A partially and its degree of membership of A increases in proportion to the proximity of $\mu_A(x)$ to 1.

[23] He was also the promoter, together with Samuel Kotz and one of the authors of this Chapter, of an international conference to commemorate two authoritative social scientists, Max Otto Lorenz and Corrado Gini. Under his guidance, the conference was held at the Certosa di Pontignano of the University of Siena (during the month of May 2005) where he had often spent periods studying, working and carrying out research at the Centre for Research on Income Distribution at the University of Siena. For this reason this place was very dear to him. In his memory, the abovementioned Centre is now dedicated to him and the people who work there remember Camilo with great affection and gratitude.

At that Conference the paper by Betti *et al.* (2006) was presented for the first time. It was the first to link the fuzzy set approach to the Generalised Gini coefficient, the basis for the new proposal of this Chapter.

The original proposed approach in Cerioli and Zani (1990) was later developed in Dagum *et al.* (1992) and in Cheli and Lemmi (1995) giving origin to the so-called *Totally Fuzzy and Relative* (TFR) approach. Both methods have been subsequently applied by a number of authors (see Berenger and Verdier-Chouchane (2007) for an excellent review of the literature). For instance, Chiappero (2000), Lelli (2001) and Qizilbash (2003) use the TFR method in order to analyse poverty or well-being according to Sen's capability approach. Baliamoune-Lutz (2006) was the first to apply the fuzzy set theory in order to construct measures at macro level. Moreover, recently McGillivray and Clarke (2006), Ruggeri *et al.* (2006) and Rojas (2006) have focused their attention on the measurement of well-being using fuzzy set theory.

2.2 From the TFR to the IFR

The TFR method was refined in Cheli (1995a) which used it to apply the fuzzy approach to poverty measurement to a dynamic context where two consecutive panel waves are available. From this point on, the methodological implementation of this approach took two directions, with a somewhat different emphasis despite their common orientation and framework. The first direction emphasized the time dimension (Cheli and Betti, 1999; Betti *et al.*, 2004) via the use of transition matrices. Another direction is taken in Betti and Verma (2008) which focuses more on capturing the multi-dimensional aspects of poverty and develops the concepts of 'manifest' and 'latent' deprivation to reflect the intersection and union of different dimensions.

In this Chapter we take into account a further development of the TFR which has led to an *Integrated Fuzzy and Relative* (IFR) approach to the analysis of poverty and social exclusion (Betti *et al.*, 2006; Lemmi and Betti, 2006).

The IFR measure is defined by taking into account the TFR approach (Cheli and Lemmi, 1995) and the approach of Betti and Verma (1999). Here we take into account only the monetary version of the IFR; let y_i be the equivalised income of individual i, the IFR measure is given by:

$$\mu = FM_i = (1-F_i)^{\alpha-1}(1-L_i) = \left(\frac{\sum_{\gamma=i+1}^{n} w_\gamma}{\sum_{\gamma=2}^{n} w_\gamma}\right)^{\alpha-1}\left(\frac{\sum_{\gamma=i+1}^{n} w_\gamma y_\gamma}{\sum_{\gamma=2}^{n} w_\gamma y_\gamma}\right); \mu_n = 0 \qquad (12.1)$$

where F_i is the income distribution function, w_γ is the sample weight of individual of rank γ (1 to n) in the ascending income distribution and finally α is a parameter chosen so that the mean of these measures equals the head count ratio (H) given by:

$$H = \left(\sum_{1=1}^{n} w_i\right)^{-1} \sum_{1=1}^{n} w_i I\{y_i < z\} \qquad (12.2)$$

where $I\{.\}$ is the indicator function, z is the poverty line.

3 The Dagum Income Distribution Model

3.1 A brief introduction to the Dagum models

In the analysis of income distribution the mathematical description of the size distribution of income to approximate the true distribution of income has been frequently considered one of the main objectives of many researchers in the formal analysis of welfare economics. In particular, we focus our attention on the parametric specification of the Dagum model.

Dagum (1977) introduced his model in order to fit the distribution of personal income and he discussed the essential and important (but not necessary) properties for a probability density function to be specified as a model of income or wealth distribution. This paper was a further generalization of the one proposed by him during the first half of the 1970s (Dagum, 1973, 1975). The Dagum model was independently specified also by Fattorini and Lemmi (1979) who derived the same model starting from a set of stochastic assumptions on the infinitesimal mean and variance of a continuous stochastic process. Then later Dagum proposed a further generalization (Dagum, 1980a) underlining the relevance of the economic meaning of the model parameters. The goodness of fit of this model (in both its four and three-parameter versions) outperformed the models most frequently applied in the literature such as the lognormal and the Gamma models (Aitchison and Brown, 1957; Salem and Mount, 1974) and also the Singh and Maddala (1976) model.

In fact, empirical applications based on income data from several countries - Canada, the United States, Italy and Argentina (Dagum, 1983, 1990; Botargues and Petrecolla, 1999) - provided strong evidence of the power of the Dagum model in producing superior descriptions of the whole range of income. Dagum and Lemmi (1989) showed that in a three-parameter version, the Dagum model provides a very flexible parametric distribution and superior performance in a considerable number of empirical results compared with the most popular and widespread interpretative models (Kleiber, 1996). Recently, Dastrup *et al.* (2007) explored the impact of taxes and transfer payments on the distribution of income across 13 countries for different years using the Luxembourg Income Study data, and discovered that the three-parameter Dagum distribution shows one of the best fits for earnings in almost all countries.

In the statistics literature, the Dagum model belongs to a classification system drawn up by Burr (1942) and, in particular, the three-parameter version is known as Burr III distribution (Tadikamalla, 1980). This form is the simplest of the Dagum models and is known as Dagum type I. Considering the great performance of the fitting of the three-parameter Dagum distribution, in this Chapter, we refer only to this one.

Let y be the random variable with cdf given by:

$$F(y) = \left\{ \begin{array}{ll} \frac{1}{\left(1+\lambda y^{-\delta}\right)^{\beta}} & ,y > 0 \\ 0 & ,y \leq 0 \end{array} \right\} \tag{12.3}$$

where $\lambda > 0,\ \delta > 0,\ \beta > 0$. The mathematical specification in (12.3) describes a Dagum type I distribution.

The corresponding probability density function (pdf) is given by:

$$f(y) = \beta\lambda\delta y^{-\delta-1}\left(1+\lambda y^{-\delta}\right)^{-\beta-1}. \tag{12.4}$$

As we said above, Dagum (1977, 1980b) explained the well-defined economic meaning of the three parameters: i) λ is a scale parameter, moreover, $\lambda^{\left(-\frac{1}{\delta}\right)}$ has the same dimension as income y. For this reason, it accounts for the monetary scale, henceforth it is a tool to adjust for inflation (Dagum and Lemmi, 1989) and to facilitate cross-country comparisons of income distribution that are expressed in different monetary units; ii) δ and β are shape parameters and the Gini ratio is a decreasing function of both; for this reason, they are also interpreted as equality parameters. Moreover, they are scale free, thus, the scale parameter λ does not affect the measurement of the parametric Gini ratio.

Model (12.3) also has an explicit mathematical expression for inequality measures; thus, for example, the Lorenz curve is given by:

$$L(y) = \frac{B\left(t^{\frac{1}{\beta}};\beta+\frac{1}{\delta},1-\frac{1}{\delta}\right)}{B\left(\beta+\frac{1}{\delta},1-\frac{1}{\delta}\right)} \tag{12.5}$$

where $t = F(y)$ and $\mathrm{B}(.)$ is the Beta function.

Moreover, this model also yields an explicit mathematical solution for location measures (median, mode, mean). For example, the median is given by:

$$\xi_{0.5} = \lambda^{\frac{1}{\delta}}\left[\left(\frac{1}{0.5}\right)^{\frac{1}{\beta}}-1\right]^{\frac{-1}{\delta}}. \tag{12.6}$$

Using equation (12.6) consequently allows estimating a poverty line z_D as a quantile poverty line. This means that $z_D = \eta\xi_q$, where ξ_q is the theoretical quantile of order q of the distribution in (12.6) and η is a parameter greater than zero. Given z_D we can then compute a poverty measure as the head count ratio H_D as:

$$H(z_D) = F(z_D|\lambda,\beta,\delta) = \frac{1}{\left(1+\lambda z_D^{-\beta}\right)^{\delta}} \tag{12.7}$$

that is the proportion of the population below the poverty line.

3.2 The use of the Dagum model in the IFR approach

The IFR measure is based on two main considerations: i) it takes into account both the proportion of individuals less poor than the person concerned, and the share of the total equivalised income received by all those less poor than the person concerned; ii) both c.d.f. and Lorenz curve are estimated from the data and so they accommodate large flexibility for modelling personal income data.

On the other hand, the empirical estimation of $F(.)$ and $L(.)$ (see equation (12.1)) has some limitations which can be overcome by a parametric specification for the size distribution of income. In fact, the researcher generally believes that the probability distribution underlying a process that generates income data is reasonably smooth. Therefore, by smoothing the data using a parametric model, one expects to obtain better estimates rather than if one just uses the empirical data. Moreover, the smoothing can be very useful in order to recover relevant information which are currently unavailable or which have been lost by focusing on published measures that are not calculated coherently over time by government agencies. Furthermore they can facilitate both the mathematical analysis of the basic structure of income and harmonize information from two or more sources.

Starting from this consideration, here we propose to model $F(.)$ using a Dagum type I distribution whose parameters have a well-defined economic meaning. Such a model in fact summarizes in three parameters the regularities discovered in empirical distribution. Also, they can be employed to compute summary measures that can be compared spatially and temporally. This is because, as we explained in Section 3.1, its location, poverty and inequality measures can be expressed in terms of its distributional parameters. Moreover, the sufficient flexibility of the Dagum type I model provides a suitable fit for observed personal income distribution of developed as well as developing countries. Following this perspective, equation (12.1) is computed in the same way, as explained in Section 2, but $F(.)$ is assumed to follow a Dagum type I distribution whose parameter estimates are obtained using maximum likelihood estimation by maximizing the log-likelihood function, given by:

$$\ln L = \sum_{i=1}^{n} \omega_i \left(\ln\beta + \ln\lambda + \ln\delta + (-\delta - 1)\ln y_i - (\beta + 1)\ln\left(1 + \lambda y_i^{-\delta}\right)\right) \quad (12.8)$$

where ω_i is the sample weight of individual i.[24] In order to distinguish the Integrated Fuzzy and Relative measure based on Dagum type I model with respect to the one expressed in equation (12.1), we term this measure as FM_i^D.

3.3 The use of the heterogeneous Dagum model in the IFR approach

In inequality and poverty study, one of the main economics aims of analysts is to make spatial or temporal (or both) comparisons in order to discover poverty (inequality) differences among sub-groups of population or poverty (inequality) persistence across time. In this Chapter, we only referred to spatial comparison.

[24] Also Cheli (1995b) used the Dagum model in the TFR approach.

In Section 3.2 we have introduced the Dagum model in the IFR approach and we have shown how the α parameter is computed so that the mean of FM_i equals the head count ratio H. In this way, we are able to make comparisons across sub-groups of the population (as for example different regions) between the traditional measure H and the IFR measure defined in equation (12.1). However, following this perspective we do not use all the capacities of the Dagum model as the head count ratio can be estimated using equation (12.7). If the model fits well we expect that the observed H and the estimated H_D from the specified distribution are very similar.

Taking into account that the aim of the study is to compare different population sub-groups we can generalize equation (12.3) in order to allow the form of income distribution to vary with personal characteristics. In this manner, we define a heterogeneous Dagum model (Biewen and Jenkins, 2005; Quintano and D'Agostino, 2006)[25] because each model parameter may be made heterogeneous as follows:

$$\beta_i = \exp(x_i\gamma_1);\, \delta_i = \exp(x_i\gamma_2);\, \lambda_i = \exp(x_i\gamma_3) \tag{12.9}$$

where x_i is a $1\times m$ vector of personal characteristics and $\gamma_1, \gamma_2, \gamma_3$ are $m\times 1$ unknown parameter vectors to be estimated. Thus equation (12.3) becomes:

$$F(y;x;\lambda;\beta;\delta) = \left\{ \begin{array}{ll} \frac{1}{\left(1+\exp(x_i\gamma_3)y^{-\exp(x_i\gamma_2)}\right)^{\exp(x_i\gamma_1)}} & ,y>0 \\ 0 & ,y\le 0 \end{array} \right\} \tag{12.10}$$

The maximum likelihood estimation of parameters are then obtained by maximizing the weighted log-likelihood function, given by:

$$\ln L = \sum_{i=1}^{n} \omega_i \left(\ln\beta_i + \ln\lambda_i + \ln\delta_i + (-\delta_i - 1)\ln y_i - (\beta_i+1)\ln\left(1+\lambda_i y_i^{-\delta_i}\right)\right) \tag{12.11}$$

The parameters $\gamma_1, \gamma_2, \gamma_3$ do not have a direct meaning but they can be used in order to calculate synthetic measures of the estimated income distribution such as the median, poverty line, head count ratio, etc. Suppose, for example, we have one categorical variable with five items. Once the model parameters have been estimated, we can compute the median, the poverty line and the head count ratio relative to each item of the considered variable. This means that if we are interested in studying personal differences with respect to macro-regions, the heterogeneous Dagum model allows overcoming the estimation of different Dagum models for each macro-region. In each macro-region $k(k = 1,2,...,K)$ the head count ratio with respect to the national poverty line z_D defined in Section 3.1 is computed as:

$$H_D^k(z_D, \widetilde{x}) = F(z_D|\gamma_1, \gamma_2, \gamma_3, \widetilde{x}) = \frac{1}{\left(1+\exp(\widetilde{x}\gamma_3)\, z_D^{-\exp(\widetilde{x}\gamma_2)}\right)^{\exp(\widetilde{x}\gamma_1)}} \tag{12.12}$$

[25] For heterogeneous income models see also Pudney (1999).

where $\widetilde{x}$ identifies the vector of covariates x for macro-region k.

Following this approach we can now compute the IFR measure using equation (12.5) where in this case α is chosen so that the mean of these measures, FM_D, equals the head count ratio H_D and $F(.)$ follows a Dagum type I distribution without personal heterogeneity.

4 Empirical Analysis

4.1 The data set: EU-SILC year 2004

The European Union collection of Statistics on Income and Living Conditions, EU-SILC has been developed as a flexible yet comparable instrument covering data and data sources of various types: cross-sectional and longitudinal; household-level and person-level; economic and social; from registers and interview surveys; from new and existing national surveys or other sources. It envisages the creation of one or more micro-data base(s) in each country to be used for the follow-up and monitoring of income and social exclusion at the EU and national level (Verma and Betti, 2006).

The empirical analysis conducted in the present Chapter is based on the first wave of the Italian sub-sample of the EU-SILC survey conducted in 2004 and which has collected information on income for the reference year 2003. The sample is composed of 61,429 individuals representative of the Italian population.

4.2 Results for Italian Macro-regions, year 2004

Table 12.1 illustrates the maximum likelihood estimates for the Dagum type I model respectively specified in (12.3) and in its heterogeneous version in (12.10). The Dagum type I model is made heterogeneous by considering a regional effect that is represented by the four dummy variables in Table 12.1. They are so defined: ITC (1 if north-western regions), ITD (1 if north-eastern regions), ITF (1 if southern regions), ITG (1 if island regions), the reference category is the central Italian regions. The log-likelihood ratio test statistics (LR) suggest a statistical improvement of the heterogeneous Dagum type I distribution.

With regard to the goodness of the overall fit of the two models we refer to Cox-Snell residuals (Cox and Snell, 1968).[26] The Cox-Snell residuals are shown respectively in Figure 12.1 for the Dagum type I model and in Figure 12.2 for the heterogeneous Dagum type I model. The results indicate that the fit of the Dagum distribution is sufficiently good. In fact, very few observations deviated from the 45-degree line. In fact, only 0.9 percent of deviations is greater than 7 in Figure 12.1

[26] See Quintano and D'Agostino (2006) for a methodological explanation of the use of Cox-Snell residuals in parametric income distribution.

Table 12.1: Results of estimation of parameters of Dagum Type I Model (1,000 euro, 2004)

Dagum type I without individual heterogeneity		
β	δ	λ
0.7436	3.2736	6319.6294
-LogL=211658.271		

Dagum type I with individual heterogeneity	Variables	Estimates	S.E.	p-value
γ_1	Intercept	-0.1684	0.0301	0.0001
	ITC	+0.0853	0.0395	0.0308
	ITD	+0.1952	0.0443	0.0001
	ITF	-0.1805	0.0404	0.0001
	ITG	-0.0179	0.0505	0.7232
γ_2	Intercept	+1.2361	0.0139	0.0001
	ITC	-0.0224	0.0179	0.2122
	ITD	-0.0176	0.0194	0.3664
	ITF	-0.0491	0.0194	0.0115
	ITG	-0.1388	0.0234	0.0001
γ_3	Intercept	+9.2733	0.1685	0.0001
	ITC	-0.0765	0.2175	0.7248
	ITD	-0.1661	0.2369	0.4834
	ITF	-1.3477	0.2271	0.0001
	ITG	-2.2900	0.2516	0.0001
-LogL=207500.495				
Likelihood ratio test (LR) = 8315.552 (p-value = 0.0001)				

and only 1.5 percent is greater than 4 in Figure 12.2, which are both negligible for overall fit. We can therefore conclude that the Dagum type I model can be used as a theoretical model for describing personal income distribution in both versions.

The parameter estimates, presented in Table 12.1 have then been used to calculate the IFR measure for income using the methodology described in Section 3.2 and in Section 3.3. Then the regional effect on poverty measures is studied making a comparison between the traditional indicator HCR and the IFR measure. In Table 12.2, results of the empirical analysis are presented.

Using (12.2), we find that H is equal to 0.1916 with a poverty line[27] z of 7649.3 euros and we estimate α to be equal to 4.8267, whereas using (12.7) we find that $H(z_D)$ is equal to 0.1920 with a poverty line z_D of 7618.9 euros and the estimated α for the IFR measures is equal to 4.8144. The percentage difference between the two poverty lines z and z_D is almost negligible being equal to 0.40.

[27] The poverty line z is also defined as a quantile poverty line as z_D, but in this case ξ_q is the empirical quantile of order $q = 0.5$ of the distribution of equivalised income, i.e. $\xi_q = sup\{y|F(y) \leq q\}$) if $\eta = 0.6$ the poverty line adopted by Eurostat is defined (60% of the median equivalised income).

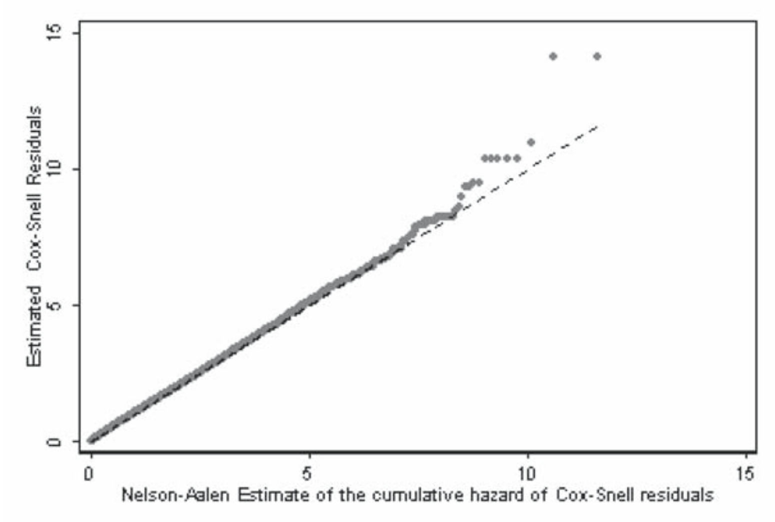


Fig. 12.1: Cox-Snell Residuals of Dagum type I

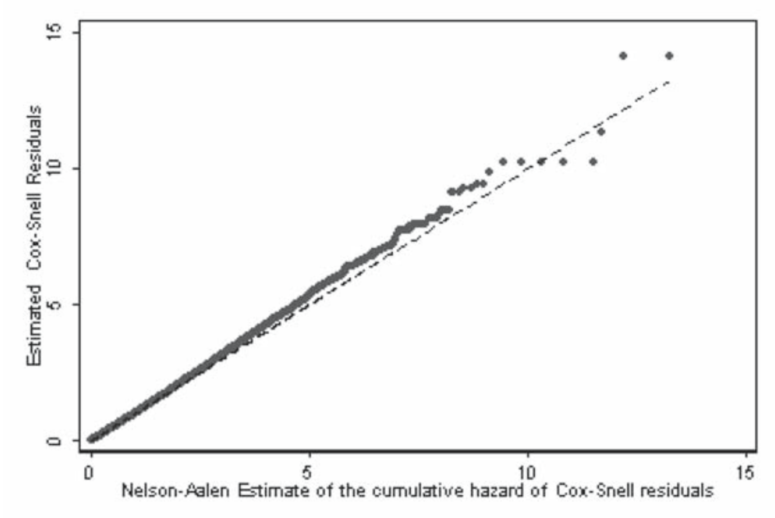


Fig. 12.2: Cox-Snell Residuals of heterogeneous Dagum type I

Table 12.2: Poverty measures among macro-regions

HCR computed on observed data and IFR using estimates from Dagum type I model			
Macro-region	H		FM^D
North-West	0.1031		0.1173
North-East	0.0860		0.1079
Centre	0.1285		0.1434
South	0.3396		0.3139
Islands	0.3621		0.3223
National H	0.1916		
α		4.8267	
z		7649.3	

HCR computed on heterogeneous Dagum type I model and IFR using estimates from Dagum Type I model			
Macro-region	$H(z_D)$		FM^D
North-West	0.1047		0.1176
North-East	0.0904		0.1083
Centre	0.1338		0.1439
South	0.3426		0.3146
Islands	0.3570		0.3229
National $H(z_D)$	0.1920		
α		4.8144	
z_D		7618.9	
$(z\text{-}z_D)$		0.40 %	

Moreover, the differences between H and $H(z_D)$ at national and at macro-regions level are very small. This confirms the excellent goodness of fit of the Dagum type I model. Taking the latter into account the interpretative and economic discussion is based on the results presented at the bottom of Table 12.2 as both the head count ratio and the integrated fuzzy and relative measure are based on the Dagum type I model. As expected, the head count ratio in southern regions and in islands (34.26 percent and 35.70 percent respectively) is much higher than the ones estimated in the other macro-regions. In particular, North-east regions show the lowest value of the head count ratio, only 9.04 percent of individuals are below the poverty line. These empirical results have an obvious interpretation, as it is well known that the socio-economic development in Italy is characterized by a very heavy regional effect that shows the Northern regions at the top of the classification and the Southern and Island regions at the bottom. The same picture is confirmed by the IFR measure. In fact, the degree of membership to the set of poor increases if individuals live in the South of Italy or in the Island regions. For example, people living in North-west regions have a degree of poverty equal to 0.1176 which is much lower than those who live in the Southern regions (0.3146).

However, the differences are generally smoothed with respect to the traditional measure and the difference between North-west and North-east regions is especially much less marked. This last result is also confirmed in Betti and Verma (2008), based on the last wave (2001) of the European Community Households Panel (ECHP).

As a final remark, it is also very interesting to note that the estimated values of FM_D presented in Table 12.2 are also very similar to the FM based on equation (12.1) that, as we explained in Section 2, represent the empirical evaluation of the membership function. In fact, we find FM equal to 0.1119 in North-west regions, FM equal to 0.1015 in North-east regions, FM equal to 0.1389 in Centre regions, FM equal to 0.3232 in Southern regions and finally FM equal to 0.3335 in Island regions, with an estimated α equal to 11.9548.

5 Conclusions

This Chapter showed how the Dagum model can be used in order to fit the empirical cumulative distribution in the IFR approach. The model parameters were also specified as a function of a macro-region variable so that spatial comparisons in poverty analysis can be made. The method was applied to the first wave of the Italian sample of the EU-SILC survey conducted in 2004. Different degrees of poverty were found across macro-regions. North-western regions showed the lowest degree of poverty and the difference with the southern and island regions was remarkably evident.

The goodness of fit of the Dagum distribution was good and the model therefore represented very well the underlying process generating income data. This evidence confirms that the Dagum type I model, also easily implemented in the IFR approach, provides a useful way for discovering the true pattern of the process that generates the degree of poverty. Moreover, the estimated model and the corresponding membership function in 2004 can then be adapted to the following year taking into account the forecasted inflation rate between the two years using the scale parameter λ without having to estimate the model again in 2005. Logically, the potential of the Dagum model in the IFR approach can be further explored by, for example, introducing more than one covariate into the model parameters. Both individual and macroeconomic factors can be considered and the relative changes in the average of the membership function can be explored over time. This could be a good starting point for future research.

References

Aitchison, J. and J. A. C. Brown (1957) *The Lognormal Distribution with Special References to Its Uses in Economics*, Cambridge University Press, Cambridge.

Baliamoune-Lutz, M. (2006) On the Measurement of Human Well-being: Fuzzy-set Theory and Sen's Capability Approach, in M. McGillivray and M. Clarke (eds.) *Understanding Human Well-being*, pp. 119–138, Brooking Institution Press, Washington D.C.

Berenger, V. and A. Verdier-Chouchane (2007) Multidimensional Measures of Well-Being: Standard of Living and Quality of Life Across Countries, *World Development*, **35**, 1259–1276.

Betti, G., B. Cheli and R. Cambini (2004) A Statistical Model for the Dynamics between Two Fuzzy States: Theory and an Application to Poverty Analysis, *Metron*, **62**, 391–411.

Betti, G., B. Cheli, A. Lemmi and V. Verma (2006) Multidimensional and Longitudinal Poverty: an Integrated Fuzzy Approach, in A. Lemmi and G. Betti (eds.) *Fuzzy Set Approach to Multidimensional Poverty Measurement*, pp. 111–137, Springer, New York.

Betti, G. and V. Verma (1999) Measuring the Degree of Poverty in a Dynamic and Comparative Context: A Multi-dimensional Approach Using Fuzzy Set Theory, Proceedings ICCS-VI, 11:289-301, Lahore, Pakistan.

Betti, G. and V. Verma (2008) Fuzzy Measures of the Incidence of Relative Poverty and Deprivation: A Multi-dimensional Perspective, *Statistical Methods and Applications*, forthcoming.

Biewen, M. and S. P. Jenkins (2005) A Framework for the Decomposition of Poverty Differences with an Application to Poverty Differences between Countries, *Empirical Economics*, **30**, 331–358.

Botargues, P. and D. Petrecolla (1999) Estimaciones Paramétricas y no Paramétricas de la Distribución del Ingreso de los Ocupados del Gran Buenos Aires, 1992-1997, *Economica (National University of La Plata)*, **45**.

Burr, I. W. (1942) Cumulative Frequency Functions, *Annals of Mathematical Statistics*, **13**, 215–232.

Cerioli, A. and S. Zani (1990) A Fuzzy Approach to the Measurement of Poverty, in C. Dagum and M. Zenga (eds.) *Income and Wealth Distribution, Inequality and Poverty, Studies in Contemporary Economics*, pp. 272–284, Springer, Berlin.

Cheli, B. (1995a) Totally Fuzzy and Relative Measures in Dynamic Context, *Metron*, **53**, 183–205.

Cheli, B. (1995b) Totally Fuzzy and Relative Measures of Poverty in Dynamic Context. An Application to the British Households Panel Survey, ISER Working Papers, 1995-95-13. Institute for Social and Economic Research, University of Essex, Colchester, UK.

Cheli, B. and G. Betti (1999) Fuzzy Analysis of Poverty Dynamics on an Italian Pseudo Panel, 1985-1994, *Metron*, **57**, 83–103.

Cheli, B. and A. Lemmi (1995) A Totally Fuzzy and Relative Approach to Multidimensional Analysis of Poverty, *Economic Notes*, **24**, 115–134.

Chiappero, M. E. (2000) A Multidimensional Assessment of Well-being Based on Sen's Functioning Approach, *Rivista Internazionale di Scienze Sociali*, **108**, 207–239.

Cox, D. R. and E. J. Snell (1968) A General Definition of Residuals, *Journal of the Royal Statistical Society, Series B (Methodological)*, **30**, 248–275.

Dagum, C. (1973) Un Modèle non Linéaire de Répartition Fonctionnelle du Revenu, *Economie Appliquée*, **26**, 843–876.

Dagum, C. (1975) A Model of Income Distribution and the Conditions of Existence of Moments of Finite Order, Proceedings of the 40th Session of the International Statistical Institute, Vol. XLVI, Book 3, Warsaw, pp 199-205.

Dagum, C. (1977) A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, **30**, 413–436.

Dagum, C. (1980a) Inequality Measures between Income Distributions with Applications, *Econometrica*, **48**, 1791–1803.

Dagum, C. (1980b) The Generation and Distribution of Income, the Lorenz Curve and the Gini Ratio, *Economie Appliquée*, **33**, 327–367.

Dagum, C. (1983) Medida de la Deferencial del Ingreso Entre Familias Blancas, Negras y de Origen Hipanico en los Estados Unidos, *El Trimestre Economico*, **50**(2).

Dagum, C. (1990) A Model of Net Wealth Distribution Specified for Negative, Null and Positive Wealth. A Case Study: Italy, in C. Dagum and M. Zenga (eds.) *Income and Wealth Distribution, Inequality and Poverty*, pp. 42–56, Spinger-Verlag, Berlin, Heidelberg.

Dagum, C. and M. Costa (2004) Analysis and Measurement of Poverty. Univariate and Multivariate Approaches and Their Policy Implications. A Case Study: Italy, in C. Dagum and G. Ferrari (eds.) *Household Behaviour, Equivalence Scales, Welfare and Poverty*, pp. 221–272, Physica-Verlag, Heidelberg.

Dagum, C., R. Gambassi and A. Lemmi (1992) New Approaches to the Measurement of Poverty, in *Poverty Measurement for Economies in Transition in Eastern European Countries, Warsaw: Polish Statistical Association and Central Statistical Office*, pp. 201–225.

Dagum, C. and A. Lemmi (1989) A Contribution to the Analysis of Income Distribution and Income Inequality and a Case Study: Italy, in D. J. Slottje (ed.) *Advances in Econometrics*, pp. 123–157, JAI Press, Greenwich, CT.

Dastrup, S. R., R. Hartshorn and J. B. McDonald (2007) The Impact of Taxes and Transfer Payments on the Distribution of Income: A Parametric Comparison, *Journal of Economic Inequality*, **5**, 353–369.

Dubois, D. and H. Prade (1980) *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York.

Fattorini, L. and A. Lemmi (1979) Proposta di un Modello Alternativo per l'Analisi della Distribuzione Personale del Reddito, atti delle Giornate di Lavoro AIRO, Bologna, pp. 89-117.

Kleiber, C. (1996) Dagum vs. Singh-Maddala Income Distributions, *Economic Letters*, **53**, 265–268.

Lelli, S. (2001) Factor Analysis vs. Fuzzy Sets Theory: Assessing the Influence of Different Techniques on Sen's Functioning Approach, Discussion Paper Series DPS 01.21, Center for Economic Studies, Catholic Univeristy of Leuven, Belgium.

Lemmi, A. and G. Betti (2006) *Fuzzy Set Approach to Multidimensional Poverty Measurement*, Springer, New York.

McGillivray, M. and M. Clarke (2006) *Understanding Human Well-being*, Brooking Institution Press, Washington D.C.

Pudney, S. (1999) On Some Statistical Methods for Modelling the Incidence of Poverty, *Oxford Bulletin of Economics and Statistics*, **61**, 385–408.

Qizilbash, M. (2003) Vague Language and Precise Measurement: The Case of Poverty, *Journal of Economic Methodology*, **10**, 41–58.

Quintano, C. and A. D'Agostino (2006) Studying Inequality in Income Distribution of Single Person Household in Four Developed Countries, *Review of Income and Wealth*, **52**, 525–546.

Rojas, M. (2006) Well-being and the Complexity of Poverty: A Subjective Well-being Approach, in M. McGillivray and M. Clarke (eds.) *Understanding Human Well-being*, pp. 182–206, Brooking Institution Press, Washington D.C.

Ruggeri, L. C., R. Saith and F. Stewart (2006) Does it Matter that We do not Agree on the Definition of Poverty? A Comparison of Four Approaches, in M. McGillivray and M. Clarke (eds.) *Understanding Human Well-being*, pp. 19–53, Brooking Institution Press, Washington D.C.

Salem, A. B. and T. D. Mount (1974) A Convenient Descriptive Model of Income Distribution: The Gamma Density, *Econometrica*, **42**, 1115–1127.

Singh, S. K. and G. S. Maddala (1976) A Function for Size Distribution of Incomes, *Econometrica*, **44**(5), 963–970.

Tadikamalla, P. R. (1980) A Look at the Burr and Related Distributions, *International Statistical Review*, **48**, 337–344.

Verma, V. and G. Betti (2006) EU Statistics on Income and Living Conditions (EU-SILC): Choosing the Survey Structure and Sample Design, *Statistics in Transition*, **7**, 935–970.

Zadeh, L. A. (1965) Fuzzy Sets, *Information and Control*, **8**, 338–353.

CHAPTER 13

Modelling Lorenz Curves: Robust and Semi-parametric Issues

Frank A. Cowell[†] and Maria-Pia Victoria-Feser[‡]

Abstract

Modelling Lorenz curves (LC) for stochastic dominance comparisons is central to the analysis of income distributions. It is conventional to use non-parametric statistics based on empirical income cumulants which are used in the construction of LC and other related second-order dominance criteria. However, although attractive because of its simplicity and its apparent flexibility, this approach suffers from important drawbacks. While no assumptions need to be made regarding the data-generating process (income distribution model), the empirical LC can be very sensitive to data particularities, especially in the upper tail of the distribution. This robustness problem can lead in practice to "wrong" interpretation of dominance orders. A possible remedy for this problem is the use of parametric or semi-parametric models for the data-generating process and robust estimators to obtain parameter estimates. In this paper, we focus on the robust estimation of semi-parametric LC and investigate issues such as sensitivity of LC estimators to data contamination (Cowell and Victoria-Feser, 2002), trimmed LC (Cowell and Victoria-Feser, 2006), and inference for trimmed LC (Cowell and Victoria-Feser, 2003), robust semi-parametric estimation for LC (Cowell and Victoria-Feser, 2007), selection of optimal thresholds for (robust) semi-parametric modelling (Dupuis and Victoria-Feser, 2006), and use both simulations and real data to illustrate these points.

[†] STICERD and Economics Department, London School of Economics and Political Science.

[‡] HEC, Faculty of Economics and Social Sciences, University of Geneva. Partially supported by Swiss National Science Foundation, grant # PP001-106465.

1 Introduction

The Lorenz curve is central to the analysis of income distributions, embodying fundamental intuition about inequality comparisons (Dagum, 1985; Cowell and Victoria-Feser, 2007). Ranking theorems based on Lorenz dominance and the associated concept of stochastic dominance are fundamental to the theoretical welfare economics of distributions. But formal welfare propositions can only be satisfactorily invoked for empirical constructs if sample data can be taken as a reasonable representation of the underlying income distributions under consideration. In practice income-distribution data may be contaminated by recording errors, measurement errors and the like and, if the data cannot be purged of these, welfare conclusions drawn from the data can be seriously misleading. Indeed, it has been formally shown that Lorenz and stochastic dominance results are non-robust (Cowell and Victoria-Feser, 2002). This means that small amounts of data contamination in the wrong place can reverse unambiguous ranking orders: the "wrong place" usually means in the upper tail of the distribution. This is of particular interest in view of a burgeoning recent literature that has focused on empirical issues concerning the upper tail of both income distributions and wealth distributions (Atkinson, 2004; Kopczuk and Saez, 2004; Moriguchi and Saez, 1991; Piketty, 2001; Piketty and Saez, 2003; Saez and Veall, 2005). So it is important to have an approach that enables one to control for the distortionary effect of upper-tail contamination in a systematic fashion. This paper addresses the problem by introducing a robust method of estimating Lorenz curves and implementing stochastic-dominance criteria. To this end we have assembled some recent research on this issue, mainly drawing on the results of Cowell and Victoria-Feser (2006) and Cowell and Victoria-Feser (2007).

Our approach is organized as follows. We begin, in section 2, by setting out the formal background to the Lorenz curve and the estimation problems associated with extreme values. Section 3 develops the semi-parametric approach to modelling Lorenz curves and section 4 discusses the practical problem of parameter choice in implementing the method. Section 5 applies the method to UK data and section 6 concludes.

2 Background

We may set out the formal representation of the Lorenz curve using the following simple framework. Let $\mathfrak{F}$ be the set of all univariate probability distributions and X be a random variable with probability distribution $F \in \mathfrak{F}$ and support $\mathfrak{X} \subseteq \mathbb{R}$. F can be thought of as a parametric model F_θ. We shall write statistics of any distribution $F \in \mathfrak{F}$ as a functional $T(F)$; in particular we write the mean as $\mu(F) := \int x dF(x)$. A key distributional concept derived from F is given by the q^{th} *cumulative functional* $C : \mathfrak{F} \times [0,1] \mapsto \mathfrak{X}$:

$$C(F;q) := \int_{\underline{x}}^{Q(F;q)} x dF(x) = c_q. \tag{13.1}$$

where $\underline{x} := \inf \mathfrak{X}$ and

$$Q(F;q) = \inf\{x|F(x) \geq q\} = x_q \tag{13.2}$$

is the quantile functional. The importance of this concept is considerable in the practical analysis of income distributions: for a given $F \in \mathfrak{F}$, the graph of $C(F,q)$ against q describes the *generalized Lorenz curve* (GLC); normalizing by the mean functional $\mu(F) = C(F,1)$ one has the *Relative Lorenz curve* (RLC) (Lorenz, 1905):

$$L(F;q) := \frac{C(F;q)}{\mu(F)} \tag{13.3}$$

The GLC and RLC are fundamental to a number of theorems drawing welfare-conclusions from income-distribution data and other types of data.

Now consider the problem of estimating Lorenz curves. There are broadly three approaches.

1. *Nonparametric methods.* Cumulative functionals can obviously be estimated by replacing F in (13.1) by the empirical distribution of a sample of incomes $x_1,\ldots,x_n$

$$F^{(n)}(y) = \frac{1}{n}\sum_{i=1}^{n} \iota\,(y \leq x_i)$$

   where $\iota\,(\cdot)$ is the *indicator function*. However, this can lead to misleading conclusions when it comes to comparing distributions in terms of their cumulative functionals when there is data contamination (Cowell and Victoria-Feser, 2002). One way of avoiding the potential bias induced by extreme data in the tails is to rely on the concept of trimmed Lorenz curves: basically, F in (13.1) is replaced by the trimmed distribution $\tilde{F}_\alpha$ given by:

$$\tilde{F}_\alpha(x) := \begin{cases} 0 & \text{if } x < Q(F,\underline{\alpha}) \\ \frac{F(x)-\underline{\alpha}}{1-\alpha} & \text{if } Q(F,\underline{\alpha}) \leq x < Q(F,\overline{\alpha}) \\ 1 & \text{if } x \geq Q(F,\overline{\alpha}) \end{cases}$$

   with $\underline{\alpha} + \overline{\alpha} = \alpha$. Using $\tilde{F}_\alpha$ instead of $F^{(n)}$ amounts to trimming the sample data below $Q(F,\underline{\alpha})$ and above $Q(F,\overline{\alpha})$, and then compute empirical cumulants. The theoretical aspects are handled in Cowell and Victoria-Feser (2006).
2. *Parametric modelling.* Alternatively, one can estimate F using a model (a functional form) such as the one proposed by Dagum (1977).[28] The parameters should obviously be estimated in a robust fashion (see e.g. Victoria-Feser and Ronchetti (1994), Victoria-Feser (1995)), but as has been discussed in Cowell

[28] Other models can be found in Dagum (1980), Dagum (1983) and McDonald (1984) and an excellent overview is provided by Kleiber and Kotz (2003).

and Victoria-Feser (2007), a full parametric estimation forces the data into the mould of a functional form that may not be suitable for comparisons.

3. *Semi-parametric approach.* The problem that a single, tractable functional form may not be appropriate for the data motivates the use of an approach in which the data above a threshold x_0 are (robustly) fitted to a parametric distribution, while the rest of the data are treated nonparametrically. The semi-parametric approach is of particular interest because of its *ad hoc* use in practical treatment of problems associated with the upper tails of distributions. For example a Pareto tail is sometimes fitted to data in cases where data are sparse in order to provide better estimates of upper tail probabilities or higher quantiles.

It is this third estimation method, the semi-parametric approach, that forms the focus of the present paper.

3 Semi-parametric robust estimation of Lorenz curves

If the range of X is bounded below – 0 is a typical value – the problems with contaminated data occur in the upper tail of the distribution (Cowell and Victoria-Feser, 2002). A case can therefore be made for using parametric modelling only in the upper tail and estimating the parameter of the upper-tail model robustly. The rest of the distribution is estimated using the empirical distribution function. If no restriction is imposed on the range of the random variable of interest, then the results below can easily be extended accordingly.

Cowell and Victoria-Feser (2007) proposed an approach which is suitable for any parametric model for the upper tail of the distribution. They however choose a model that is of special relevance empirically, that is the Pareto distribution given by

$$F_\theta(x) = 1 - \left[\frac{x}{x_0}\right]^{-\theta}, \qquad x > x_0 \tag{13.4}$$

with density $f(x;\theta) = \theta x^{-(\theta+1)} x_0^\theta$. The parameter of interest is θ.[29] A semi-parametric approach will combine a non-parametric RLC for say the $100(1-\alpha)\%$ lower incomes and a parametric RLC based on the Pareto distribution for the $100\alpha\%$ upper incomes. Therefore x_0 is determined by the $1-\alpha$ quantile $Q(F;1-\alpha)$ defined in (13.2). The method for a suitable choice of x_0 is given in section 4. The full semi-parametric distribution $\widetilde{F}$ of the income variable X is

$$\widetilde{F}(x) = \begin{cases} F(x) & x \le x_0 \\ F(x_0) + (1 - F(x_0))F_\theta(x) & x > x_0 \end{cases}$$

where F could be in principle any suitable parametric distribution, but in our case will be estimated by the empirical distribution. With $x_0 = Q(F;1-\alpha)$, we have

[29] θ is assumed to be greater than 2 for the variance to exist.

$$\widetilde{F}(x) = \begin{cases} F(x) & x \leq Q(F;1-\alpha) \\ 1-\alpha\left[\frac{x}{Q(F;1-\alpha)}\right]^{-\theta} & = x > Q(F;1-\alpha) \end{cases} . \tag{13.5}$$

For $x > Q(F;1-\alpha)$, the density $\widetilde{f}$ is

$$\widetilde{f}(x;\theta) = \alpha\theta Q(F;1-\alpha)^{\theta} x^{-\theta-1} .$$

In particular

$$\widetilde{f}(x_{1-\alpha};\theta) = \frac{\alpha\theta}{x_{1-\alpha}} . \tag{13.6}$$

The quantile functional is then obtained using (13.5) and is given by

$$Q(\widetilde{F},q) = \begin{cases} Q(F,q) & q \leq 1-\alpha \\ Q(F;1-\alpha)\left(\frac{1-q}{\alpha}\right)^{-1/\theta} & q > 1-\alpha \end{cases}$$

Hence the cumulative income functional defining the semi-parametric GLC becomes

$$C(\widetilde{F};q) = \int_{\underline{x}}^{Q(\widetilde{F};q)} x d\widetilde{F}(x) \tag{13.7}$$

$$= \begin{cases} \int_{\underline{x}}^{Q(F;q)} x dF(x) & q \leq 1-\alpha \\ \int_{\underline{x}}^{Q(F;1-\alpha)} x dF(x) & \\ +\alpha \int_{Q(F;1-\alpha)}^{Q(F;1-\alpha)\left(\frac{1-q}{\alpha}\right)^{-1/\theta}} x dF_{\theta} & q > 1-\alpha \end{cases} \tag{13.8}$$

$$= \begin{cases} \int_{\underline{x}}^{Q(F;q)} x dF(x) & q \leq 1-\alpha \\ \int_{\underline{x}}^{Q(F;1-\alpha)} x dF(x) & \\ +\alpha \frac{\theta}{1-\theta} Q(F;1-\alpha)\left[\left(\frac{1-q}{\alpha}\right)^{\frac{\theta-1}{\theta}} - 1\right] & q > 1-\alpha \end{cases} \tag{13.9}$$

where $\underline{x} := \inf \mathfrak{X}$. The mean of the semi-parametric distribution is given by:

$$\begin{aligned} C(\widetilde{F};1) &= \int_{\underline{x}}^{Q(F;1-\alpha)} x dF(x) - \alpha Q(F;1-\alpha)\frac{\theta}{1-\theta} \\ &= c_{1-\alpha} - \alpha x_{1-\alpha}\frac{\theta}{1-\theta} \\ &= \mu(\widetilde{F}) \end{aligned} \tag{13.10}$$

The semi-parametric RLC is simply

$$L(\widetilde{F};q) = \frac{C(\widetilde{F};q))}{\mu(\widetilde{F})} \tag{13.11}$$

The cumulative income function (13.9) obviously needs to be estimated. The (unknown) distribution F is replaced by the empirical distribution $F^{(n)}$ and an estimate for α will be discussed in Section 4. To estimate the Pareto model, hence θ, for the upper tail of the distribution, one can use the maximum likelihood estimator (MLE). Unfortunately, the MLE for the Pareto model is known to be very sensitive to data contamination (Victoria-Feser and Ronchetti, 1994). This is also the case for other models such as Dagum (1977) model (see Victoria-Feser (1995)). Cowell and Victoria-Feser (2007) propose using a robust estimator in the class of M-estimators (Huber, 1981). For a sample of k observations x_i, a general M-estimator is defined as the solution in θ of

$$\frac{1}{n}\sum_{i=1}^{k} \psi(x_i;\theta) = 0$$

with some (mild) conditions on the function ψ. This function is chosen so that the resulting estimator is consistent at the model F_θ and also that it is robust to slight model deviations (for a discussion, see e.g. Hampel *et al.* (1986)). The latter condition is satisfied if the ψ-function is bounded, which is the case for so-called weighted MLE (WMLE), i.e.

$$\frac{1}{n}\sum_{i=1}^{k} w(x_i;\theta)\,[s(x_i;\theta) - a(\theta)] = 0 \tag{13.12}$$

where $w(x;\theta)$ is a weight function with value in $[0,1]$ insuring the robustness of the estimator, $s(x;\theta) = \partial \log f(x;\theta)/\partial\theta$ is the score function and $a(\theta)$ is a consistency correction factor[30]. Cowell and Victoria-Feser (2007) choose the *optimal B-robust estimators* (OBRE) (Hampel *et al.*, 1986), a robust estimator with minimal asymptotic covariance matrix (see e.g. Cowell and Victoria-Feser (2007) for details).

The resulting semi-parametric GLC (and RLC) estimates are hence robust to data contamination. They are based on the Pareto model for the upper tail and robustness is sought against deviations from the Pareto model. If the Pareto model is believed not to be suitable, then it can be changed (for example, to a generalized version of it) but the method remains the same. Cowell and Victoria-Feser (2007) also provide the asymptotic covariances of the estimators for inference with semi-parametric GLC (and RLC) which can be used for robust welfare comparison.

In section 5 an example will illustrate the performance of robust semi-parametric estimators of RLC and GLC.

[30] The correction factor does not need to be estimated simultaneously, see below.

4 Choosing α

The choice of the proportion α of data in the upper tail to be fitted to the Pareto model, or equivalently the threshold x_0 above which the data are fitted to a Pareto model, is not a problem specific to income distribution analysis. It has attracted and still attracts the attention of researchers in domains such as finance, insurance, engineering, or environmental sciences. This problem falls within the general heading of extreme value distributions (for a general reference, see e.g. Embrechts *et al.* (1997)). To estimate the threshold, a compromise should be sought between bias and variance: choosing a threshold too close to the central data will cause bias in the Pareto model estimator since only the tail can be assumed to be Pareto distributed, and selecting too extreme a threshold will yield large variances for the estimator since it will be based on a small sample. A common practice is to use the Pareto quantile plot (see e.g. Beirlant *et al.* (1996)). Indeed, rearranging (13.4) one gets

$$\log\left(\frac{x}{x_0}\right) = -\frac{1}{\theta}\log\left(1 - F_\theta(x)\right), \quad x > x_0 \tag{13.13}$$

showing that there is a linear relationship between the log of the $x > x_0$ and the log of the survival function. This relationship was actually found empirically by Pareto (1896) and led him to the construction of his model (see also Dagum (1983)). Let $x^*_{[i]}$, $i = 1,\ldots,k$, be the ordered largest k observations, so that $x^*_{[i]} = Q(F^{*(n)}; i/(k+1))$, with $F^{*(n)}$ the empirical distribution of $x^*_{[i]}$. The empirical counterpart of (13.13) is the Pareto quantile plot

$$\log\left(\frac{Q(F^{*(n)}; i/(k+1))}{x_0}\right) = -\frac{1}{\theta}\log\left(\frac{k+1-i}{k+1}\right), \quad i = 1,\ldots,k. \tag{13.14}$$

Therefore, given a sample of n income data $x_i, 1,\ldots,n$ and by letting $x_{[i]}$ denote the ith order statistic, the plot of $\log\left(x_{[i]}\right)$ versus $-\log\left((n+1-i)/(n+1)\right)$, $i = 1,\ldots,n$ is the Pareto quantile plot that is used to detect graphically the quantile $x_{[i]}$ above which the Pareto relationship is valid, i.e. the point above which the plot yields a straight line. We note that there is a clear relationship between x_0 and k in that $k = \sum_{i=1}^{n} \iota(x_{[i]} \geq x_0)$.

More formally, a general approach in determining k is the minimization of an estimate of the asymptotic mean squared error (AMSE) of the estimator of θ. If a classical estimator such as the MLE is chosen, then the determination of k can be influenced by extreme data in the upper tail (see Dupuis and Victoria-Feser (2006)). Note that here extreme is used relatively to the Pareto model: if it is assumed to fit the upper tail, then extreme data represent deviations for this assumption that can appear in the Pareto quantile plot as data that do not fit the straight line.

In order to choose k, or equivalently x_0 in a robust fashion, Dupuis and Victoria-Feser (2006) use another criterion, namely a prediction error criterion that is estimated robustly (see also Ronchetti and Staudte (1994)), named the *RC*-criterion.

Let $Y_i = \log\left(x^*_{[i]}/x_0\right), i = 1,\ldots,k$, $\hat{Y}_i = -1/\hat{\theta}\log\left[(k+1-i)/(k+1)\right], i = 1,\ldots,k$ where $\hat{\theta}$ is an estimator of θ, and

$$\widehat{\sigma}_i^2 = \text{var}(Y_i) = \sum_{j=1}^{i} \frac{1}{\widehat{\theta}^2(k-i+j)^2}$$

the (estimated) *RC*-criterion is given by

$$C_R(x_0) = \frac{1}{k}\sum_{i=1}^{k}\hat{w}_i^2\left(\frac{Y_i-\hat{Y}_i}{\hat{\sigma}_i}\right)^2 + \frac{2}{k}\sum_{i=1}^{k}\frac{1}{\hat{\sigma}_i^2}\text{cov}\left[\hat{w}_iY_i, \hat{w}_i\hat{Y}_i\right] - \frac{1}{k}\sum_{i=1}^{k}\frac{1}{\hat{\sigma}_i^2}\text{var}\left[\hat{w}_iY_i\right] \tag{13.15}$$

where each $\hat{w}_i, 0 \leq \hat{w}_i \leq 1$, is the fitted weight of the i^{th} observation, provided by a robust fit of the Pareto model, using a WMLE given in (13.12). For suitable estimates of $\text{cov}\left[\hat{w}_iY_i, \hat{w}_i\hat{Y}_i\right]$ and $\text{var}[\hat{w}_iY_i]$, see Dupuis and Victoria-Feser (2006). The effect of extreme observations on the calculation of $C_R(x_0)$ is controlled by the weights $\hat{w}_i$. The criterion is minimized over possible values for x_0. Obviously, at the minimum, we have that $Y_i \approx \hat{Y}_i$, hence $\log\left(x^*_{[i]}/x_0\right) \approx -1/\hat{\theta}\log\left[(k+1-i)/(k+1)\right]$.

For the choice of the WMLE, Dupuis and Victoria-Feser (2006) propose an estimator which downweights observations that are "far" from the Pareto model in terms of the size of the residuals with respect to the Pareto regression model, i.e.

$$w(x^*_{[i]};\theta) = \begin{cases} 1 & \text{if } |r_i| \leq c \\ c/|r_i| & \text{if } |r_i| > c \end{cases} \tag{13.16}$$

with $r_i = (Y_i - \hat{Y}_i)/\sigma_i$ and c is a constant regulating the amount of robustness (for more details, see Dupuis and Victoria-Feser (2006)).

In the following section, an empirical example will illustrate the method.

5 Data analysis

Let us put the semi-parametric method into practice using a typical income distribution. The data for our illustration are for household disposable incomes in the UK, 1981 ($n = 7470$)[31].

A Pareto quantile plot of the data together with fitted regression lines are given in Figure 13.1 (see also Dupuis and Victoria-Feser (2006)). The fits are provided by WMLE estimates with residual weights (13.16) for two values of c as well as the classical MLE. The optimal values for x_0 are obtained using $C_R(x_0)$ in which the weights $\hat{w}_i$ and $\hat{Y}_i$ are obtained using the different estimators. For the MLE, $\hat{w}_i = 1, \forall i$. The fit for the MLE (and hence the corresponding optimal value for

[31] The data set is Households Below Average Income which, despite its name, actually provides a representative sample of households over the whole income range – see Department of Social Security (1992) for details.

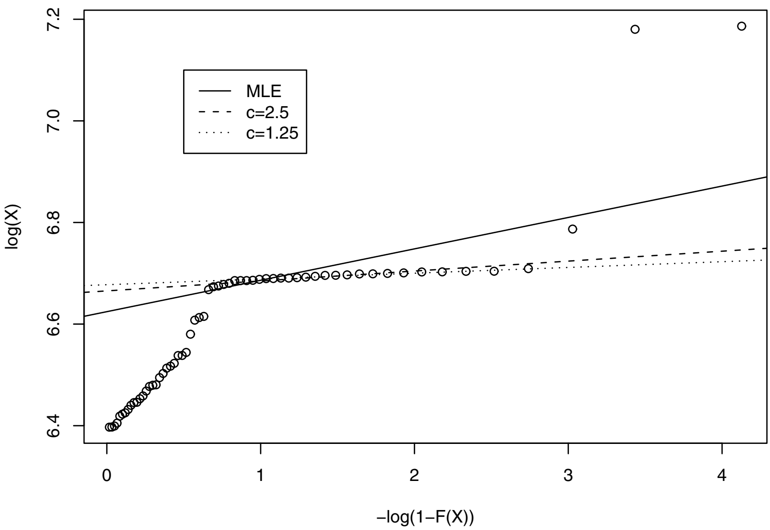


Fig. 13.1: Pareto regression plot. Fitted regression line based on classical and robust \$RC-\$ criteria added. Only incomes above 600 are shown for clarity.

x_0) are not adequate, probably because of a few very extreme observations. Both robust fits seem on the other hand appropriate. For the latter, the optimal value of x_0 corresponds to $k = 22$ selected upper incomes ($k = 32$ for the MLE). Figure 13.2 (see also Dupuis and Victoria-Feser (2006)) shows observations above the robustly selected threshold $x_0 = 803.3$ and arrows indicate the downweighted observations. The striking feature is that not only the largest observations are downweighted, but also the smallest.

To estimate the Pareto parameter, we hence choose $k = 22$. The value for the MLE is $\hat{\theta} = 17.5$ (with standard error 3.73) and the one for the OBRE with $c = 2$[32] is $\hat{\theta} = 76.65$ (17.62). We use these two estimates to build estimated RLC (see (13.9) and (13.10)). These curves (corresponding to the 0.5% top incomes) are presented in Figure 13.3 together with the empirical RLC estimate. Even if it is small, one can see a difference between the three estimates, in that the MLE follows the empirical RLC up to roughly the 0.1% of the top distribution, while the OBRE leads to an estimated RLC showing less inequality on the entire 0.5% top range.

[32] One can note that a different robust estimator is used to estimate the Pareto parameter. For the choice of k a WMLE based on residual weights is a reasonable choice, whereas the more efficient robust estimator (OBRE) for the Pareto parameter given a value for k is also a reasonable choice.

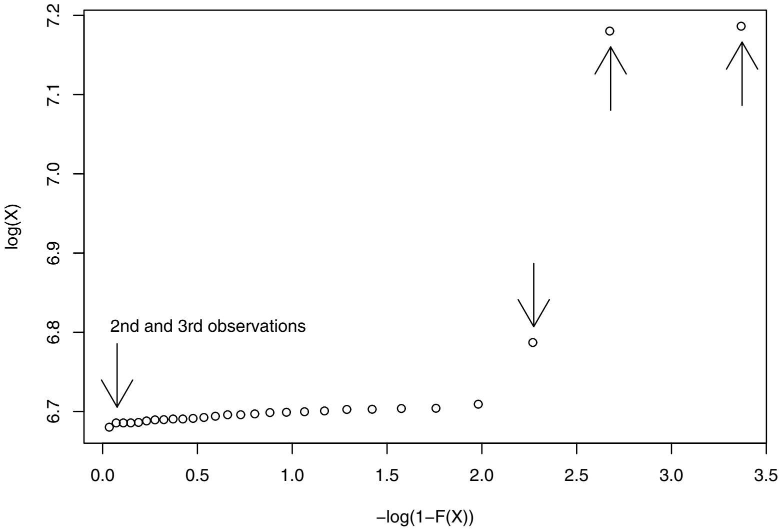


Fig. 13.2: Pareto quantile plot of income data above robustly chosen threshold. Downweighted observations (with WMLE, $c = 1.25$) are identified.

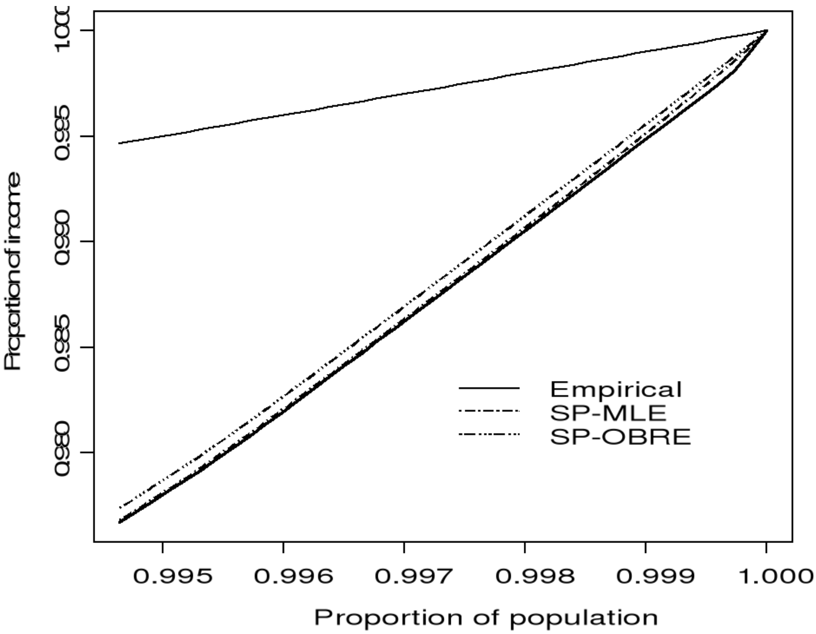


Fig. 13.3: RLC (top 0.5%) estimates (empirical and semi-parametric with MLE and OBRE with $c = 2$) of the UK income data

6 Conclusion

Using ranking criteria to compare distributions is of immense theoretical advantage and practical convenience. In welfare economics they provide a connection between the philosophical basis of welfare judgments and elementary statistical tools for describing distributions. In practical applications they suggest useful ways in which simple computational procedures may be used to draw inferences from collections of empirical distributions. However, since it has been shown that second order rankings are not robust to data contamination, especially in the upper tail of the distribution, it is important to provide the empirical researcher with computational devices which can be used to draw inferences about the properties of distributional comparisons in a robust fashion.

One way forward might be to estimate Lorenz curves through an appropriately specified parametric model and to estimate the model parameters robustly. However, this approach is too restrictive because tractable parametric models are unlikely to be sufficiently flexible to capture some of the essential nuances of Lorenz comparisons. For example, in order for Lorenz curves to be able to cross, a parametric model would usually need to incorporate at least three parameters, which itself may lead to serious estimation complications.

The method proposed here is a semi-parametric approach in that the upper tail of the distribution is robustly fitted using the Pareto model and a semi-parametric Lorenz curve is then built which combines non-parametric cumulative functionals and estimated ones. Simulated examples have proved not only that a few extreme data can reverse the ranking order, but also that the robust parametric Lorenz curve restores the initial ordering. Inference can be made for comparing two distributions even in the semi-parametric setting, by extending the general setting provided in Cowell and Victoria-Feser (2007). For variances too, a robust approach provides reasonable estimates when there is contamination. Note however, that inference has been developed for a fixed value for the proportion α of data in the upper tail, and when it is estimated as is done in this paper, inference that takes into account the variability of an estimator of α is still an open question.

Finally note that although we took the Pareto distribution as a suitable parametric model for the upper tail, and although we considered the (most common) case of a range of definition for the variable bounded below, our results can be extended to other models and to two-tail modelling in a relatively straightforward manner.

References

Atkinson, A. B. (2004) Income tax and Top Incomes over the Twentieth Century, *Hacienda Pública Española*, **168**, 123–141.

Beirlant, J., P. Vynckier and J. L. Teugels (1996) Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics, *Journal of the American Statistical Association*, **91**, 1651–1667.

Cowell, F. A. (2007) Inequality: Measurement, in L. Blume and S. Durlauf (eds.) *The New Palgrave*, Palgrave Macmillan, Basingstoke, Hampshire, UK.

Cowell, F. A. and M.-P. Victoria-Feser (2002) Welfare Rankings in the Presence of Contaminated Data, *Econometrica*, **70**, 1221–1233.

Cowell, F. A. and M.-P. Victoria-Feser (2003) Distribution-Free Inference for Welfare Indicies under Complete and Incomplete Information, *Journal of Economic Inequality*, **1**, 191–219.

Cowell, F. A. and M.-P. Victoria-Feser (2006) Distributional Dominance with Trimmed Data, *Journal of Business & Economic Statistics*, **24**, 291–300.

Cowell, F. A. and M.-P. Victoria-Feser (2007) Robust Stochastic Dominance: A Semi-Parametric Approach, *Journal of Economic Inequality*, **5**, 21–37.

Dagum, C. (1977) A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, **30**, 413–436.

Dagum, C. (1980) Generating Systems and Properties of Income Distribution Models, *Metron*, **38**, 3–26.

Dagum, C. (1983) Income Distribution Models, in D. L. Banks, C. B. Read and S. Kotz (eds.) *Encyclopedia of Statistical Sciences, Vol 4*, pp. 27–34.

Dagum, C. (1985) Lorenz Curve, in D. L. Banks, C. B. Read and S. Kotz (eds.) *Encyclopedia of Statistical Sciences, Vol 5*, pp. 156–161.

Department of Social Security (1992) *Households Below Average Income: A Statistical Analysis, 1979-1988/9*, London: HMSO.

Dupuis, D. J. and M.-P. Victoria-Feser (2006) A Robust Prediction Error Criterion for Pareto Modeling of Upper Tails, *Canadian Journal of Statistics*, **34**, 639–658.

Embrechts, P., C. Kl*ü*pperberg and T. Mikosch (1997) *Modelling Extremal Events. Applications of Mathematics: Stochastic Modelling and Applied Probability*, Springer-Verlag, Berlin, Heidelberg.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel (1986) *Robust Statistics: the Approach Based on Influence Functions*, John Wiley, New York.

Huber, P. J. (1981) *Robust statistics*, John Wiley, New York.

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

Kopczuk, W. and E. Saez (2004) Top Wealth Shares in the United States, 1916-2000: Evidence from Estate Tax Returns, *National Tax Journal*, **57**, 445–487.

Lorenz, M. O. (1905) Methods of Measuring the Concentration of Wealth, *Quarterly Publications of the American Statistical Association*, **9 (New Series, No. 70)**, 209–219.

McDonald, J. B. (1984) Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52**, 647–663.

Moriguchi, C. and E. Saez (1991) The Evolution of Income Concentration in Japan, 1886-2002: Evidence from Income Tax Statistics, NBER Working Paper 12558, National Bureau of Economic Research, Cambridge, Massachusetts.

Pareto, V. (1896) La Courbe de la Répartition de la Richesse, Reprinted 1965 in G. Busoni (ed.): Œuvres Complètes de Vilfredo Pareto, Tome 3: Ecrits sur la Courbe de la Répartition de la Richesse, Geneva: Librairie Droz. English translation in *Rivista di Politica Economica*, 87 (1997), 645-700.

Piketty, T. (2001) *Les Hauts Revenus en France au 20eme Siècle - Inégalités et Redistributions, 1901-1998*, Editions Grasset, Paris.

Piketty, T. and E. Saez (2003) Income Inequality in the United States, *Quarterly Journal of Economics*, **118**, 1–39.

Ronchetti, E. and R. G. Staudte (1994) A Robust Version of Mallow's C_p, *Journal of the American Statistical Association*, **89**, 550–559.

Saez, E. and M. Veall (2005) The Evolution of High Incomes in Northen America: Lessons from Canadian Evidence, *American Economic Review*, **95**, 831–849.

Victoria-Feser, M.-P. (1995) Robust Methods for Personal Income Distribution Models with Application to Dagum's Model, in C. Dagum and A. Lemmi (eds.) *Research on Economic Inequality, Volume 6: Income Distribution, Social Welfare,Inequality and Poverty*, pp. 225–239, JAI Press, Greenwich.

Victoria-Feser, M.-P. and E. Ronchetti (1994) Robust Methods for Personal Income Distribution Models, *The Canadian Journal of Statistics*, **22**, 247–258.

CHAPTER 14

Modelling Inequality with a Single Parameter*

J. M. Henle,† N. J. Horton,‡ and S. J. Jakus§

Abstract

We argue that the Lorenz curve for income is well-modelled by members of the one-parameter family of functions:

$$\{y = (1-(1-r)^k)^{\frac{1}{k}}\}.$$

We justify this statement with data from the Luxembourg Income Study. The family of curves arises from a dynamic model of income growth, in which the parameter k has a direct economic interpretation.

1 Introduction

The unequal distribution of a resource is captured in all its variety by the Lorenz curve which charts, given the rank r ($0 \leq r \leq 1$) of an entity (based on the entity's level of the resource in a given population), the proportion $L(r)$ of the resource belonging to all those of lower rank. For most of this paper, the entity will be the family and the resource will be income.

In theory, the Lorenz curve, and hence inequality in a society, is a multifaceted phenomenon. The curve is subject only to the constraints that it pass through $(0,0)$ and $(1,1)$ and that its derivative be non-decreasing. In practice, however, real Lorenz

* The authors would like to thank Peter Lambert for his generous advice and encouragement and the editor and referees for many helpful suggestions.

† Department of Mathematics and Statistics, Smith College, USA, email: jhenle@math.smith.edu. The first author would like to dedicate his contribution to this paper to his father, the economist Peter Henle.

‡ Department of Mathematics and Statistics, Smith College, USA.

§ Department of Mathematics and Statistics, Smith College, USA.

curves appear to follow a very distinct pattern and in nearly every case the Lorenz curve is well-modelled by a member of a one-parameter family of curves, the Lamé curves of the form:

$$\{y = (1-(1-r)^k)^{\frac{1}{k}}\}.$$

In section 1 we introduce our family of curves and use it to model Lorenz curves for a number of countries and years, chiefly for income data. In section 2 we develop two economic models based on "trickle-up" theories. Both yield Lamé curves. In section 3 we consider a number of reality checks on our model and its consequences. In section 4 we pose a few questions. For the most part, we will restrict our attention to inequality of income. The Luxembourg Income Study (LIS) provides excellent income data for many countries and many years. Data on the distribution of wealth are less reliable or comparable.

We will use r to denote the rank ($0 \leq r \leq 1$) of a family in terms of its income and $I(r)$ for the income of a family at rank r. We will use N for the number of families and A for the aggregate income of all families (i.e. $A = N\int_0^1 I(x)\,dx$). The Lorenz curve, $L(r)$, is the fraction of income earned by families of rank $\leq r$, that is, $L(r) = \frac{\int_0^r I(x)\,dx}{\int_0^1 I(x)\,dx} = \frac{N}{A}\int_0^r I(x)\,dx$. Alternatively, we can write: $L'(r) = \frac{I(r)}{\frac{A}{N}}$, that is, the slope of the Lorenz curve at every rank is equal to the ratio of the income of a family at that rank to the mean income for all families.[33] For background on this and on inequality in general, see Lambert (2001).

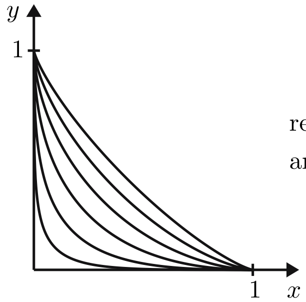
2 Modelling the Lorenz Curve

The problem of modelling the Lorenz curve has a history going back at least 40 years. Early models range from simple, $L = 1-(1-r)^k$ (Quandt, 1966), $L = re^{-k(1-r)}$ (Kakwani and Podder, 1973), $L = \frac{e^{kr}-1}{e^k-1}$ (Chotikapanich, 1993) to quite elaborate $(1-(1-r)^j)^k$ (Rasche *et al.*, 1980) [RGKO], $r^l(1-(1-r)^j)^k$ (Sarabia *et al.*, 1999), $\frac{1}{\sqrt{2}}(L+r) = k\left(\frac{1}{\sqrt{2}}(L-r)\right)^j\left(\sqrt{2}-\frac{1}{\sqrt{2}}(L-r)\right)^l$ (Kakwani and Podder, 1976).

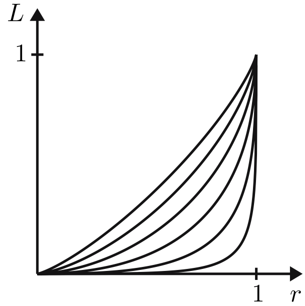
Our curves are a special case of [RGKO], though we arrived at them from a different direction, as a special case, $x^k+y^k=1$, of the Lamé curves, $\left(\frac{x}{a}\right)^k+\left(\frac{y}{b}\right)^k = 1$. These were studied by the Danish engineer and designer Piet Hein and have been called, when $k>1$, "superellipses."

[33] In particular, $L'(.5)$ is equal to the ratio of median income to mean income.

Our family of functions results from taking $a = b = 1$ and $k < 1$,

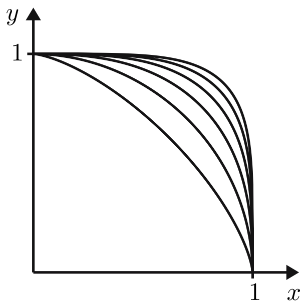


replacing x with $1 - r$
and replacing y with L,

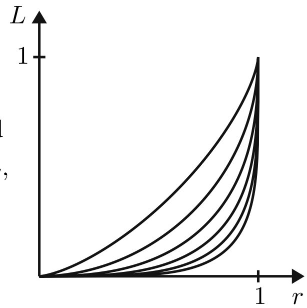


giving us $L(r) = (1-(1-r)^k)^{1/k}$.

A second family can be formed by taking $a = b = 1$ and $k > 1$,



replacing x with r and
replacing y with $1 - L$,



giving us $L(r) = 1-(1-r^k)^{1/k}$.

This second family has basically the same properties as the first. For this reason and since it appears to fit the data no better and no worse, we will concentrate on the first family.

The curves satisfy the four conditions for representing a Lorenz curve as stated in Kakwani and Podder (1973): (a) $L(0) = 0$, (b) $L(1) = 1$, (c) $L(r) \leq r$, and (d) $L', L'' > 0$. It is clear that $L = (1-(1-r^k))^{1/k}$, $0 < k < 1$, satisfies (a) and (b). For (d), if we differentiate $L^k + (1-r)^k = 1$ implicitly, we find that $L' = \left(\frac{1-r}{L}\right)^{k-1}$ is positive, since $r < 1$. Differentiating this yields $L'' = (1-k)\left(\frac{1-r}{L}\right)^{k-2} \cdot \frac{L+(1-r)L'}{L^2}$, also positive, since $k < 1$. Condition (c) follows from the other three.

The family $L(r) = (1-(1-r)^k)^{1/k}$ is easily seen to be symmetric about the line $L = 1-r$, that is, whenever a point (a,b) is on a curve, so is the point $(1-b, 1-a)$. Members of the family do not intersect (see section 4A).

The Gini coefficient of $L = (1-(1-r)^k)^{1/k}$ is not easily computable. Following [RGKO], it is $1 - \frac{2}{k}B(\frac{1}{k}, \frac{1}{k}+1)$, where B is the Beta distribution.

An expression for the income distribution, since it is a multiple of the reciprocal of the second derivative of L, is relatively easy to compute. We have

$$\left(\frac{d^2L}{dr^2}\right)^{-1} = \frac{1}{1-k}(1-r)^{2-k}(1-(1-r)^k)^{2-1/k}.$$

As mentioned earlier, our family is a special case of the two parameter family proposed in [RGKO], $(1-(1-r)^j)^k$. Necessarily, because of the additional parameter, they achieve a better fit. McDonald (1984) catalogued a hierarchy of probability models (ranging from one to four parameters) for the size distribution of income. We are struck, however, by how well real Lorenz curves can be modelled without additional degrees of freedom. In addition, for many situations where data are limited (i.e. estimates are available only at the decile level), it is less clear that the additional flexibility introduced with more than one parameter is worth increased complexity of interpretation.

We tested our family of Lorenz curves, $L(r) = (1-(1-r)^k)^{1/k}$, on 89 sets of data from LIS. Each set consisted of decile data for a country and year, specifically, Austria (4 years), Australia (4 years), Belgium (4 years), Canada (8 years), Denmark (4 years), Finland (4 years), France (4 years), Germany (8 years), Ireland (4 years), Israel (4 Years), Italy (3 years), Mexico (6 years), Netherlands (4 years), Norway (4 years), Taiwan (4 years), Sweden (7 years), United Kingdom (8 years), and the United States (5 years). We also took two sets of data for the United States from Ryu and Slottje (1996) which in addition to decile points included values at $r = .91$, $.92, \ldots .99$. The years considered ranged from 1967 to 2000; median: 1991. The nonlinear least squares regression function `nl` in *Stata* version 9.1 was used for estimation.

We compared three measures of goodness of fit for each of the deciles within each observation (country/year). These included the root mean square error (square root of the average square of the residual), the mean absolute deviation of the observed and predicted value within a country/year as well as the maximum absolute deviation within a country/year.

The results for the 91 observations are impressive:

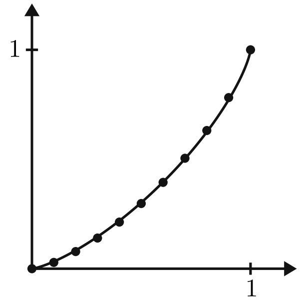
Variable	Mean	Std. Dev.	Min.	Max.
root mean square error (MSE)	.0043318	.0032447	.0004838	.0200169
mean absolute deviation	.0034550	.0026484	.0003716	.0153033
maximum absolute deviation	.0074414	.0060083	.0007958	.0446752

The average root mean square error for the models overall was 0.0043. The maximum absolute deviation of the predicted value from any observed value was 0.045, and the largest MSE for any country/year combination was 0.020 (Italy in 1991 for both).[34] The largest MSE for any other country/year combination (not including US Sarabia) was 0.010 (US 1991) with corresponding maximum absolute deviation of 0.018.

35.2% of the models for country/year combinations yielded a max absolute deviation of less than 0.005; 80.2% were always within 0.01 of the observed value.

The proportion of variance accounted for by the single parameter model was quite high (all R^2 values ≥ 0.998). While the addition of a second parameter may lead to a statistically significant better fit, it is less clear whether this is of practical significance.

The figure below shows a typical example, LIS data for Canada in 1997, together with the graph of $y = (1-(1-r)^{.752})^{1/.752}$.

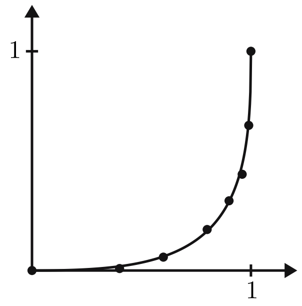


r	actual L	predicted L
0.1	0.028702916	0.032586514
0.2	0.07829257	0.083439991
0.3	0.139985671	0.145979888
0.4	0.212949199	0.218817948
0.5	0.297715202	0.301811658
0.6	0.394426981	0.395656756
0.7	0.504552741	0.502055849
0.8	0.631282881	0.624556416
0.9	0.781668058	0.771777539

The results from the one parameter model explain 99.99% of the variability, with mean absolute deviation of 0.0046 and maximal deviation of 0.0098. A plot of the residuals indicates that while these deviations are of relatively small magnitude, the primary lack of fit is due to the symmetry assumption of the one-parameter model. The two parameter model of Sarabia and colleagues provides an even better fit (mean absolute deviation of 0.0003 and maximal deviation of 0.0008) but at the cost of potentially overfitting the data, and with less readily interpretable parameters.

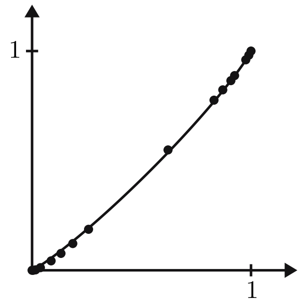
Wealth, which is more unequally distributed than income, was also well modelled by members of this family of curves. Shown below is wealth data for the United States in 1983 from Wolff (2000) with $y = (1-(1-r)^{.417})^{1/.417}$.

[34] The numbers for Italy in 1991 are suspect. The first decile is negative (-.01), the only example in all the data from LIS. In addition, the first decile was positive five years earlier (+.03) and positive again four years later (+.02).



r	actual L	predicted L
0.4	0.009	0.019077141
0.6	0.061	0.063880125
0.8	0.187	0.179748417
0.9	0.318	0.314327595
0.96	0.439	0.483775604
0.99	0.662	0.683838571

Education, which is more equally distributed than income, again fits the pattern. The figure below shows data for years of educational attainment among U.S. citizens 15 years and older, modelled by $y = (1-(1-r)^{.8862})^{1/.8862}$. The data, from the U.S.



r	actual L	predicted L
0.016	0.002	0.008217649
0.038	0.012	0.021841201
0.087	0.043	0.055800117
0.132	0.077	0.089597842
0.186	0.122	0.13245114
0.258	0.187	0.192658887
0.621	0.549	0.537404018
0.831	0.776	0.769843507
0.871	0.823	0.818250894
0.908	0.865	0.864894806
0.925	0.888	0.887113172
0.976	0.96	0.958697528
0.99	0.981	0.980963166

Census Bureau (1998), was broken down into enough categories to yield 16 points on the Lorenz curve (see Appendix A).

3 Modelling the Redistribution of Wealth

One justification for our family of Lorenz curves, $L(r) = (1-(1-r)^k)^{1/k}$, is its success in matching real Lorenz curves. We have a second justification. The family of curves is the solution to a simple dynamic model of income growth which we present here.

We start by viewing I and L as functions of two variables, rank and time. We imagine income rising or falling for each family, which in turn affects the Lorenz curve.

We also adopt a sort of "trickle-up" theory. This theory posits that families earn money off families of lower rank. The higher the rank of a family, the faster its income will grow. In other words, we assume that $\frac{\partial I}{\partial t}$ is related to

$$\frac{\frac{A}{N}L}{N(1-r)} = \frac{AL}{N^2(1-r)}.$$

$\frac{A}{N}L$ is the aggregate income of families of lower rank. $N(1-r)$ is the number of families of higher rank (a family at rank r must share development rights on poorer families with all richer families).

How is $\frac{\partial I}{\partial t}$ related to $\frac{\frac{A}{N}L}{N(1-r)} = \frac{AL}{N^2(1-r)}$? It seems reasonable to assume that $\frac{AL}{N^2(1-r)}$, and $\frac{\partial I}{\partial t}$ are simultaneously zero or simultaneously non-zero. Thus we can allow for considerable possibilities by assuming that their logarithms satisfy a linear equation. This leads to

$$\log\left(\frac{\partial I}{\partial t}\right) = B\log\left(\frac{AL}{N^2(1-r)}\right) + C,$$

or,

$$\frac{\partial I}{\partial t} = e^C\left(\frac{AL}{N^2(1-r)}\right)^B,$$

for some constants B and C.

Now, for a small interval of time Δt, we have $\Delta I = e^C\left(\frac{AL}{N^2(1-r)}\right)^B \Delta t$. For a fixed rank r, we have:

$$L + \Delta L = \frac{\int_0^r (I+\Delta I)\,dx}{\int_0^1 (I+\Delta I)\,dx} = \frac{\frac{A}{N}L + \int_0^r e^C\left(\frac{AL}{N^2(1-x)}\right)^B \Delta t\,dx}{\frac{A}{N} + \int_0^1 e^C\left(\frac{AL}{N^2(1-x)}\right)^B \Delta t\,dx}.$$

We are interested in shape of L in the steady-state, that is, when $\Delta L = 0$. This reduces the equation to:

$$L\int_0^1 \left(\frac{L}{1-x}\right)^B dx = \int_0^r \left(\frac{L}{1-x}\right)^B dx.$$

The integral, $\int_0^1 \left(\frac{L}{1-x}\right)^B dx$ is a constant; we will call it H. Taking the derivative of both sides with respect to r, we have:

$$H\frac{dL}{dr} = \left(\frac{L}{1-r}\right)^B$$

We can solve this equation by separation of variables:

$$H\int L^{-B}dL = \int(1-r)^{-B}dr$$

$$\tfrac{H}{1-B}L^{1-B} = \tfrac{-1}{1-B}(1-r)^{1-B} + F$$

$$HL^{1-B} = -(1-r)^{1-B} + F(1-B).$$

If we relabel $k = 1-B$, this simplifies to

$$HL^k + (1-r)^k = Fk.$$

In practice, $k > 0$. Substituting the points $r = 1$, $L = 1$ and $r = 0$, $L = 0$, gives us that $H = 1$ and $F = \frac{1}{k}$ and we are left with

$$L^k + (1-r)^k = 1, \text{ or, } L = (1-(1-r)^k)^{1/k}.$$

We can attempt a corresponding "trickle-down" theory by assuming that $\frac{\partial I}{\partial t}$ depends on $\frac{\frac{A}{N}(1-L)}{Nr}$—a family at rank r developing, with those of lower rank (Nr), the wealth of those of higher rank ($\frac{A}{N}(1-L)$). From

$$\frac{\partial I}{\partial t} = e^C\left(\frac{A(1-L)}{N^2r}\right)^B$$

we derive the second family of Lamé curves mentioned earlier, $L = 1-(1-r^k)^{1/k}$. But in this case, the constant $k = 1-B$ is greater than 1, meaning B is negative. In other words, we are left with another trickle-up theory, which one might describe as a dollar in the hands of someone at rank r sharing development rights on families of lower rank with all the dollars in the hands of those of higher rank.

These trickle-up theories suggest two additional theories, one in which dollars develop dollars (with $\frac{\partial I}{\partial t}$ proportional to $\frac{1-L}{L}$) and one in which people develop people (with $\frac{\partial I}{\partial t}$ proportional to $\frac{1-r}{r}$). Both of these result in one-parameter families, but the families lack closed-form expressions because the differential equations can't be solved analytically. The MSEs associated with these approaches are of the same order of magnitude as those associated with the other approaches.

A discussion of these alternative trickle-up theories is beyond the scope of this paper. Initial investigation, however, has persuaded the authors that it would be difficult to argue that any one is significantly better than the others.

4 Checks and Balances

The success of the Lamé curves suggests that there is something fundamentally one-dimensional about inequality. That is a radical hypothesis that should be treated with caution. We explore the hypothesis and its ramifications here.

A. Lamé curves, as solutions to a differential equation, do not cross. But Lorenz curves do cross. Kakwani (1984) reports that in a collection of Lorenz curves 21% of the pairs intersected. Does this falsify our hypothesis?

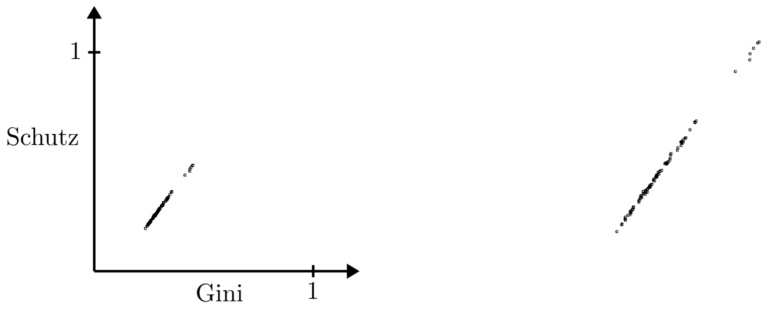
We don't believe it does. Consider what we might find if the Lorenz curve for a particular time and place were computed from two independently collected data sets. The curves would follow the same basic arc but would vary up and down. The two would almost certainly cross several times. For countries whose Lorenz curves are close, it doesn't seem surprising that they would cross.

B. We have just defended the hypothesis by appealing to possible errors or random variation in the data. But the data are also the basis for our argument. Is that a difficulty?

The LIS data, we understand, is the gold-standard for income data, yet we did experience some difficulties with it. The program supplied for computing deciles, for example, had a bug. Even after dealing with that, we found at least one set of numbers that raised suspicions.[35] But unless the data have systemic biases, it seems a reasonable source on which to base our models. Note that our confidence in the data does have limits. We have four different one-parameter families (the two presented in the Introduction and the two noted at the end of the previous section), all of which model Lorenz curves well, but we don't feel we can distinguish among them.

C. If Lorenz curves were fundamentally Lamé curves, then all monotonic measures of inequality would be equivalent in the sense that knowing one measure gives you all the others. Suppose, for example, we knew the Gini coefficient g of a Lorenz curve. Then we could find the unique Lamé curve with Gini coefficient g. From that we could compute the Schutz index. Conversely, given the Schutz index, we could recover the Gini coefficient. Further, if the computations of two measures are continuous, then plotting the measures against each other should result in a connected curve.

Indeed, that seems to be the case. Here is the plot for Gini vs. Schutz, from the LIS data. The graph on the right is a magnification.



[35] Italy, 1991, as mentioned in the previous section.

Another measure of inequality is suggested by the trickle-up theory, the exponent B in the partial differential equation,

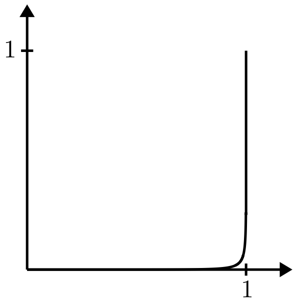
$$\frac{\partial I}{\partial t} = e^C \left(\frac{AL}{N^2(1-r)} \right)^B .$$

Since $B = 1 - k$, B can be determined from the best-fitting model of the Lorenz curve from the family $L(r) = \{(1-(1-r)^k)^{1/k}\}$. We could call this measure the "sensitivity factor" since it reflects how sensitive income growth for an individual is to the incomes of others. Perfect equality occurs when sensitivity is zero ($B = 0$, $k = 1$):

$$L(r) = 1 - (1-r)^1)^{1/1} = r.$$

In that case, all incomes grow at the same absolute rate: $\left(\frac{\partial I}{\partial t} = e^C \right)$. At steady-state, where the Lorenz curve doesn't change, incomes can still grow, but they must all grow proportionally. The only way incomes can grow at the same absolute rate and the same proportional rate is if they are all equal.

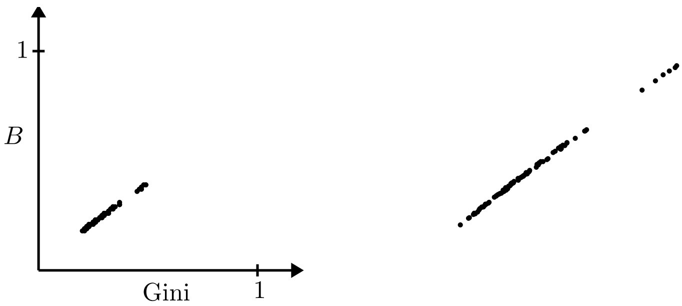
Similarly, as k approaches 0, the Lamé curves approach absolute inequality.



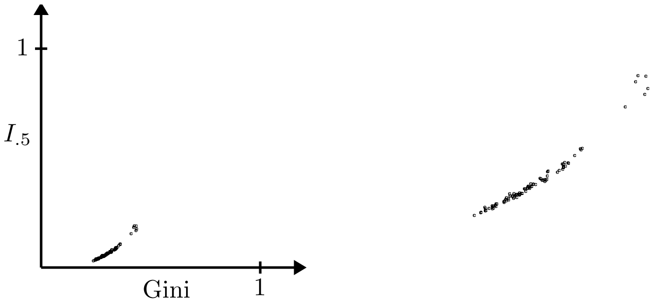
$L(r) = (1-(1-r)^{.2})^{1/.2}$

This reflects a growth rate directly proportional to what we might call the "opportunity for development," $\frac{AL}{N^2(1-r)}$.

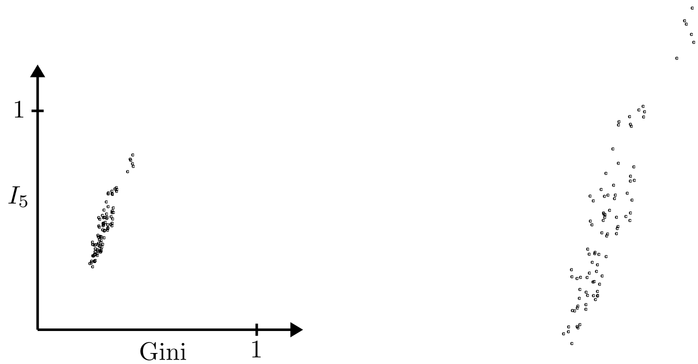
Compared to other measures, the sensitivity factor is, perhaps, less abstract and more directly meaningful. It also tracks well with the Gini coefficient (LIS data).



D. Harvey (2005) explores the relationship between the Gini coefficient and several of the Atkinson indices I_r. His plots show large scattering which would seem to refute our hypothesis. We computed for the LIS data two Atkinson indices, one where the relationship is well-behaved in Harvey's paper, $I_{.5}$,



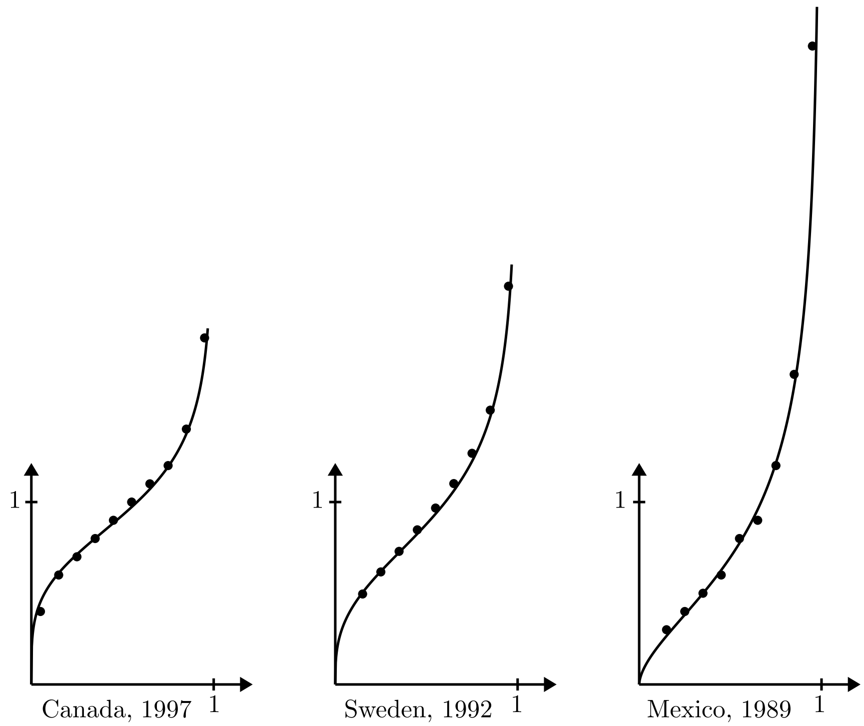
and also the one where the points are most scattered, I_5.



The picture for I_5 is not as clearly a curve as the picture for $I_{.5}$, but it's more organized and linear than the picture in Harvey (2005). The higher subscript tends to exaggerate differences at the low end of the Lorenz curve. An error, for example, of ε in the calculation of $L(.1)$ can change I_5 by more than 10ε.

E. Finally, we are modelling a curve that is confined to a small space, a curve that must go from $(0,0)$ to $(1,1)$ with a constantly increasing derivative. Under those circumstances, modelling closely with a carefully chosen family of curves may seem unspectacular.

We considered this and thought to test how well the derivatives of our curves matched the derivatives of the Lorenz curves (Pen's Parade, since L' is proportional to income). This is a significantly greater challenge, since the derivative is theoretically unbounded. The following graphs show the derivatives of the Lamé curves for countries with varying degrees of inequality. The points are the difference quotients formed from consecutive decile data.



The model passes this test surprisingly well.

5 Questions

We have presented evidence that Lorenz curves for income taken at different times in different countries are well-modelled by curves from a one-parameter family of functions. Of course, additional parameters produce better fits. Modern economies are subject to countless disturbances which must vary the Lorenz curve in local but significant ways.

But the closeness of the approximations produced by a single parameter implies that the distribution of income in a society is largely characterized by a single number. This raises some related questions.

1. Is there a single economic variable that drives inequality?

2. What are the ways in which the sensitivity factor can be changed through economic policy?

3. What does the success of trickle-up theories have to say about how governments should stimulate the economy?

4. Can $B = 1 - k$ be seen as a measure of the efficiency of an economy? If so, does this suggest an explicit trade-off between efficiency and equality?

5. Inequality in the United States decreased from 1950 to 1970 (Henle, 1972) and increased from 1979 to 2000 (the Gini coefficient increased steadily from .301 to .368). Can the framework of this paper help explain these trends?

6. The relationship between the Gini coefficient and the sensitivity factor appears almost linear. Does this mean that the Gini coefficient has a concrete interpretation? That is, does the Gini coefficient tell us something definite about the relation between the rate of growth of one's income and the income of those who earn less?

7. We were not able to distinguish among the four models of income growth ($\frac{L}{1-r}$, $\frac{r}{1-L}$, $\frac{L}{1-L}$, $\frac{r}{1-r}$). Is there a way of determining which leads to the best for model for Lorenz curves?

Appendix A

On the education data

We found a Lorenz curve for educational attainment using data from the U.S. Census Bureau. In the table below, we have noted the number of years we attached to each category.

The data gave us (with $(0,0)$ and $(1,1)$) 16 points on the Lorenz curve.

Table 14.1: Educational Attainment of Persons 15 Years Old and Over (all races, both sexes, in thousands)

none (years: 0)	1st-4th (years: 2.5)	5th-6th (years: 5.5)	7th-8th (years: 7.5)	9th (years: 9)	10th (years: 10)	11th (years: 11)	H.S. grad (years: 12)	Some college (years: 13)	Associate degree (years: 14)	Bachelor's degree (years: 16)	Master's degree (years: 17)	Professional Degree (years: 18)	Doctorate Degree (years: 22)
887	2091	3911	9039	8212	9795	12993	66210	38315	13998	3090	9295	2586	1869

Appendix B

Computational details

For interest, we report here on the techniques we used to compute (a) the Gini coefficient and (b) the Schutz index.

(a) We computed the Gini coefficient from quintile data using a Newton-Cotes formula.

Given the value of a function f at three values, a, $a+.5(b-a)$, b, Simpson's Rule approximates the integral of f on $[a,b]$ by integrating the quadratic passing through the three points. Given the value of f at more points the Newton-Cotes formulae find more accurate approximations by integrating polynomials of higher degree. The particular formula we used (appropriate for the six points given by quintile data) approximates $\int_a^b f(x)\,dx$ by $\frac{95}{288}f(a)+\frac{125}{96}f(a+.2(b-a))+\frac{125}{144}f(a+.4(b-a))+\frac{125}{144}f(a+.6(b-a))+\frac{125}{96}f(a+.8(b-a))+\frac{95}{288}f(b)$.

(b) The Schutz index is the greatest distance between the Lorenz curve and the straight line from the origin to $(1,1)$. The difficulty is determining this given only decile data for the Lorenz curve.

A little calculus tells us that the point where this distance is greatest is where $L'(r)=1$. For most Lorenz curves, that comes when r is between .6 and .7. We then approximate L with the cubic passing through the points, $(.5,L(.5))$, $(.6,L(.6))$, $(.7,L(.7))$, $(.8,L(.8))$ and use this to find a such that $L'(a)=1$. and then to evaluate $a-L(a)$ (the Schutz index).

References

Chotikapanich, D. (1993) A Comparison of Alternative Functional Forms for the Lorenz Curve, *Economic Letters*, **41**(2), 129–138.

Harvey, J. (2005) A Note on the 'Natural Rate of Subjective Inequality' Hypothesis and the Approximate Relationship between the Gini Coefficient and the Atkinson Index, *Journal of Public Economics*, **89**, 1021–1025.

Henle, P. (1972) Exploring the Distribution of Earned Income, *Monthly Labor Review*, pp. 16–27.

Kakwani, N. (1984) Welfare Ranking of Income Distributions, *Advances in Econometrics*, **3**, 191–213.

Kakwani, N. C. and N. Podder (1973) On the Estimation of Lorenz Curves from Grouped Observations, *International Economic Review*, **14**, 278–291.

Kakwani, N. C. and N. Podder (1976) Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations, *Econometrica*, **44**(1), 137–148.

Lambert, P. (2001) *The Distribution and Redistribution of Income*, 3rd ed., Manchester University Press.

McDonald, J. B. (1984) Some Generalized Functions for the Size Distribution of Income, *Econometrica*, **52**, 647–663.

Quandt, R. (1966) Old and New Methods of Estimation and the Pareto Distribution, *Metrica*, **X**, 55–82.

Rasche, R. H., J. Gaffney, A. Koo and N. Obst (1980) Functional Forms for Estimating the Lorenz Curve, *Econometrica*, **48**, 1061–1062.

Ryu, H. K. and D. J. Slottje (1996) Two Flexible Functional Form Approaches for Approximating the Lorenz Curve., *Journal of Econometrics*, **72**, 251–274.

Sarabia, J. M., E. Castillo and D. J. Slottje (1999) An Ordered Family of Lorenz Curves, *Journal of Econometrics*, **91**, 43–60.

U.S. Census Bureau (1998) Educational Attainment in the United State, Update.

Wolff, E. N. (2000) Recent Trends in Wealth Ownership, 1983-1998, Working Paper No. 300, Jerome Levy Economics Institute.

CHAPTER 15

Lorenz Curves and Generalised Entropy Inequality Measures

Nicholas Rohde[†]

Abstract

Lorenz curves and Generalised Entropy (GE) measures are popular tools for analyzing income inequality. This paper seeks to connect these techniques by demonstrating that GE inequality measures may be derived directly from the Lorenz curve. The paper provides analytical expressions for Theil's T and L inequality measures, half the square of the coefficient of variation and Atkinson's utility based measure in terms of the Lorenz curve. Mathematical expressions for common GE measures are derived for three simple parametric specifications. The results are empirically illustrated and shown to be consistent with Lorenz dominance.

1 Introduction

Much of the current analysis of income inequality involves the use of both Lorenz curves and measures derived from the notion of Generalised Entropy. These two techniques are generally used alongside each other in most empirical studies, including prominent works by Milanovic (2002) and Sala-i Martin (2002). Despite their common application, little is known about the relationship between these methods of inequality measurement.

Of these two approaches, the Lorenz curve is more intuitive and plays a more fundamental role in inequality measurement. Graphically, the Lorenz curve gives the proportion of total societal income accruing to the lowest earning proportion of income earners, and the representation may be used to make comparisons between

[†] School of Economics, The University of Queensland, Brisbane, Australia. The author expresses his gratitude to Prasada Rao for his support and encouragement.

different distributions of income. If the Lorenz curve for distribution A lies nowhere below, and is at some point above the Lorenz curve for distribution B, the Lorenz curve for distribution A is said to dominate the Lorenz curve for distribution B. When Lorenz dominance occurs, unambiguous statements about the relative inequality levels within the distributions can be made. If distribution A Lorenz dominates distribution B, then the inequality level within distribution A is necessarily less than that within distribution B, and any inequality measure that satisfies the transfer principle will provide the same ranking (Atkinson, 1970). If Lorenz curves intersect however, Lorenz dominance fails to provide a definitive ordering and numerical measures must be used to complete the ranking.

As the most intuitive numerical inequality measures are derived from the Lorenz curve, their relationship with Lorenz dominance is straightforward. In the case of intersecting Lorenz curves, the Gini coefficient will rank the distribution which has less area enclosed between the Lorenz curve and the egalitarian line as the more equal. The Kakwani (1980) inequality measure uses the length of the Lorenz curve to determine inequality and will rank the shorter Lorenz curve as more egalitarian. Similarly the Schutz (1951) coefficient will prefer the Lorenz curve that has the lower maximum difference between the Lorenz curve and the egalitarian line.

As GE measures are related to the concept of information theory, they have no direct interpretation in terms of the Lorenz curve. This family of measures includes popular inequality indices such as Theil's T (Theil, 1967) and L (Theil, 1979) measures, as well as half the square of the coefficient of variation. While this group of measures may be criticized for lacking an intuitive appeal they possess decomposability characteristics that make them particularly useful for analyzing inequality (Cowell, 1995). Atkinson's (Atkinson, 1970) measure is included with GE measures in this analysis, as Cowell (1995) shows that the measure is closely related to the family of GE measures.

The main objective of this paper is to improve the understanding of the relationship between these techniques by providing expressions for common GE measures in terms of the Lorenz curve. The results given here are all consistent with Lorenz dominance and provide a basis for studying the behavior of GE measures when Lorenz curves intersect. The results also provide an exact method for calculating these measures in the instance where a parametrically specified Lorenz curve is known but little other data are available. These results have the same practical implications as expressions for GE measures in terms of a parametrically specified density function. One advantage to calculating GE measures from parametrically specified Lorenz curves rather than grouped data is that parametric Lorenz curves may be used for interpolation and thus inequality indices calculated from them do not understate total inequality by ignoring the variation within subgroups.

This paper is divided into five major sections. Section 2 provides a quick overview of the notation and techniques used within the paper. Section 3 discusses the discrete versions of Theil's T, Theil's L, Atkinson's measure and half the squared coefficient of variation, and shows that the concepts may be extended to a continuous Lorenz curve. Section 4 provides analytical expressions for these measures in terms of the Kakwani and Podder (1973), Gupta (1984) and Chotikapanich (1993) functional forms. Unfortunately expressions for Atkinson's measure in terms of the Gupta and Kakwani-Podder Lorenz curves are not included as they were found to be prohibitively difficult to compute. Section 5 provides an empirical illustration of the results and Section 6 gives some concluding comments.

2 Notation and Basic Concepts

The paper considers the distribution of income as it accrues unevenly across a population of j individuals. We assume that all incomes are non-negative, and denote the income of the kth individual to be x_k. If the population is arbitrarily partitioned such that we have n groups, the income share and population share of the ith group are denoted q_i and p_i respectively. The income share of group i may be calculated as the total income accruing to persons within income group i, divided by the total income of the population. Similarly the population share of group i is the proportion of total population contained within that group. For income group i that contains b_i individuals, the income and population shares may be calculated as

$$q_i = \frac{\sum_{k_i=1}^{b_i} x_{k(i)}}{\sum_{k=1}^{j} x_k} \tag{15.1}$$

$$p_i = \frac{b_i}{j} \tag{15.2}$$

Clearly

$$\sum_{i=1}^{n} q_i = 1 \tag{15.3}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{15.4}$$

Generalised Entropy inequality measures are easily calculated from income and population share data. These measures take the general form

$$GE(\alpha) = \frac{1}{\alpha^2 - \alpha}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{q_i}{p_i}\right)^{\alpha} - 1\right] \tag{15.5}$$

where α is a non-negative sensitivity parameter which dictates the emphasis the inequality measure places on higher and lower ends of the distribution. Low values for α will place extra emphasis on the lower end of the distribution, while higher values will place emphasis on the higher incomes. Taking the limits of this equation as $\alpha \to 0, 1$ gives Theil's L and T inequality measures respectively

$$L = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right) \tag{15.6}$$

$$T = \sum_{i=1}^{n} q_i \ln\left(\frac{q_i}{p_i}\right) \tag{15.7}$$

Half the square of the coefficient of variation is given when $\alpha = 2$. It may be written as

$$\frac{1}{2}CV^2 = \frac{1}{2}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{q_i}{p_i}\right)^2 - 1\right] \tag{15.8}$$

Atkinson's measure is given as

$$A = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{q_i}{p_i}\right)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} \tag{15.9}$$

where ε is an inequality aversion parameter and lies between zero and infinity. The higher the value that ε takes, the more harmful inequality is to the considered society. Cowell (1995) shows that this measure is ordinally equivalent to GE measures when $\alpha = 1 - \varepsilon$.

If the individual incomes are placed in ascending order such that $x_1 < x_2 .... < x_j$ a Lorenz curve may be constructed. The Lorenz curve is a plot of the cumulative income share of the lowest earning k individuals against the cumulative population share of the same group. If expressed as a continuous function, the Lorenz curve is typically given as

$$\eta = f(\pi) \tag{15.10}$$

where π is the cumulative population share and $0 \leq \pi \leq 1$. η is the cumulative income share and $0 \leq \eta \leq 1$. Here $\frac{d\eta}{d\pi} > 0$, $\frac{d^2\eta}{d\pi^2} > 0$, $\eta\,(0) = 0$, $\eta\,(1) = 1$.

3 Income Shares and the Lorenz Curve

We start with a linear piecewise Lorenz curve constructed from n equal sized segments, each describing successive income blocks. Data is often presented in this form, usually with $n = 10$ equal sized income groups. These aggregate income shares are represented as $(\eta_1, \eta_2, ...\eta_{10})$ and may be plotted against aggregate population shares$(\pi_1, \pi_2, ...\pi_{10})$ to form the Lorenz curve. Clearly the point (π_{10}, η_{10}) represents the termination point $(1,1)$ on the Lorenz curve, while the origin would be represented as(π_0, η_0). The population share of income group i, denoted p_i, is given as $p_i = \pi_i - \pi_{i-1}$ where i is the income group under consideration. If we are dealing with decile data, each $p_i = 0.1$. The income data may be similarly disaggregated using the formula $q_i = \eta_i - \eta_{i-1}$ to give income shares corresponding to each population share.

Since GE measures are easily calculated from income and population share data, the task now is to express the GE measures such that they can be interpreted as a discrete form integral or Riemann sum. The Riemann sum approximates a definite integral by measuring the area under a curve. This is done by filling the space under the curve with a series of rectangles and summing these to measure the enclosed area. These rectangles may be infinitesimal in size, yielding a good approximation to the area under a curve. If the area under the curve is in the range from zero to one (as for a Lorenz curve) and is divided into n evenly sized partitions, the Riemann sum may be written as

$$\int_0^1 f(x)dx \approx \sum_{i=1}^{n} f(x_i)\frac{1}{n} \tag{15.11}$$

where $f(x_i)$gives the height of partition i and $\frac{1}{n}$ gives the width. The product of these two terms gives the area of the enclosed rectangle; the sum of these terms gives an approximation of the total area under the curve. An attractive property of the Riemann approximate integral is that the limit of this sum as $n \to \infty$ is the definite integral. This allows us to dispense with the approximate symbol in equation (15.11) when considering the limit, and allows us to refer to the enclosed area as a Riemann integral.

When the population is divided into n equal sized partitions, each population share p_i is equal to $\frac{1}{n}$. Replacing this term in equation (15.6) allows Theil's L measure to be written as

$$L = -\sum_{i=1}^{n} \frac{1}{n} \ln\left(\frac{q_i}{p_i}\right) \tag{15.12}$$

Equation (15.7) may also be expressed as an approximate form integral. Multiplying and dividing through by the scaling factor$\frac{1}{n}$ gives

$$T = \sum_{i=1}^{n} \frac{1}{n} \left(\frac{q_i}{1/n}\right) \ln\left(\frac{q_i}{p_i}\right) \tag{15.13}$$

which may be written as

$$T = \sum_{i=1}^{n} \frac{1}{n} \left(\frac{q_i}{p_i}\right) \ln\left(\frac{q_i}{p_i}\right) \tag{15.14}$$

Equations (15.8) and (15.9) for half the squared coefficient of variation and Atkinson's measure are already written in the form given by equation (15.11) and may easily be interpreted as a Riemann sum.

Our focus now turns to the limit of the $\left(\frac{q_i}{p_i}\right)$ ratios that make up equations (15.8), (15.9), (15.12) and (15.14) as $n \to \infty$. As we obtained the income and population shares q_i and p_i from disaggregating the Lorenz curve, these shares may be interpreted as the "rise" and "run" of each piecewise segment. As we increase the number of segments to the Lorenz curve we get a parallel to Newton's Difference Quotient for the derivative. This is

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{15.15}$$

which states that the derivative of f at x is given by the limit of the difference quotient as $h \to 0$. This formula may be applied to the Lorenz curve. If two points are chosen on the π axis, the "run" or gap between the two points is given as $\pi_i - \pi_{i-1}$, which is denoted h in the difference quotient formula, and defined as the population share p_i. Similarly the "rise" is expressed as$f(x+h) - f(x)$ which is given by $\eta_i - \eta_{i-1}$ in Lorenz curve notation and is equal to income share q_i. Substituting these expressions into equation (15.15) produces the result that the derivative of η at π is

$$\eta'(\pi_i) = \lim_{n \to \infty} \frac{q_i}{p_i} \tag{15.16}$$

where $h = \frac{1}{n}$. This result is consistent with Kakwani's (Kakwani, 1980) finding that the slope of the Lorenz curve is equal to $\frac{x(F)}{\mu}$, where $x(F)$ is the inverse function of the income CDF and lies on the interval [0,1). This may be seen by examining the discrete income data outlined in the previous section, where we have j non-negative incomes ordered such that $x_1 < x_2 .... < x_j$ where x_k represents the income accruing to the kth individual.

If F is taken to be equal to $\frac{k}{j}$, then $x(F)$, which may be interpreted as the income accruing to the person earning at the $\frac{k}{j}$ proportion of income earners will be equal to x_k. The mean income level is $\mu = \frac{1}{j}\sum_{k=1}^{j} x_k$, giving the result that the slope of the Lorenz curve using discrete data at point $\pi = \frac{k}{j}$ is equal to

$$\eta' = \frac{x_k}{\left(\frac{1}{j}\sum_{k=1}^{j} x_k\right)} \tag{15.17}$$

The result used in this paper shows that the derivative of a linear segment is equal to $\frac{q_i}{p_i}$. We may calculate q_i from discrete data using the formula

$$q_i = \frac{x_k}{\sum_{k=1}^{j} x_k} \tag{15.18}$$

where the numerator is the income earned by the individual determining income "group" i and is equal to x_k . The denominator is the total population income.

The population share is given as

$$p_i = \frac{1}{j} \tag{15.19}$$

Using equations (15.18) and (15.19) to compute $\frac{q_i}{p_i}$, which is an input component for all GE measures, gives the result that the slope of the Lorenz curve at $\pi = \frac{k}{j}$ is equal to

$$\eta' = \frac{j x_k}{\left(\sum_{k=1}^{j} x_k\right)} \tag{15.20}$$

which is clearly equal to the result given by Kakwani in equation (15.17).

Applying the results from equations (15.11) and (15.16) to the discrete GE measures given in equations (15.12), (15.14) and (15.8) and (15.9) respectively gives the following results for the selected GE measures in terms of the continuous Lorenz curve. Theil's L and T measures are denoted L and T respectively, while Atkinson's measure is denoted A.

$$L = \lim_{n\to\infty} -\sum_{i=1}^{n} \frac{1}{n} \ln\left(\frac{q_i}{p_i}\right) = -\int_0^1 \ln \eta'(\pi) d\pi \tag{15.21}$$

$$T = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n} \left(\frac{q_i}{p_i}\right) \ln\left(\frac{q_i}{p_i}\right) = \int_0^1 \eta'(\pi) \ln(\eta'(\pi)) d\pi \tag{15.22}$$

$$\frac{1}{2}CV^2 = \lim_{n\to\infty} \frac{1}{2} \sum_{i=1}^{n} \frac{1}{n} \left(\left(\frac{q_i}{p_i}\right)^2 - 1\right) = \frac{1}{2} \int_0^1 \left(\eta'(\pi)^2 - 1\right) d\pi \tag{15.23}$$

$$A = \lim_{n\to\infty} 1 - \left[\sum_{i=1}^{n} \frac{1}{n} \left(\frac{q_i}{p_i}\right)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} = 1 - \left[\int_0^1 \eta'(\pi)^{1-\varepsilon} d\pi\right]^{\frac{1}{1-\varepsilon}} \tag{15.24}$$

Equations (15.21)-(15.24) show that the four inequality measures may be calculated directly from the Lorenz curve. Thus in the absence of any other data, the Lorenz curve is sufficient for the calculation of these GE measures.

4 Calculating Analytical Expressions for GE Measures from Simple Lorenz Curves

In this section we derive analytical expressions for the given inequality measures associated with three selected functional specifications for Lorenz curves. The functional specifications are taken from Chotikapanich (1993), Gupta (1984) and Kakwani and Podder (1973). Direct analytical solutions for Atkinson's measure are not provided for the Gupta and Kakwani-Podder Lorenz curves as they were found to be difficult to determine.

4.1 The Chotikapanich specification

The Chotikapanich Lorenz curve is a simple functional form that has some useful properties. This functional form was used in a recent inequality analysis by Dowrick and Akmal (Dowrick and Akmal, 2005) and is particularly attractive for this purpose as it is easily integrable. The functional form for the specification of the Lorenz curve, $\eta = f(\pi)$ is

$$\eta = \frac{e^{k\pi} - 1}{e^k - 1}, \qquad k > 0 \tag{15.25}$$

The first and second order derivatives of the Lorenz curve with respect to π are given by:

$$\eta' = \frac{ke^{k\pi}}{e^k - 1} > 0 \tag{15.26}$$

$$\eta'' = \frac{k^2 e^{k\pi}}{e^k - 1} > 0 \tag{15.27}$$

Using equations (15.21)-(15.24) we are able to determine analytical solutions for the given GE measures in terms of parameter k (and ε in the case of Atkinson's measure).

$$L = \frac{k}{2} - \ln\left(\frac{ke^k}{e^k - 1}\right) \tag{15.28}$$

$$T = k + \ln k - \ln(e^k - 1) - \left(\frac{e^k - k - 1}{e^k - 1}\right) \tag{15.29}$$

$$\frac{1}{2}CV^2 = \frac{1}{2}\left[\frac{k\left(e^{2k} - 1\right)}{2\left(e^k - 1\right)^2} - 1\right] \tag{15.30}$$

$$A = 1 - \left[\frac{1}{k(\varepsilon - 1)}\left(\left(\frac{k}{e^k - 1}\right)^{1-\varepsilon} - \left(\frac{ke^k}{e^k - 1}\right)^{1-\varepsilon}\right)\right]^{\frac{1}{1-\varepsilon}} \tag{15.31}$$

4.2 The Gupta specification

Gupta provides another single parameter Lorenz curve given by the following equation.

$$\eta = \pi a^{\pi - 1}, \qquad a > 0 \tag{15.32}$$

It can be seen that the first and second order derivatives of the Lorenz curve in Equation 15.32 are given by

$$\eta' = a^{\pi - 1}(1 + \pi \ln a) > 0 \tag{15.33}$$

$$\eta'' = a^{\pi - 1} \ln a (2 + \pi \ln a) > 0 \tag{15.34}$$

The solutions for the GE measures in terms of parameter a are

$$L = 1 + \ln\left(\frac{a}{\ln(a)+1}\right) - \frac{\ln(\ln(a)+1)}{\ln(a)} - \frac{\ln(a)}{2} \tag{15.35}$$

$$T = \ln(1+\ln a) - \left[\left(\frac{1}{\ln a} - \frac{\gamma + \ln(\ln a + 1) + \sum_{k=1}^{\infty}\frac{(\ln a+1)^k}{kk!}}{a e \ln^2 a}\right)\ln a + \left(\frac{\gamma + \sum_{k=1}^{\infty}\frac{1}{kk!}}{a e \ln^2 a}\right)\ln a\right] \tag{15.36}$$

$$\frac{1}{2}CV^2 = \frac{2\ln(a)(\ln(a)+1) - a^{-2} + 1}{8\ln(a)} - \frac{1}{2} \tag{15.37}$$

The integral required to express Theil's T measure in terms of parameter estimate a does not have a closed form solution. The analytical equation given here uses series approximations as a substitute for exponential integral (Hildebrand, 1962). An exponential integral is defined as

$$E_i(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt \tag{15.38}$$

and may be used to generate integrals of functions where no clear anti-derivative exists. The exponential integrals are represented here with the series

$$E_i(x) = \gamma + \ln x + \sum_{k=1}^{\infty} \frac{x^k}{kk!} \tag{15.39}$$

where γ is the Euler gamma constant and is approximately equal to 0.57721. This constant is defined as the limiting difference between the harmonic series and the log function. As a result the computation of Theil's T measure based on equation (15.36) requires a series approximation and is therefore not a closed form solution.

4.3 The Kakwani and Podder specification

Kakwani and Podder proposed the following functional form for the Lorenz curve which was used in their 1973 Australian income study. The Kakwani Podder Lorenz curve is specified as

$$\eta = \pi e^{-\beta(1-\pi)}, \qquad \beta > 0 \tag{15.40}$$

The first and second order derivatives of the Lorenz curve are given by

$$\eta' = e^{-\beta(1-\pi)}(1+\pi\beta) > 0 \tag{15.41}$$

$$\eta'' = \beta e^{-\beta(1-\pi)}(2+\pi\beta) > 0, \tag{15.42}$$

and the above GE measures are

$$L = \frac{\beta+2}{2} - \ln(\beta+1)\left(1+\frac{1}{\beta}\right) \tag{15.43}$$

$$T = \ln(1+\beta) - \left[\frac{e^{-\beta-1}\left(\beta e^{\beta+1} - \left[\left(\gamma+\ln(\beta+1)+\sum_{k=1}^{\infty}\frac{(\beta+1)^k}{kk!}\right)-\left(\gamma+\sum_{k=1}^{\infty}\frac{1}{kk!}\right)\right]\right)}{\beta}\right] \tag{15.44}$$

$$\frac{1}{2}CV^2 = \frac{2\beta(\beta-1)-e^{-2\beta}+1}{8\beta} \tag{15.45}$$

Again the expression for Theil's T measure in terms of the Lorenz curve parameter requires a series approximation for an exponential integral. This need not be considered a problem however, as the series rapidly converges to the true value defined by the exponential integral. Equations (15.28)-(15.31), (15.35), (15.37), (15.43) and (15.45) demonstrate the feasibility of producing closed form solutions for GE inequality measures from Lorenz curve parameters.

5 Empirical Illustration of Results

In this section the empirical validity of the equations giving GE measures directly in terms of the parameters of the Lorenz curve is examined using a simulation experiment. To conduct this experiment a realistic range of parameter values are chosen for each Lorenz curve. For each parameter value the set of implied decile income shares is generated. Approximate GE measures are calculated from these decile shares, and represent what is commonly reported as GE inequality estimates. Exact inequality measurements are then calculated using the analytic solutions provided in the paper and are compared to the commonly used approximate values. The validity and accuracy of the results are reflected in the closeness of the approximate and exact inequality measurements.

The following tables show the results from the simulation for six different parameter values for the Chotikapanich, Gupta and Kakwani-Podder Lorenz curves. Column 1 of the tables gives the selected parameter values for each Lorenz curve. Columns 2 and 3 give the approximation from decile shares and analytical values for Theil's L measure based on the selected parameter value. Columns 4 and 5 give the corresponding approximate and analytic measurements from Theil's T measure, while columns 6 and 7 repeat the process for half the squared coefficient of variation. An extra table is included to demonstrate the behavior of Atkinson's measure with respect to the parameter k from the Chotikapanich specification. This is done for three chosen values for ε, and the analytical results are again presented alongside approximate values calculated from decile shares.

5.1 Results from Simulations

Parameter values ranging from one to six have been chosen for Chotikapanich and Kakwani-Podder functional forms to demonstrate the effectiveness of the analytic solutions over a wide range of values. In practice most parameter estimates for the Chotikapanich Lorenz curve lie between two and four, while most parameter estimates for the Kakwani-Podder functional form lie between one and three. Parameter estimates for the Gupta Lorenz curve are typically higher, usually ranging between 4 and 12. All analytical solutions accurately match the discrete approximations even when using extreme parameter values. As expected, the values calculated using the analytical expressions are all slightly higher than the values from the decile share data. This is because data presented in grouped form ignores the inequality within each group. The analytic solutions do not face this drawback. The results are all consistent with expectations of Lorenz dominance with higher parameter values giving larger inequality estimates in all cases.

Table 15.1: Chotikapanich Lorenz curve: $\eta = \frac{e^{k\pi}-1}{e^{k}-1}$

Parameter (k) value	Theil L decile	Theil L analytic	Theil T decile	Theil T analytic	$\frac{1}{2}CV^2$ decile	$\frac{1}{2}CV^2$ analytic
1	0.040908	0.041324	0.040227	0.040911	0.040537	0.040988
2	0.159773	0.161439	0.149902	0.151910	0.154337	0.156517
3	0.346570	0.350318	0.303075	0.307247	0.322434	0.328593
4	0.588561	0.595220	0.472698	0.479863	0.523701	0.537314
5	0.873405	0.883801	0.639681	0.650656	0.741207	0.766959
6	1.190802	1.205758	0.794193	0.809783	0.963801	1.007454

Analytic values for this table are based on algebraic expressions (15.28)-(15.30). Decile values represent what is commonly reported for these indices.

Table 15.2: Gupta Lorenz curve: $\eta = \pi a^{\pi-1}$

Parameter (k) value	Theil L decile	Theil L analytic	Theil T decile	Theil T analytic	$\frac{1}{2}CV^2$ decile	$\frac{1}{2}CV^2$ analytic
2	0.059654	0.060277	0.057742	0.058393	0.057899	0.058539
4	0.193880	0.196019	0.175846	0.177790	0.178681	0.181106
6	0.292866	0.296211	0.254285	0.257195	0.261675	0.265765
8	0.369756	0.374082	0.310943	0.314617	0.323430	0.329033
10	0.432575	0.437733	0.354771	0.359078	0.372403	0.379390
12	0.485722	0.491602	0.390270	0.395122	0.412918	0.421181

Analytic values are based on algebraic expressions (15.35)-(15.37).

Table 15.3: Kakwani-Podder Lorenz curve: $\eta = \pi e^{-\beta(1-\pi)}$

Parameter (k) value	Theil L decile	Theil L analytic	Theil T decile	Theil T analytic	$\frac{1}{2}CV^2$ decile	$\frac{1}{2}CV^2$ analytic
1	0.112503	0.113705	0.106020	0.107239	0.106801	0.108083
2	0.348035	0.352081	0.295247	0.298786	0.306199	0.311355
3	0.643483	0.651607	0.488006	0.494619	0.528921	0.541563
4	0.974912	0.988202	0.663555	0.673970	0.756373	0.781239
5	1.330424	1.349888	0.818505	0.833475	0.982144	1.024998
6	1.703177	1.729771	0.954457	0.974748	1.203314	1.270833

Analytic values are based on algebraic expressions (15.43)-(15.45).

Table 15.4: Atkinson's measure from Chotikapanich Lorenz curve

Parameter (k) value	Atkinson decile ($\varepsilon = 1.5$)	Atkinson analytic ($\varepsilon = 1.5$)	Atkinson decile ($\varepsilon = 2$)	Atkinson analytic ($\varepsilon = 2$)	Atkinson decile ($\varepsilon = 3$)	Atkinson analytic ($\varepsilon = 3$)
1	0.059637	0.060223	0.078558	0.079325	0.113784	0.114890
2	0.214621	0.216580	0.273521	0.275937	0.364951	0.368115
3	0.410688	0.413991	0.499997	0.503729	0.610141	0.614492
4	0.596713	0.600721	0.691836	0.695911	0.784632	0.788880
5	0.744444	0.748400	0.825670	0.829257	0.889354	0.892738
6	0.848010	0.851384	0.907597	0.910320	0.945995	0.948351

Analytic values are based on equation (15.31).

6 Conclusions

The paper has provided a formal link between the popular Generalised Entropy inequality measures and the Lorenz curve, which forms the basis for a number of other inequality measures including the Gini coefficient. The paper provides mathematical expressions for these measures in terms of the Lorenz curve and its parameters. The main result of the paper shows that the Lorenz curve can be seen as the basis for most of the commonly used inequality measures, including GE measures which stem form information theory. Analytical expressions for GE inequality statistics are derived for some specific Lorenz curve functional forms and results from the simulation experiments demonstrate the validity of the analytical expressions provided.

References

Atkinson, A. B. (1970) On the Measurement of Inequality, *Journal of Economic Theory*, **2**, 244–263.

Chotikapanich, D. (1993) A Comparison of Alternative Functional Forms for the Lorenz Curve, *Economics Letters*, **41**, 129–138.

Cowell, F. (1995) *Measuring Inequality*, 2nd ed., Harvester Wheatsheaf. Hemel Hempstead.

Dowrick, S. and M. Akmal (2005) Contradictory Trends in Global Income Inequality: A Tale of Two Biases, *Review of Income and Wealth*, **51**, 201–229.

Gupta, M. (1984) Functional Form for Estimating the Lorenz Curve: Note, *Econometrica*, **52**, 1313–1314.

Hildebrand, F. (1962) *Advanced Calculus for Applications*, Prentice Hall, New Jersey.

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

Kakwani, N. and N. Podder (1973) On the Estimation of Lorenz Curves from Grouped Observations, *International Economic Review*, **14**, 278–292.

Milanovic, B. (2002) True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone, *Economic Journal*, **112**, 51–92.

Sala-i Martin, X. (2002) The World Distribution of Income (Estimated from Individual Country Distributions), *NBER Working Paper 8905*, pp. 1–68.

Schutz, R. R. (1951) On the Measurement of Income Inequality, *American Economic Review*, **41**, 107–122.

Theil, H. (1979) World Income Inequality and its Components, *Economics Letters*, **2**, 99–102.

Theil, H. (1967) *Economics and Information Theory*, North Holland, Amsterdam.

CHAPTER 16

Estimating Income Distributions Using a Mixture of Gamma Densities

Duangkamon Chotikapanich†
and William E Griffiths‡

Abstract

The estimation of income distributions is important for assessing income inequality and poverty and for making comparisons of inequality and poverty over time, countries and regions, as well as before and after changes in taxation and transfer policies. Distributions have been estimated both parametrically and non-parametrically. Parametric estimation is convenient because it facilitates subsequent inferences about inequality and poverty measures and lends itself to further analysis such as the combining of regional distributions into a national distribution. Non-parametric estimation makes inferences more difficult, but it does not place what are sometimes unreasonable restrictions on the nature of the distribution. By estimating a mixture of gamma distributions, in this paper we attempt to benefit from the advantages of parametric estimation without suffering the disadvantage of inflexibility. Using a sample of Canadian income data, we use Bayesian inference to estimate gamma mixtures with two and three components. We describe how to obtain a predictive density and distribution function for income and illustrate the flexibility of the mixture. Posterior densities for Lorenz curve ordinates and the Gini coefficient are obtained.

† Department of Econometrics and Business Statistics, Monash University, Australia.
Email: chotikap@buseco.monash.edu.au

‡ Department of Economics, University of Melbourne, Australia.
Email: b.griffiths@unimelb.edu.au

1 Introduction

The estimation of income distributions has played a major role in economic analysis. Information from such estimations is used to measure welfare, inequality and poverty, to assess changes in these measures over time, and to compare measures across countries, over time and before and after specific policy changes, designed, for example, to alleviate poverty. Typical inequality measures are the Gini coefficient and Atkinson's inequality measure. Measures of poverty are based on the proportion of population below a threshold or the expected value of a function over that part of the income distribution below a threshold. See, for example, Kakwani (1999). Estimates of these quantities and the Lorenz curve, a fundamental tool for measuring inequality, depend on the income distribution and how it is estimated. Thus, the estimation of income distributions is of central importance for assessing many aspects of the well being of society. A convenient reference for accessing the literature on the various dimensions of inequality measurement, and how they relate to welfare in society is Silber (1999).

A large number of alternative distributions have been suggested in the literature for estimating income distributions. See Kleiber and Kotz (2003) for a review of many of them, one of which is the Dagum distribution, whose inventor is being honoured by this volume. Further reviews of alternative income distributions appear elsewhere in this volume. After an income distribution model has been selected and estimated, probability distributions are used to draw inferences about inequality and poverty measures. These probability distributions can be sampling distributions for estimators of inequality and poverty, or Bayesian posterior distributions for inequality and poverty measures. In each case the required probability distributions are derived from corresponding probability distributions for the parameters (or their estimators) of the assumed income distribution. This parametric approach to the analysis of income distributions can be applied to a sample of individuals, typically obtained via household surveys, or to more limited grouped data which may be the only form available. An advantage of the parametric approach is the ease with which probability distributions for inferences about inequality and poverty can be derived from those for the income distribution parameters. Also, in the case of more limited grouped data, the parametric approach gives a complete picture of the income distribution by allowing for within-group inequality. For an example of where the latter advantage is utilized, see Chotikapanich *et al.* (2007) who estimated generalized beta distributions from grouped data.

Assuming a particular parametric distribution also has disadvantages. Inferences about inequality can depend critically on what distribution is chosen. This was evident in the work of Chotikapanich and Griffiths (2006) who found the posterior probabilities for Lorenz and stochastic dominance were sensitive to the choice of a Singh-Maddala or Dagum income distribution. To avoid the sensitivity of inferences to choice of income distribution, nonparametric approaches are frequently used. See Cowell (1999) and Barret and Donald (2003) for examples of nonparametric sampling theory approaches and Hasegawa and Kozumi (2003) for a Bayesian approach.

One way of attempting to capture the advantages but not the disadvantages of a parametric specification of an income distribution is to use a functional form that is relatively flexible. This paper represents an attempt in this direction. Mixtures of distributions can provide flexible specifications and, under certain conditions, can approximate a distribution of any form. With these characteristics in mind, we consider a mixture of gamma distributions; the gamma density is convenient one and it has been widely used for estimating income distributions. Our approach is Bayesian. Using data on before-tax income for Canada in 1978, taken from the Canadian Family Expenditure Survey and kindly provided by Gary Barrett, we find (i) posterior densities for the parameters of a gamma mixture, (ii) an estimate of the income distribution and 95% probability limits on the distribution, (iii) the posterior density for the Gini coefficient and (iv) an estimate of the Lorenz curve and 95% probability limits on this curve.

In Section 2 we specify the Gamma mixture and describe the Markov chain Monte Carlo algorithm (MCMC) for drawing observations from the posterior density for the parameters of the mixture. The data set and our selection of prior parameters is given in Section 3. Section 4 contains the results and a summary of the expressions used to obtain those results. Goodness-of-fit comparisons with other functional forms for the income distribution are given in Section 5. Some concluding remarks appear in Section 6.

2 Estimating the Gamma Mixture Model

An income distribution that follows a gamma mixture with k components can be written as

$$f(x\,|\,\mathbf{w},\boldsymbol{\mu},\boldsymbol{\nu}) = \sum_{z=1}^{k} w_z G(x\,|\,\nu_z,\nu_z/\mu_z) \tag{16.1}$$

where x is a random draw of income from the probability density function (pdf) $f(x\,|\,\mathbf{w},\boldsymbol{\mu},\boldsymbol{\nu})$, with parameter vectors, $\mathbf{w} = (w_1,w_2,\ldots,w_k)'$, $\boldsymbol{\mu} = (\mu_1,\mu_2,\ldots,\mu_k)'$, and $\boldsymbol{\nu} = (\nu_1,\nu_2,\ldots,\nu_k)'$. The pdf $G(x\,|\,\nu_z,\nu_z/\mu_z)$ is a gamma density with mean $\mu_z > 0$ and shape parameter $\nu_z > 0$. That is,

$$G(x\,|\,\nu_z,\nu_z/\mu_z) = \frac{(\nu_z/\mu_z)^{\nu_z}}{\Gamma(\nu_z)} x^{\nu_z-1} \exp\left(-\frac{\nu_z}{\mu_z}x\right) \tag{16.2}$$

Including the mean μ_z as one of the parameters in the pdf makes the parameterization in (16.2) different from the standard textbook one, but it is convenient for later analysis. The parameter w_z is the probability that the i-th observation comes from the z-th component in the mixture. To define it explicitly, let $\mathbf{x} = (x_1,x_2,\ldots,x_n)$ be a random sample from (16.1), and let $Z_1,Z_2,\ldots,Z_n$ be indicator variables such that $Z_i = z$ when the i-th observation comes from the z-th component in the mixture. Then,

$$P(Z_i = z \,|\, \mathbf{w}) = w_z \quad \text{for } z = 1, 2, \ldots, k$$

with $w_z > 0$ and $\sum_{z=1}^{k} w_z = 1$. Also, conditional on $Z_i = z$, the distribution of x_i is $G\left(v_z, v_z/\mu_z\right)$.

To use Bayesian inference, we specify prior distributions on the unknown parameters $\mathbf{w}$, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$, and then combine these pdfs with the likelihood function defined by (16.1) to obtain a joint posterior pdf for the unknown parameters. This joint posterior pdf represents our post-sample knowledge about the parameters and is the source of inferences about them. However, as is typically the case in Bayesian inference, the joint posterior pdf is analytically intractable. This problem is solved by using MCMC techniques to draw observations from the joint posterior pdf and using these draws to estimate the quantities required for inference. Because we are interested in not just the parameters, but also the income distribution, the Gini coefficient, and the Lorenz curve, the parameter draws are also used in further analysis to estimate posterior information about these quantities.

The MCMC algorithm used to draw observations from the posterior density for $(\boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{w})$ is taken from Wiper *et al.* (2001). In the context of other problems, Wiper *et al.* consider estimation for both a known and an unknown k. We will assume a known value of k that is specified *a priori*. In our empirical work we considered k = 3 and k = 2 but settled on k = 2 as an adequate formulation. The MCMC algorithm is a Gibbs sampling one where draws are taken sequentially and iteratively from the conditional posterior pdfs for each of the parameters. Because only the conditional posterior pdfs are involved in this process, it is not necessary to specify the complete joint posterior pdf. The relevant conditional posterior pdfs are sufficient; they are specified below after we introduce the prior pdfs.

Following Wiper *et al.* (2001), the prior distributions used for each of the parameters are

$$f(\mathbf{w}) = D(\varphi) \propto w_1^{\phi_1 - 1} w_2^{\phi_2 - 1} \ldots w_k^{\phi_k - 1} \qquad \text{(Dirichlet)} \qquad (16.3)$$

$$f(v_z) \propto \exp\{-\theta v_z\} \qquad \text{(exponential)} \qquad (16.4)$$

$$f(\mu_z) = GI(\alpha_z, \beta_z) \propto \mu_z^{-(\alpha_z + 1)} \exp\left\{-\frac{\beta_z}{\mu_z}\right\} \qquad \text{(inverted gamma)} \qquad (16.5)$$

$$\text{for } z = 1, 2, \ldots, k$$

The Dirichlet distribution is the same as a beta distribution for $k = 2$ and a multivariate extension of the beta distribution for $k > 2$. Its parameters are $\varphi = (\phi_1, \phi_2, \ldots, \phi_k)'$. To appreciate the relationship between the gamma and inverted gamma pdfs, note that if $y \sim G(\alpha, \beta)$, then $q = (1/y) \sim GI(\alpha, \beta)$. The pdfs in (16.3), (16.4) and (16.5) are chosen because they combine nicely with the likelihood function for derivation of the conditional posterior pdfs, and because they are sufficiently flexible to represent vague prior information which can be dominated by the sample data. In addition to the above prior pdfs, the restriction $\mu_1 < \mu_2 < \cdots < \mu_k$ is imposed *a priori* to ensure identifiability of the posterior distribution. Settings for the prior parameters $(\varphi, \theta, \alpha_z, \beta_z)$ are discussed in Section 3.

After completing the algebra necessary to combine the prior pdfs with the likelihood function in such a way that isolates the conditional posterior densities for use in a Gibbs sampler, we obtain the following conditional posterior pdfs.

The posterior probability that the i-th observation comes from the z-th component in the mixture, conditional on the unknown parameters, is the discrete pdf

$$P(Z_i = z \,|\, \mathbf{x}, \mathbf{w}, \boldsymbol{\nu}, \boldsymbol{\mu}) = \frac{p_{iz}}{p_{i1} + p_{i2} + \cdots + p_{ik}} \tag{16.6}$$

where

$$p_{iz} = w_z \frac{(\nu_z/\mu_z)^{\nu_z}}{\Gamma(\nu_z)} x_i^{\nu_z - 1} \exp\left\{ -\frac{\nu_z x_i}{\mu_z} \right\}$$

The posterior pdf for the mixture-component probabilities $\mathbf{w}$, conditional on the other parameters and on the realized components for each observation $\mathbf{z} = (z_1, z_2, \ldots, z_n)'$, is the Dirichlet pdf

$$f(\mathbf{w} \,|\, \mathbf{x}, \mathbf{z}, \boldsymbol{\nu}, \boldsymbol{\mu}) = D(\boldsymbol{\varphi} + \boldsymbol{n}) \tag{16.7}$$

where $\mathbf{n} = (n_1, n_2, \ldots, n_k)'$, with n_z being the number of observations for which $Z_i = z$. Thus, $\sum_{z=1}^{k} n_z = n$.

The posterior pdfs for the means of the component densities μ_z, conditional on the other parameters and on $\mathbf{z}$, are the inverted gamma pdfs

$$f(\mu_z \,|\, \mathbf{x}, \mathbf{z}, \mathbf{w}, \boldsymbol{\nu}) = GI(\alpha_z + n_z \nu_z,\ \ \beta_z + S_z \nu_z) \tag{16.8}$$

where $S_z = \sum_{i:Z_i=z} x_i$.

The form of the posterior pdfs for the scale parameters of the component densities ν_k, conditional on the other parameters and on $\mathbf{z}$, is not a common recognizable one. It is given by

$$f(\nu_z \,|\, \mathbf{x}, \mathbf{z}, \mathbf{w}, \boldsymbol{\mu}) \propto \frac{\nu_z^{n_z \nu_z}}{[\Gamma(\nu_z)]^{n_z}} \exp\left\{ -\nu_z \left(\theta + \frac{S_z}{\mu_z} + n_z \log \mu_z - \log P_z \right) \right\} \tag{16.9}$$

where $P_z = \prod_{i:Z_i=z} x_i$.

A Gibbs sampling algorithm that iterates sequentially and iteratively through the conditional posterior pdfs can proceed as follows:

1. Set $t = 0$ and initial values $\mathbf{w}^{(0)}, \boldsymbol{\mu}^{(0)}, \boldsymbol{\nu}^{(0)}$.
2. Generate $\left(\mathbf{z}^{(t+1)} \,|\, \mathbf{x}, \mathbf{w}^{(t)}, \boldsymbol{\nu}^{(t)}, \boldsymbol{\mu}^{(t)} \right)$ from (16.6).
3. Generate $\left(\mathbf{w}^{(t+1)} \,|\, \mathbf{x},\ \mathbf{z}^{(t+1)}, \boldsymbol{\nu}^{(t)}, \boldsymbol{\mu}^{(t)} \right)$ from (16.7).
4. Generate $\left(\mu_z^{(t+1)} \,|\, \mathbf{x}, \mathbf{z}^{(t+1)}, \boldsymbol{\nu}^{(t)}, \mathbf{w}^{(t+1)} \right)$ from (16.8), for $z = 1, 2, \ldots, k$.
5. Generate $\left(\nu_z^{(t+1)} \,|\, \mathbf{x}, \mathbf{z}^{(t+1)}, \boldsymbol{\mu}^{(t+1)}, \mathbf{w}^{(t+1)} \right)$ from (16.9), for $z = 1, 2, \ldots, k$.

6. Order the elements for $\boldsymbol{\mu}^{(t+1)}$ such that $\mu_1 < \mu_2 < \ldots < \mu_k$ and sort $\mathbf{w}^{(t+1)}$ and $\boldsymbol{\nu}^{(t+1)}$ accordingly.
7. Set $t = t+1$ and return to step 2.

To describe each of the generation steps in more detail, first consider (16.6). In this case we divide the interval (0,1) into k sub-intervals with the length of the z-th sub-interval equal to $P(Z_i = z \,|\, \mathbf{x}, \mathbf{w}, \boldsymbol{\nu}, \boldsymbol{\mu})$. A uniform random number is generated from the (0,1) interval. The value assigned to Z_i is the sub-interval in which the uniform random number falls. To generate observations from the Dirichlet density in (16.7), we first generate k gamma random variables, say γ_z, $z = 1,2,\ldots,k$ from $G(\phi_z + n_z,\, 1)$ densities, and then set $w_z = \gamma_z \Big/ \sum_{j=1}^{k} \gamma_j$. To generate μ_z from (16.8), we generate a random variable from a $G(\alpha_z + n_z v_z,\; \beta_z + S_z v_z)$ density and then invert it.

Generating v_z from equation (16.9) is more complicated, requiring a Metropolis step. We draw a candidate $\tilde{v}_z^{(t+1)}$ from a gamma density with mean equal to the previous draw $v_z^{(t)}$. That is, a candidate $\tilde{v}_z^{(t+1)}$ is generated from a $G\left(r,\, r\Big/v_z^{(t)}\right)$ distribution and is accepted as $v_z^{(t+1)}$ with probability

$$\min\left\{1,\, \frac{f\left(\tilde{v}_z^{(t+1)} \,|\, \mathbf{x}, \mathbf{z}^{(t+1)}, \mathbf{w}^{(t+1)}, \boldsymbol{\mu}^{(t+1)}\right)}{f\left(v_z^{(t)} \,|\, \mathbf{x}, \mathbf{z}^{(t+1)}, \mathbf{w}^{(t+1)}, \boldsymbol{\mu}^{(t+1)}\right)} \frac{p\left(\tilde{v}_z^{(t+1)}, v_z^{(t)}\right)}{p\left(v_z^{(t)}, \tilde{v}_z^{(t+1)}\right)}\right\}$$

where $p\left(v_z^{(t)}, \tilde{v}_z^{(t+1)}\right)$ is the gamma density used to generate $\tilde{v}_z^{(t+1)}$. Non-acceptance of $\tilde{v}_z^{(t+1)}$ implies $v_z^{(t+1)} = v_z^{(t)}$. The value of r is chosen by experimentation to give an acceptance rate of approximately 0.4.

3 Data Characteristics and Prior Parameters.

Characteristics of the sample of incomes from the 1978 Canadian Family Expenditure Survey are presented in Figure 16.1. The units are thousands of Canadian dollars. There are 8526 observations with values ranging from 0.281 to 173.8. Sample mean income is 35.5 and the sample median income is 32.4. The histogram reveals two modes, one at approximately 23 and the other at approximately 32. The Gini coefficient computed from the sample is 0.3358.

In choosing values for the parameters of the prior densities, our objective was to have proper but relatively uninformative priors so that posterior densities would be dominated by the sample data. We initially tried a mixture of $k = 3$ components but encountered identification problems and then reduced the number of components to $k = 2$.

We set $\phi_z = 1$ for all z, thus implying a uniform prior for the weights on each component. For the exponential prior on the scale parameters v_z we set $\theta = 0.02$. A 95% probability interval for this prior is (0.5, 161) implying a large range of values

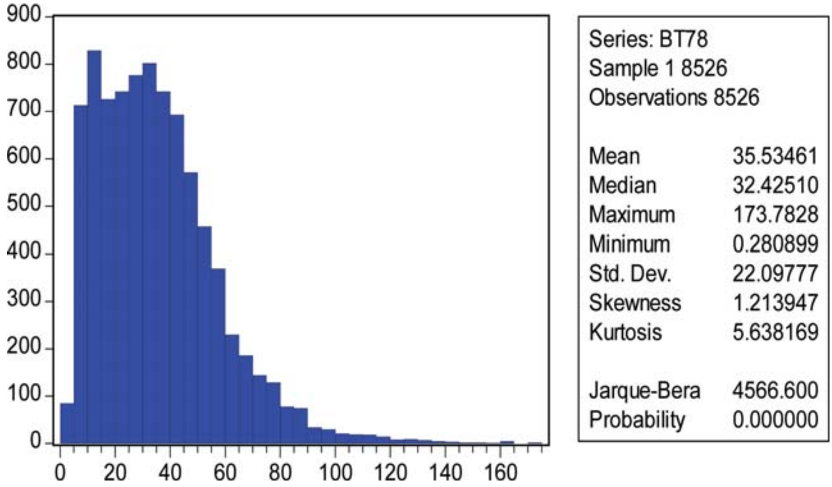


Fig. 16.1: Characteristics of Canadian income data

are possible. For the μ_z we initially set $\alpha_z = 2.2$ for $z = 1,2,3$ and $\beta_1 = 24$, $\beta_2 = 54$, $\beta_3 = 120$. Then, when we proceeded with $k = 2$, we set $\beta_1 = 30$ and $\beta_2 = 95$. From this latter setting, and ignoring the truncation $\mu_1 < \mu_2$, 95% prior probability intervals for μ_1 and μ_2 are, respectively, (5, 98) and (16, 306). In light of the sample mean of 35.5, these intervals suggest priors that are relatively uninformative.

4 Results

The algorithm described in Section 2 was used to generate 200,000 observations from the joint posterior density for the parameters $(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\nu})$ and the first 100,000 were discarded as a burn in. In our first attempts with $k = 3$ there appeared to be an identification problem with the second and third components. For separate identification of these two components, we require $\mu_2 < \mu_3$. If $\mu_2 = \mu_3$, some other mechanism is required for identification (Wiper *et al.* 2001). The two-dimensional plot of the draws for μ_2 and μ_3 given in Figure 16.2 shows a large number of observations on the boundary where $\mu_2 = \mu_3$. Other evidence is the bimodal distributions for v_2 and v_3 (Figure 16.3), the very high correlation between w_2 and w_3 (Figure 16.4) and the fact that the marginal posterior densities for w_2 and w_3 were mirror images of each other.

These issues led us to consider instead a model with two components ($k = 2$). In this case there was no apparent identification problem, and the Gibbs sampler

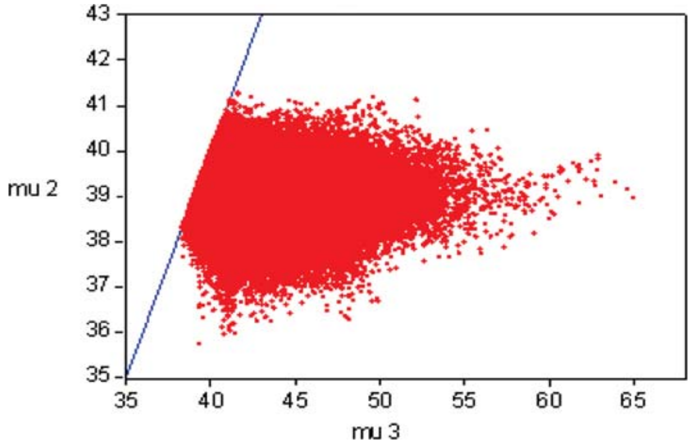


Fig. 16.2: Posterior observations on μ_2 and μ_3 for $k = 3$

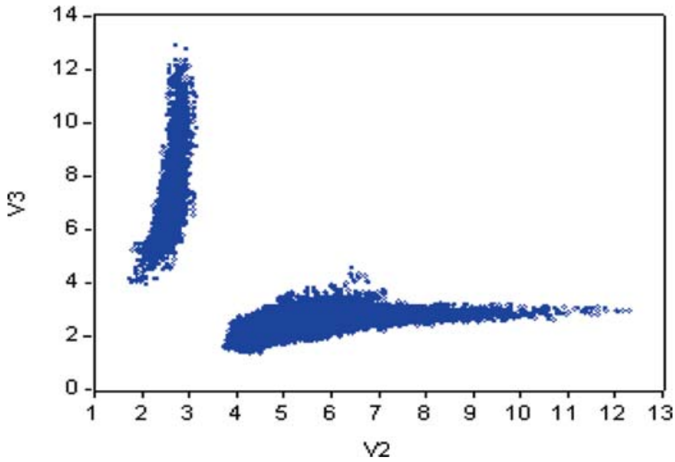


Fig. 16.3: Posterior observations on v_2 and v_3 for $k = 3$

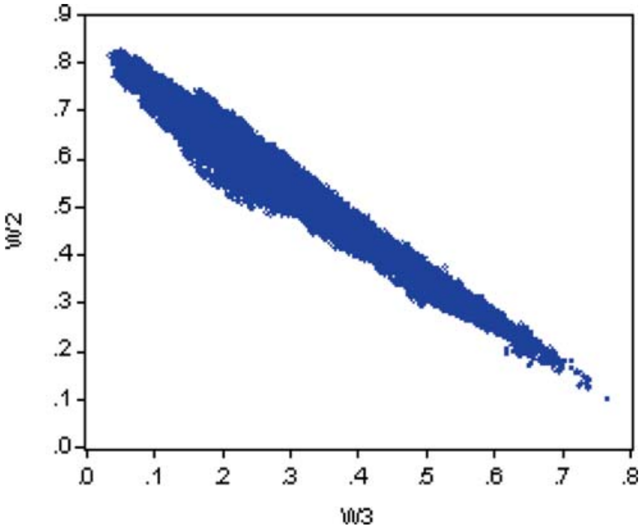


Fig. 16.4: Posterior observations on w_2 and w_3 for $k = 3$

showed evidence of converging. Summary statistics for the draws on the parameters are given in Table 16.1. There is relatively large weight (about 0.9) on the second component and a relatively small weight (about 0.1) on the first component. The posterior mean for the mean of the first component is relatively small (compared to the sample mean) and, likely, serves to help capture the first mode of the income distribution.

Table 16.1: Posterior Summary Statistics for Parameters

Name	Mean	St.Dev	Min	Max
μ_1	9.6134	0.3569	7.8906	11.130
μ_2	38.704	0.4207	36.903	40.768
w_1	0.1090	0.0121	0.0608	0.1566
w_2	0.8910	0.0121	0.8434	0.9392
v_1	7.4761	0.8087	5.2314	12.653
v_2	3.3616	0.1198	2.9667	3.9985

Having obtained M MCMC-generated observations from the posterior density $f(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\nu} | \mathbf{x})$, for a sample of observations $\mathbf{x}$ we can proceed to obtain estimates for the density and distribution functions for income and for the corresponding Lorenz curve as well as probability bands around these functions. Indexing an

MCMC-generated observation by a superscript (j), an estimate for the density function at a given income x is given by

$$f(x \mid \mathbf{x}) = \frac{1}{M} \sum_{j=1}^{M} \sum_{z=1}^{k} w_z^{(j)} G\left(x \mid \nu_z^{(j)}, \nu_z^{(j)} / \mu_z^{(j)}\right) \tag{16.10}$$

This function was calculated for 101 values of x from 0 to 200 such that the intervals between successive values of log x were equal. For each x 95% probability bands were found by sorting the M values of

$$f(x \mid \mathbf{w}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}) = \sum_{z=1}^{k} w_z^{(j)} G\left(x \mid \nu_z^{(j)}, \nu_z^{(j)} / \mu_z^{(j)}\right)$$

and taking the 0.025 and 0.975 percentiles of these values. The plots for the mean distribution and its probability bounds appear in Figure 16.5. The bimodal nature of the distribution has been well captured, although, as one would expect, it is at the peaks of the distribution where the greatest uncertainty is exhibited through wider bounds.

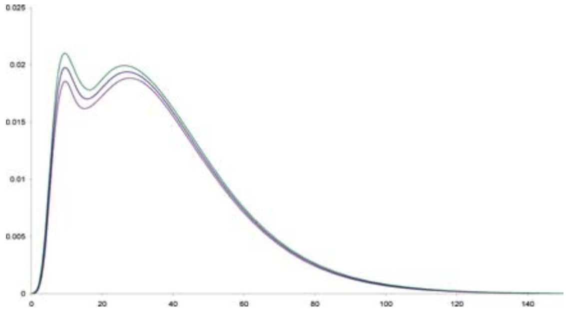


Fig. 16.5: Mean and 95% probability bounds for the predictive density for income

An estimate of the distribution function and probability bounds on that distribution can be found in a similar way. In this case the value of the distribution function for a given value x is given by

$$\begin{aligned} F(x \mid \mathbf{x}) &= \frac{1}{M} \sum_{j=1}^{M} \sum_{z=1}^{k} w_z^{(j)} \int_0^x G\left(t \mid \nu_z^{(j)}, \nu_z^{(j)} / \mu_z^{(j)}\right) dt \\ &= \frac{1}{M} \sum_{j=1}^{M} F(x \mid \mathbf{w}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}) \end{aligned} \tag{16.11}$$

This function was evaluated for the same 101 values of x. To estimate the Lorenz curve we consider for each x the M points $F\left(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)}\right)$ and the corresponding points for the first moment distribution which is given by

$$\begin{aligned}\eta\left(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)}\right) &= \frac{1}{\sum_{h=1}^{k} w_h^{(j)}\mu_h^{(j)}}\int_0^x tf\left(t\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)}\right)dt \\ &= \frac{1}{\sum_{h=1}^{k} w_h^{(j)}\mu_h^{(j)}}\sum_{z=1}^{k} w_z^{(j)}\int_0^x tG\left(t\,|\,\nu_z^{(j)},\nu_z^{(j)}/\mu_z^{(j)}\right)dt \\ &= \frac{1}{\sum_{h=1}^{k} w_h^{(j)}\mu_h^{(j)}}\sum_{z=1}^{k} w_z^{(j)}\mu_z^{(j)}\int_0^x G\left(t\,|\,(\nu_z^{(j)}+1),\nu_z^{(j)}/\mu_z^{(j)}\right)dt\end{aligned} \tag{16.12}$$

To see how to use these points to estimate a Lorenz curve and find its probability bounds it is instructive to examine a graph of the M points for $F(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})$ and $\eta(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})$ for a given value of x. Such a graph for the point $x = 35$ is given in Figure 16.6. A graph like that in Figure 16.6 could be drawn for each of the 101 x points. To estimate the Lorenz curve and draw probability bounds around it, we need to "select" three points from each graph, an estimate of the Lorenz curve for each x and its corresponding upper and lower probability bounds. As an estimate of the Lorenz curve for a given x we can take the mean values of all the points in Figure 16.6. That is, the point $[\eta(x\,|\,\mathbf{x}),F(x\,|\,\mathbf{x})]$ where

$$\eta(x\,|\,\mathbf{x}) = \frac{1}{M}\sum_{j=1}^{M}\eta(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)}) \tag{16.13}$$

and $F(x\,|\,\mathbf{x})$ is given in (16.11). Then an estimate of the complete Lorenz curve is obtained by joining these points for all x.

Finding 95% probability bounds for the Lorenz curve is more difficult than it is for the density and distribution functions because, for each x, we have a 2-dimensional space for $F(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})$ and $\eta(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})$ to consider. Two approaches were taken. In the first, for each x, we regressed the M values of $\eta(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})$ on the corresponding M values of $F(x\,|\,\mathbf{w}^{(j)},\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})$ via a least squares regression. The residuals from this regression were ordered and the 0.025 and 0.975 percentiles of the residuals were noted. Denoting them by $\hat{e}_{.025}$ and $\hat{e}_{.975}$, the bounds at a given x were taken as the points

$$[F(x\,|\,\mathbf{x}),\;\;\eta(x\,|\,\mathbf{x})+\hat{e}_{.025}] \qquad \text{and} \qquad [F(x\,|\,\mathbf{x}),\;\;\eta(x\,|\,\mathbf{x})+\hat{e}_{.925}] \tag{16.14}$$

Note that $\hat{e}_{.025} < 0$, so we add it rather than subtract it from $\eta(x\,|\,\mathbf{x})$. To obtain the lower bound on the Lorenz curve, we computed $[F(x\,|\,\mathbf{x}),\;\;\eta(x\,|\,\mathbf{x})+\hat{e}_{.025}]$ for each x, and joined these points. Similarly, to obtain the upper Lorenz bound, we computed $[F(x\,|\,\mathbf{x}),\;\;\eta(x\,|\,\mathbf{x})+\hat{e}_{.925}]$ for each x and joined these points. These bounds

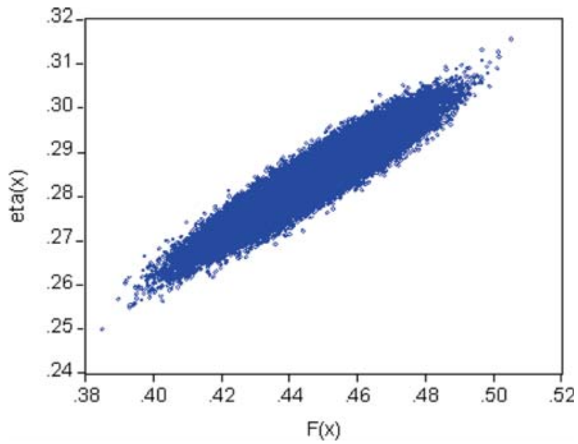


Fig. 16.6: Plots of 100,000 pairs of $F(x)$ and $\eta(x)$ for $x = 35$

and the estimated Lorenz curve are plotted in Figure 16.7. However, the bounds are so narrow that they are indistinguishable from the estimated curve. In Figure 16.8 we present a more distinct cross section of the plots for $0.4 < F(x) < 0.6$ and $0.2 < \eta(x) < 0.4$. Also, to give an idea of the width of the bounds, in Figure 16.9 we plot $\hat{e}_{.025}$ and $\hat{e}_{.975}$ against $F(x)$. The maximum width of the probability interval is less than 0.008, implying the Lorenz curve is accurately estimated.

To introduce our second approach for finding probability bounds on the Lorenz curve, first note that, in the first approach, the bounds do not correspond to one set of parameter values for all x. The upper and lower extreme 2.5% of parameter values is likely to be different for each x setting. While this is not necessarily a bad thing – it is also a characteristic of the estimated density function for income – it is interesting to examine an alternative method of obtaining bounds that "discards" the same parameter values for each x. One way to use a unique set of upper and lower 2.5% of parameter values is to order Lorenz curves on the basis of their Gini coefficients. Denoting the 101 x points as $x_1, x_2, \ldots, x_{101}$, the Gini coefficient for the j-th set of parameters can be approximated by

$$\begin{aligned}\text{Gini}\left(\mathbf{w}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}\right) = &\sum_{m=1}^{100} \eta\left(x_{m+1} \,|\, \mathbf{w}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}\right) F\left(x_m \,|\, \mathbf{w}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}\right) \\ &- \sum_{m=1}^{100} \eta\left(x_m \,|\, \mathbf{w}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}\right) F\left(x_{m+1} \,|\, \mathbf{w}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}\right)\end{aligned} \tag{16.15}$$

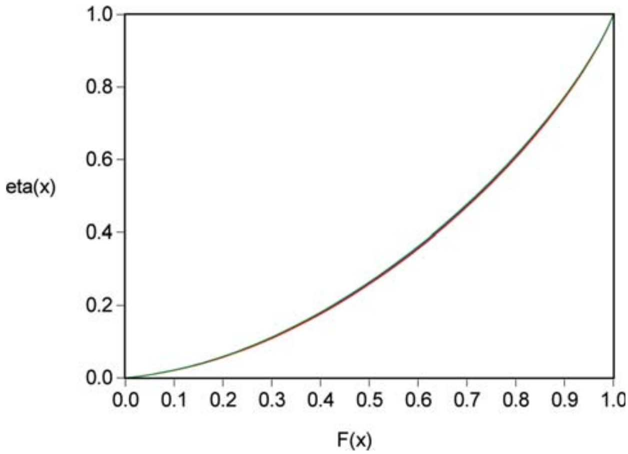


Fig. 16.7: Entire Lorenz curve and 95% probability bounds

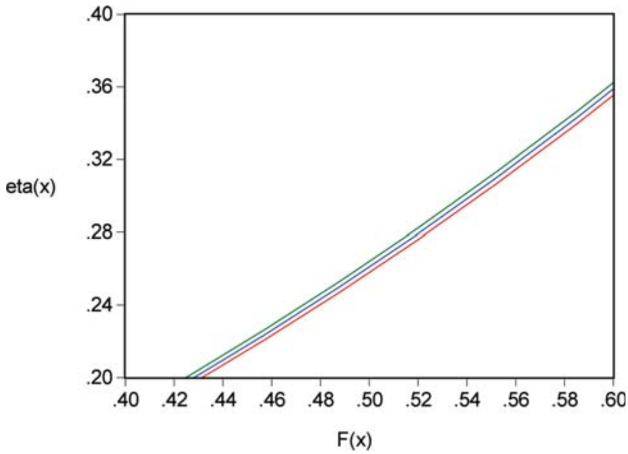


Fig. 16.8: Close up of Lorenz curve and 95% probability bounds

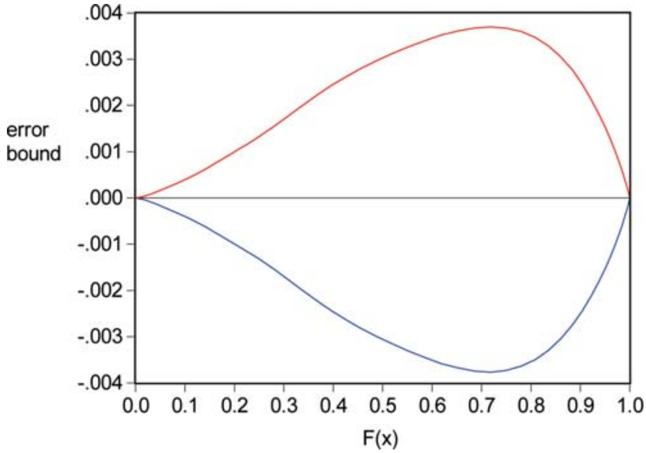


Fig. 16.9: Plots of the differences between the estimated Lorenz curve and 95% probability bounds

In this approach the probability bounds of the Lorenz curve were taken as the Lorenz curves corresponding to the parameter values that yield the 0.025 and 0.975 percentiles for the Gini coefficient. Thus, the bounds on the Lorenz curve are found by using the area under the Lorenz curve to determine a parameter ordering. Specifically, if the parameter values corresponding to the 0.025 and 0.975 percentiles of the Gini coefficient are $(\mathbf{w}_{.025}, \boldsymbol{\mu}_{.025}, \boldsymbol{\nu}_{.025})$ and $(\mathbf{w}_{.975}, \boldsymbol{\mu}_{.975}, \boldsymbol{\nu}_{.975})$, then the upper bound is the curve joining the points $[\eta\,(x\,|\,\mathbf{w}_{.975}, \boldsymbol{\mu}_{.975}, \boldsymbol{\nu}_{.975})\,, F(x\,|\,\mathbf{w}_{.975}, \boldsymbol{\mu}_{.975}, \boldsymbol{\nu}_{.975})]$ for each x, and the lower bound is the curve joining the points $[\eta\,(x\,|\,\mathbf{w}_{.025}, \boldsymbol{\mu}_{.025}, \boldsymbol{\nu}_{.025})\,, F(x\,|\,\mathbf{w}_{.025}, \boldsymbol{\mu}_{.025}, \boldsymbol{\nu}_{.025})]$ for each x.

While it is straightforward to draw the bounds in this way, it is not obvious how one might define the "errors" between the estimated Lorenz curve and its 95% probability bounds if one is interested in these values. In the regression approach, where η was treated as the "dependent" variable and F was treated as the "explanatory" variable, it was natural to define the errors as the vertical distances as specified in (16.14). In this case, however, there is no reason why they should be vertical or horizontal distances. To solve this dilemma, we define the errors as the orthogonal distances from the Lorenz curve

$$\hat{d}_U(x) = \sqrt{(F(x\,|\,\mathbf{w}_{.975}, \boldsymbol{\mu}_{.975}, \boldsymbol{\nu}_{.975}) - F(x\,|\,\mathbf{x}))^2 + (\eta\,(x\,|\,\mathbf{w}_{.975}, \boldsymbol{\mu}_{.975}, \boldsymbol{\nu}_{.975}) - \eta(x\,|\,\mathbf{x}))^2}$$

$$\hat{d}_L(x) = \sqrt{(F(x\,|\,\mathbf{w}_{.025}, \boldsymbol{\mu}_{.025}, \boldsymbol{\nu}_{.025}) - F(x\,|\,\mathbf{x}))^2 + (\eta\,(x\,|\,\mathbf{w}_{.025}, \boldsymbol{\mu}_{.025}, \boldsymbol{\nu}_{.025}) - \eta(x\,|\,\mathbf{x}))^2}$$

Once again, it turned out that the Lorenz curve is estimated very accurately with the probability bounds not discernible from the mean Lorenz curve. Rather than present another figure that appears identical to Figure 16.7, in this case we simply plot the errors $\hat{d}_U(x)$ and $\hat{d}_L(x)$ that appear in Figure 16.10. The pattern of these differences is a strange one, and, as expected, they are larger than those obtained using the regression method. Larger differences are expected because the regression method minimizes the "error" for each x. Nevertheless, the largest error is still relatively small, being less than 0.016.

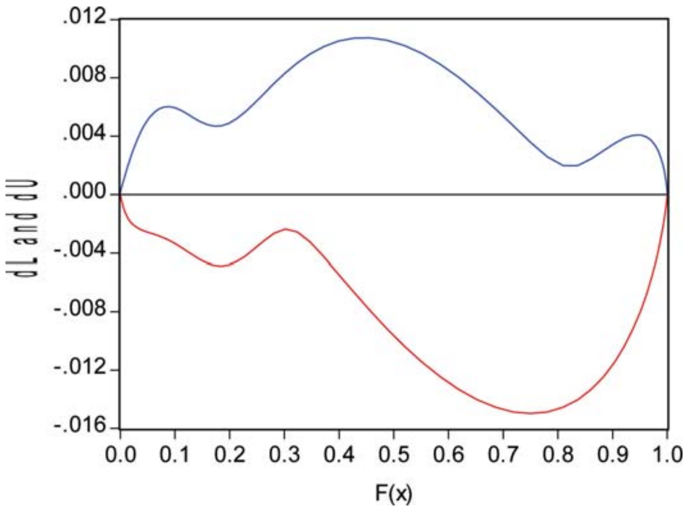


Fig. 16.10: Orthogonal differences between Lorenz curve and 95% probability bounds

Also, of interest is the Gini coefficient. Its posterior density, estimated from the 100,000 points defined by equation (16.15), is plotted in Figure 16.11. The posterior mean is 0.337 and 95% probability bounds for the Gini coefficient are 0.333 and 0.342.

5 Goodness of Fit

Given our objective was to specify a gamma mixture as a flexible parametric model for an income distribution, it is useful to assess its goodness of fit against those of some common income distributions. To do so we compare the estimated distribution function $F(x|\mathbf{x})$ with the empirical distribution function $F_0(x_j) = j/n$ where j refers to the j-th observation after ordering them from lowest to highest and n is the sample size. We compute goodness of fit using the root mean squared error

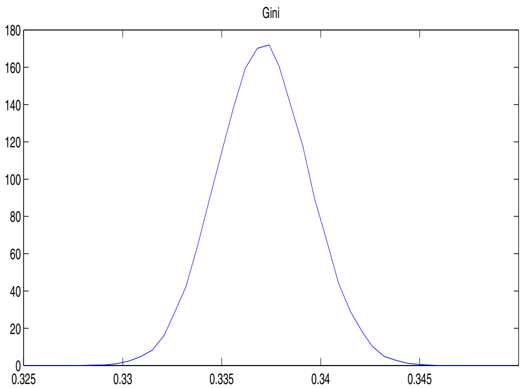


Fig. 16.11: Posterior density for the Gini coefficient

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(F\left(x_j\,|\,\mathbf{x}\right) - F_0(x_j)\right)^2}$$

In addition we perform a Kolmogorov-Smirnov test which is based on the largest difference between $F\left(x_j\,|\,\mathbf{x}\right)$ and $F_0(x_j)$. Table 16.2 contains the results for the Bayesian-estimated gamma mixture and for maximum likelihood estimates of the lognormal, beta2, Singh-Maddala and Dagum distributions. Clearly, the gamma mixture is far superior to other models in terms of goodness of fit.

Table 16.2: Goodness of Fit Comparisons

	RMSE	Max Dif(δ_n)	$\delta_n\sqrt{n}$	p-value
Gamma Mix	0.0064	0.01449	1.33795	0.055738
Log Normal	0.0414	0.07449	6.87813	0.000000
Beta2	0.0310	0.05523	5.09974	0.000000
Singh-Maddala	0.0122	0.02757	2.54571	0.000005
Dagum	0.0135	0.03146	2.90490	0.000000

6 Concluding Remarks

A mixture of gamma densities has been suggested as a model for income distributions. Mixtures have the advantage of providing a relatively flexible functional form and at the same time they retain the advantages of parametric forms that are amenable to inference. We have demonstrated how a Bayesian framework can be utilized to estimate the gamma mixture and related quantities relevant for income distributions. In addition to showing how the income distribution estimate and its 95% probability bounds can be calculated, we considered the distribution function, the Lorenz curve and the Gini coefficient. Two ways of computing 95% probability bounds for the Lorenz curve were explored. Goodness-of-fit comparisons showed the gamma mixture fits well compared to a number of commonly used income distributions.

An attempt to estimate a mixture with 3 components was not successful leading us to opt for a model with 2 components. The results for 3 components suggested a lack of identification between the second and third components. Most likely, the empirical characteristics of the distribution are well captured by 2 components, making it hard for the data to discriminate when 3 are specified. This outcome does not necessarily imply 2 will always be adequate. There could be other distributions where more components improve the specification. Also, the number of components can be treated as an unknown parameter which, in a Gibbs sampling algorithm, can vary from iteration to iteration.

Further research will focus on the use of estimated gamma mixtures in the measurement of inequality and poverty and in methodology for examining stochastic and Lorenz dominance for income distributions. Expressing uncertainty about such quantities in terms of posterior densities facilitates making inferences and probability statements about relative welfare scenarios.

References

Barret, G. F. and S. G. Donald (2003) Consisitent Tests for Stochastic Dominance, *Econometrica*, **71**(1), 71–104.

Chotikapanich, D. and W. E. Griffiths (2006) Bayesian Assessment of Lorenz and Stochastic Dominance in Income Distributions, in J. Creedy and G. Kalb (eds.) *Research on Economic Inequality, vol. 13: Dynamics of Inequality and Poverty*, pp. 297–321, Elsevier, Amsterdam.

Chotikapanich, D., W. E. Griffiths and D. S. P. Rao (2007) Estimating and Combining National Income Distributions Using Limited Data, *Journal of Business and Economics Statistics*, **25**, 97–109.

Cowell, F. A. (1999) Estimation of Inequality Indices, in J. Silber (ed.) *Handbook on Income Inequality Measurement*, Kluwer Academic Publishers, London.

Hasegawa, H. and H. Kozumi (2003) Estimation of Lorenz Curves: A Bayesian Nonparametric Approach, *Journal of Econometrics*, **115**, 277–291.

Kakwani, N. (1999) Inequality, Welfare and Poverty: Three Interrelated Phenomena, in J. Silber (ed.) *Handbook on Income Inequality Measurement*, Kluwer Academic Publishers, London.

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

Silber, J. (1999) *Handbook on Income Inequality Measurement*, Kluwer Academic Publishers, London.

Wiper, M., D. Rios Insua and F. Ruggeri (2001) Mixtures of Gamma Distributions with Applications, *Journal of Computational and Graphical Statistics*, **10**, 440–454.

CHAPTER 17

Inequality in Multidimensional Indicators of Well-Being: Methodology and Application to the Human Development Index*

Quentin Wodon† and Shlomo Yitzhaki‡

Abstract

Inequality measures for multidimensional indicators of well-being may be sensitive to the weights used for weighting the various dimensions taken into account in the overall indicator. This paper provides a general yet simple method for assessing whether the Gini index of inequality is sensitive to changes in weights. The method is applied to the Human Development Index (*HDI*). Changing the weights used to compute the *HDI* would not change significantly world inequality in human development.

1 Introduction

In order to provide measures of well-being which take into account several dimensions of well-being, multidimensional indices have been proposed in the literature. A well known example is the Human Development Index (*HDI* hereafter) published

---

* The authors benefited from comments of the editor, two anonymous referees and participants to the World Bank's Thematic group on inequality. The views expressed here are those of the authors and need not reflect those of the World Bank, its Executive Directors or the countries they represent.

† World Bank, Washington D.C. E-mail: qwodon@worldbank.org.

‡ Department of Economics, The Hebrew University, Jerusalem, Israel.

by the United Nations Development Programme (2006). The *HDI* is a weighted average of three other indices dealing respectively with life expectancy, educational attainment, and per capita GDP. Each of the three indices receives an equal weight of one third in the *HDI*. These weights (and to some extent the variables themselves) are arbitrary, so that the results obtained (for example, in terms of measuring worldwide between country inequality in human development) may be sensitive to the choice of the weights. The objective of this paper is to assess to what extent this is the case.

The choice of the variables used in defining the *HDI* and their weights has been discussed extensively, among others by Doessel and Gounder (1994); Hicks (1997); Lüchters and Menkhoff (1996); McGillivray (1991); McGillivray and White (1993); McGillivray and Noorbakhsh (2004); Noorbakhsh (1998, 2007); Ogwang (1994); Palazzi and Lauri (1998) and Ram (1992). One first set of issues revolves around the choice of the variables to be used. As noted by McGillivray and Noorbakhsh (2004), indices that were developed before the *HDI*, such as the General Index of Development (GID; see McGranahan *et al.* (1972)) and the Socio-Economic Development Index (United Nations Research Institute for Social Development (UNRISD), 1970) were criticized because they were closer to measures of structural activity rather than well-being. The choice of the variables used in the *HDI* was made in order to try to capture Sen's concept of capabilities. Still, this choice of variables remains somewhat ad hoc, and it is not clear that this can be easily avoided, not only because different observers may have an interest in different variables, but also from a practical point of view because there are constraints in terms of what can be properly measured at regular intervals by developing countries in order to be incorporated in the overall index. Difficulties in measurement are also one of the reasons why it is easier to work with country means for the selected variables (or more precisely, in the case of the *HDI*, transformations of these means taking into account minimum and maximum thresholds that the variables can reach), rather than with more complex indicators that would take into account within country inequality in indicators.

A second set of issues relates to the weight to be used in order to combine the chosen indicators in one aggregate measure of well-being. Quite a few studies have been devoted to assessing the correlations between various composite indices of well-being, between the various variables used in the *HDI* or other indices, and between the variables and the *HDI* or the indices themselves. For example, Ram (1992) calculates correlation coefficients between the *HDI* and real income inequalities and finds them to be very high. However, he notes that despite high correlations, inequality measures differ markedly, with inequality in the *HDI* being much lower than in real GDP per capita.[36]

As noted by McGillivray and Noorbakhsh (2004), when comparing different composite indices, "a fundamental weakness with these studies is that it is not entirely clear what extent of statistical association deems a new indicator empirically redundant with respect to a pre-existing one". More generally, a key issue when

---

[36] For a recent analysis of inequality in *HDI*, see Noorbakhsh (2007).

assessing weighting schemes is that the assessment of the appropriate weights, or the assessment of whether changes in weights make a substantial difference in the results obtained, depends on the purpose of the analysis. For example, in the context of work aiming to assess the level of inequality in human development, perhaps with the aim to subsequently measure whether there is convergence over time or not between countries in average levels of human development (as done by Noorbakhsh (2007)), one might be interested in testing first whether changes in the weighting system affect between-country inequality in the *HDI*, because this may in turn affect conclusions reached about convergence.

In this paper, building on work by Wodon and Yitzhaki (2003) and Yitzhaki and Wodon (2004), we provide a general yet simple method for assessing whether inequality measures for multidimensional indicators of well-being are sensitive to the weights used for the various dimensions taken into account in the overall indicator. The basic idea is the following: the higher the association between the variables included in the index is, the less important is the weighting scheme used for aggregating the various components of the index when estimating inequality in the overall index. However, measures of association may differ in their values and therefore it is important to find out the "appropriate" measure of association to be used in order to translate the association between the various variables included in the index into a correlation measure. This "appropriate" correlation indeed depends on the choice of the inequality measure (such as the Gini index, which requires the use of the Gini correlation to measure the association between the variables used for the index). When applying our method to the *HDI* and estimating the Gini index for the *HDI*, we are able to predict fairly accurately how inequality in human development would be affected by changes in weights for the various components of the *HDI*. We also find that the impact of changing weights on inequality is not very large.

2 Methodology

Let (Z_1, Z_2) be a bivariate distribution, with $F(Z_1)$ and $F(Z_2)$ as marginal cumulative distributions. It is assumed that first and second moments exist. We want to estimate the Gini index of inequality of a linear combination of Z_1 and Z_2. To simplify the derivation, we work with variables with unit means, $Y_i = Z_i/\mu_i$. To perform such an exercise we need the equivalent of the correlation coefficient which is appropriate for Gini analysis. Schechtman and Yitzhaki (1987, 1999) define the asymmetric Gini correlation between Y_1 and Y_2 as:

$$\Gamma_{12} = \mathrm{cov}(Y_1, F(Y_2))/\mathrm{cov}(Y_1, F(Y_1)) \tag{17.1}$$

The Gini correlation is asymmetric because Γ_{12} need not be equal to Γ_{21}, although both are bounded by minus one and one. If each observation k in distribution 2 is obtained by applying a monotonic increasing transformation t() on distribution 1, with $y_{2k} = t(y_{1k})$, then $\Gamma_{12} = \Gamma_{21} = 1$. Under rank reversal (the largest value in distribution 1 becomes the lowest in distribution 2, etc.), $\Gamma_{12} = \Gamma_{21} = -1$. If Y_1 and Y_2 are independent, $\Gamma_{12} = \Gamma_{21} = 0$. Although Γ_{12} need not equal Γ_{21}, one can define a symmetric Gini correlation as:

$$\Gamma = w_1\Gamma_{12} + w_2\Gamma_{21} \tag{17.2}$$

where $w_i = G_i/(G_1 + G_2), i = 1,2$, are the shares of inequality of each distribution in the sum for both distributions. Changing the role of the distributions does not affect Γ.

Let $Y_\alpha = \alpha Y_1 + (1-\alpha)Y_2$ where $0 < \alpha < 1$ is a constant. If α is known, the Gini of Y_α, denoted by G_α, can be computed directly by creating a new variable, Y_α. If α is not known with certainty, it is useful to provide lower and upper bounds for G_α. Denoting by $F(Y_\alpha)$ the cumulative distribution of Y_α, $G_\alpha = 2\text{cov}(Y_\alpha, F(Y_\alpha))$ since all variables have unit means (on this covariance formula, see Lerman and Yitzhaki, 1994). This implies:

$$G_\alpha = 2\alpha\text{cov}(Y_1, F(Y_\alpha)) + 2(1-\alpha)\text{cov}(Y_2, F(Y_\alpha)) \tag{17.3}$$

Yitzhaki and Wodon (2004) show that (17.3) is bounded by:

$$\text{Max}[0, \alpha G_1\Gamma_{12} + (1-\alpha)G_2\Gamma_{21}] \leq G_\alpha \leq \alpha G_1 + (1-\alpha)G2 \tag{17.4}$$

The proof is reproduced in appendix. For example, if Γ_{12} = 0.8, $\Gamma_{21} = 0.9, G_1 =$ 0.4, and $G_2 = 0.3$, then 0.27 + 0.05 $\alpha \leq G_\alpha \leq$ 0.3 + 0.1 α, which is an error range of less than 10 percent. The upper bound is reached when one variable is a monotonic increasing transformation of the other, so that the ranking of observations in distributions 2 and 1 are identical. If $F(Y_\alpha) = F(Y_1) = F(Y_2)$, G_α is simply the weighted average of G_1 and G_2. The lower bound depends on the Gini correlations, with low Gini correlations reducing the potential value of G_α, up to a minimum value of zero since the Gini coefficient is non-negative. Under negative Gini correlations, the variables "neutralize" each other. However, in most cases of interest, the variables will exhibit positive correlations, so that the binding factor will be the second term in the maximum function in (17.4). A special case occurs when Y_1 and Y_2 are exchangeable[37], as is the case for the bivariate normal distribution. Then, $\Gamma = \Gamma_{12} = \Gamma_{21}$. If Y_α and Y_1 are also exchangeable, as well as Y_α and Y_2, then it can be shown (the proof is reproduced in appendix) that:

$$G_\alpha^2 = \alpha^2 G_1^2 + (1-\alpha)^2 G_2^2 + 2\alpha(1-\alpha)G_1G_2\Gamma. \tag{17.5}$$

Under exchangeability, the Gini behaves like the coefficient of variation. So far, we worked with variables with unit means. To handle variables with non-equal

[37] A set of random variables is said to be exchangeable if, for all $n \geq 1$, and for every permutation of n subscripts, the joint distributions are identical (Stuart and Ord, 1987). For the two Gini correlations to be equal ($\Gamma_{js} = \Gamma_{sj}$), it is required that the variables be exchangeable up to a linear transformation, a requirement which is weaker than exchangeability (Schechtman and Yitzhaki, 1987).

means, the Gini behaves like the coefficient of variation, so that it suffices to replace α by $a = \alpha\mu_1/(\alpha\mu_1 + (1-\alpha)\mu_2)$.

3 Application to the Human Development Index

To provide evidence on the good predictive power of the above upper and lower bounds in (17.4), and of the estimates obtained under the exchangeability assumption in (17.5), we estimate measures of between country inequality in human development, using the *HDI*. The data are for the year 2004 and come from the 2006 Human Development Report (United Nations Development Program, 2006). Specifically, the variables used for computing the Human Development Index come from Table 1 on pages 283-286, while the data used to compute the population weights in the covariances and Gini indices comes from Table 5 on pages 297-300[38].

We provide results for both population weighted and non-weighted samples. The *HDI* is a weighted sum of three indices themselves based on indicators. Denoting by X the value of the indicator, each index is computed using a formula taking into account the actual value of the indicator as well as fixed minimum and maximum values.[39] The formula is such that for each country, the value of each index is between zero and one. That is, for any given country, the indices are computed as:

$$\text{Index} = (\text{Actual } X - \text{Minimum } X)/(\text{Maximum } X - \text{Minimum } X) \tag{17.6}$$

The first index is that of life expectancy at birth, for which the maximum and minimum values are respectively 25 and 85 years. The second index is that of educational attainment. It is itself a weighted average of two indices or components. The first component is the adult literacy rate index for which the minimum and maximum values are 0 and 100 percent. The second component is the combined gross enrolment ratio index for primary, secondary, and tertiary education, with minimum and maximum values also fixed at 0 and 100 percent. In the *HDI* calculation, the adult literacy index and the combined gross enrolment ratio index are given weights of 2/3 and 1/3, so that the educational attainment index is simply the weighted arithmetic mean of its two components. Finally, the third index is that of the logarithm of

[38] Due to lack of data for some variables, we deleted from the sample a number of small countries: Bhutan, Cuba, Ecuador, Haiti, Maldives, Occupied Palestinian Territories, Suriname, Timor-Leste, and Turkmenistan. For countries for which data on the combined enrolment indicator were missing (there was a slightly larger number of such countries), values of one were given for the richest countries, and for other countries, the average of the values obtained for the three countries ranked (in terms of *HDI*) just above the country with missing data were used. These hypotheses do not affect the results reported in this paper in any meaningful way, and they were used in order to keep as many countries as possible in the sample. The final sample consists of 167 countries.

[39] We are grateful to an anonymous referee who pointed out the importance of stating that the maxima and minima are subjective goalposts assigned by UNDP. This is discussed among others by McGillivray and Noorbakhsh (2004).

real GDP per capita measured using Purchasing Power Parity values in U.S. dollars, with the minimum and maximum values set at log(100) and log(40,000). According to the UNDP report, income is used as a proxy for those dimensions of well-being that are not already reflected through the other components of the *HDI*, namely life expectancy and knowledge as measured by education outcomes. It is worth noting that the way in which income enters in the *HDI* index has been modified from earlier versions of the aggregate (this is discussed for example in Anand and Sen (1999)). The *HDI* index is then computed as the simple arithmetic mean of the three indices. Real GDP, life expectancy, and educational attainment are thus given equal weights of one third in the *HDI*.

For the purpose of our analysis, it is convenient to present the *HDI* in a slightly different way. In equation (17.7), the educational attainment index is a weighted sum of the adult literacy index (*ALI*) and the combined gross enrolment index (*CGEI*), with weights γ and one minus γ (for the UNDP, γ is set at two thirds). Next, we shall refer to the non-GDP index, denoted by *NGDPI*, as the weighted sum of the life expectancy (*LEI*) and educational attainment (*EAI*) indices, with weights β and one minus β (for the UNDP, β is set at 0.5). Finally, the *HDI* itself is a weighted sum of the real GDP and non-GDP indices, with weights α and one minus α (for the UNDP, given the way the formula is set up in equation (17.7), α is set at one third). In other words, in order to add only two indices at a time to get to the *HDI*, we have grouped the life expectancy and educational attainment indices together into a new non-real GDP index. This manipulation will facilitate the analysis, and it does not change any of our conclusions.

$$
\begin{aligned}
HDI &= \alpha GDPI + (1-\alpha)NGDPI \\
\text{With } NGDPI &= \beta LEI + (1-\beta)EAI \\
\text{and } EAI &= \gamma ALI + (1-\gamma)CGEI
\end{aligned}
\tag{17.7}
$$

Table 17.1 provides summary statistics for the variables used. The non-weighted values give an equal weight to all countries regardless of their population size. The weighted values use the population shares as weights. The mean values for the world as well as other summary statistics are given. There are some differences in both the indicators and the indices according to the weighting scheme, as expected. Typically, weighted indicators and indices have lower values than unweighted variables.

In what follows, we focus on the ability of our lower and upper bounds to predict the level of inequality in the *HDI* and its various components at the world level. Table 17.2 provides the estimates of the Gini indices of inequality as well as the asymmetric and symmetric Gini correlations. As mentioned in the previous section, the two asymmetric Gini correlations need not be equal, even though in practice they tend to be close to each other. Overall, the Gini correlations between the various dimensions of the *HDI* are high. The lowest correlation is observed for the relationship between life expectancy and educational attainment, which is 0.79.

Table 17.3 gives the test of the predictive power of the lower and upper bounds in equation (17.4), as well as the estimates under the exchangeability assumption in

Table 17.1: Summary values of the variables and indices

Variable	Min	Max	Unweighted		Weighted	
			Mean	Std.Dev.	Mean	Std.Dev.
Human Development Index	0.31	0.97	0.713	0.180	0.715	0.151
Life Expectancy	31.30	82.20	65.947	12.596	67.508	9.383
Adult Literacy	19.00	100.00	82.329	19.039	80.748	17.369
Gross Enrolment	21.00	113.00	70.383	18.652	69.697	15.395
Gross Domestic Product	561.00	69961.00	10442.390	11137.770	8968.777	10382.870
Life Expectancy Index	0.10	0.95	0.683	0.210	0.707	0.156
Enrolment Index	0.23	0.99	0.780	0.187	0.767	0.166
Gross Domestic Product Index	0.29	1.00	0.676	0.194	0.667	0.165
Population	0.10	1308.00	37.334	135.172	523.804	550.010

Source: Authors' estimation from UNDP (2006) data.

Table 17.2: Inequality and Gini correlations for the various components of the *HDI*

Real GDP and non-Real GDP indices		*Life expectancy and educational attainment indices*		*Adult literacy and combined gross enrolment indices*	
Non-weighted					
Mean for *NGDPI* (Y_2)	0.731	Mean for *EAI* (Y_2)	0.780	Mean for *CGEI* (Y_2)	0.704
Mean for *GDPI* (Y_1)	0.676	Mean for *LEI* (Y_1)	0.683	Mean for *ALI* (Y_1)	0.823
Gini for *NGDPI* (Y_2)	0.139	Gini for *EAI* (Y_2)	0.130	Gini for *CGEI* (Y_2)	0.149
Gini for *GDPI* (Y_1)	0.166	Gini for *LEI* (Y_1)	0.167	Gini for *ALI* (Y_1)	0.122
Γ_{21}	0.857	Γ_{21}	0.795	Γ_{21}	0.846
Γ_{12}	0.868	Γ_{12}	0.790	Γ_{12}	0.856
Γ	0.863	Γ	0.792	Γ	0.850
Weighted					
Mean for *NGDPI* (Y_2)	0.737	Mean for *EAI* (Y_2)	0.767	Mean for *CGEI* (Y_2)	0.697
Mean for *GDPI* (Y_1)	0.667	Mean for *LEI* (Y_1)	0.707	Mean for *ALI* (Y_1)	0.807
Gini for *NGDPI* (Y_2)	0.112	Gini for *EAI* (Y_2)	0.120	Gini for *CGEI* (Y_2)	0.121
Gini for *GDPI* (Y_1)	0.137	Gini for *LEI* (Y_1)	0.116	Gini for*ALI* (Y_1)	0.116
Γ_{21}	0.926	Γ_{21}	0.865	Γ_{21}	0.851
Γ_{12}	0.924	Γ_{12}	0.791	Γ_{12}	0.892
Γ	0.925	Γ	0.829	Γ	0.871

Source: Authors' estimation from UNDP (2006) data.

Table 17.3: Inequality bounds, mid-points, and actual values for the indices used in the *HDI*

	Para-meters	*UNDP parameter values* α	*Adjusted parameter values* a *(*)*	*Lower bound*	*Upper bound*	*Mid point*	*Estimate under exchan-geability*	*Actual Gini using UNDP parameter values*
Non-weighted								
GDPI and *NGDPI*	α	0.333	0.317	0.127	0.147	0.137	0.143	0.143
LEI and *EAI*	β	0.500	0.480	0.117	0.148	0.132	0.140	0.139
ALI and *CGEI*	γ	0.667	0.701	0.111	0.130	0.120	0.125	0.126
Weighted								
GDPI and *NGDPI*	α	0.333	0.312	0.111	0.147	0.129	0.118	0.118
LEI and *EAI*	β	0.500	0.520	0.097	0.149	0.123	0.113	0.112
ALI and *CGEI*	γ	0.667	0.699	0.104	0.130	0.117	0.114	0.115

Source: Authors' estimation from UNDP (2006) data. (*) The adjusted parameter values are based on the replacement of α by $a = \alpha\mu_1/(\alpha\mu_1 + (1-\alpha)\mu_2)$, and similarly for β and γ, as mentioned in section 2 of the paper.

equation (17.5). Consider first the inequality in educational attainment, which is a function of the adult literacy and gross enrolment indices. Using the UNDP weight of 0.667, the actual between country Gini index of inequality for the educational attainment index is 0.126 when the variables are not weighted by population, and 0.115 when they are weighted. The lower and upper bounds provide fairly small intervals around these values, with the mid point of the interval being less than one percentage point away from the true value of the Gini index. The estimates under the exchangeability hypothesis perform even better, with an error below one tenth of a percentage point. The same can be observed for the *NGDPI* (non-per capita GDP) index computed using the life expectancy and educational attainment sub-indices. And again, the same is observed for the overall *HDI*, which is very well predicted under the exchangeability assumption.

Table 17.4 provides the intervals obtained for the various indices and sub-indices using different weighting schemes. All these intervals are fairly small overall due to high Gini correlations between the various indices entering into the *HDI*. For extreme weighting schemes such as α equal to zero or one, the analyst of course knows directly what the Gini of the weighted index of well-being will be, since in such cases the upper bound applies, and this is apparent in the estimates under the exchangeability assumption. But even for other weighting schemes such as those used by the UNDP, the predictions will be good. Also, it is worth noting that in most cases, changing the weights of the *HDI* would not affect between country inequality in human development very much.

4 Conclusion

Using the property that the Gini coefficient can be decomposed in a way which resembles the decomposition of the coefficient of variation, we have provided a method for testing the sensitivity of inequality measures based on multidimensional indicators to changes in the weights used to combine the various indicators. When applying the technique to the *HDI*, we found that the Gini correlations among the various components of the *HDI* are fairly high, and therefore, the impact of changing the weights of the various variables can be predicted with accuracy. More generally, while we have applied our method to the *HDI*, the method is quite general, and it can be applied to any multidimentional indicator when inequality is measured with the Gini index.

It is worth mentioning that while changing the weights in the *HDI* would not have a large impact on between country inequality in human development, this does not mean that there are no consequences from changing those weights. For example some governments tend to be sensitive to their country's precise ranking, as these rankings tend to be widely used. Even if in the aggregate inequality in the *HDI* is not affected much by changes in weights, particular country rankings may well be.

Table 17.4: Inequality bounds and mid points for alternative linear combinations of indices

	Real GDP and non-Real GDP indices				*Life expectancy and educational attainment indices*				*Adult literacy and combined gross enrolment indices*		
1-α	Upper bound	Lower bound	Exchan-geability	1-β	Upper bound	Lower bound	Exchan-geability	1-γ	Upper bound	Lower bound	Exchan-geability
Non-weighted											
0.0	0.144	0.166	0.166	0.0	0.132	0.167	0.167	0.0	0.104	0.122	0.122
0.1	0.141	0.163	0.161	0.1	0.129	0.163	0.160	0.1	0.106	0.124	0.122
0.2	0.139	0.160	0.157	0.2	0.126	0.159	0.154	0.2	0.108	0.127	0.123
0.3	0.136	0.157	0.153	0.3	0.123	0.155	0.148	0.3	0.110	0.129	0.125
0.4	0.133	0.154	0.149	0.4	0.120	0.151	0.143	0.4	0.112	0.132	0.127
0.5	0.131	0.152	0.146	0.5	0.117	0.147	0.139	0.5	0.114	0.134	0.129
0.6	0.128	0.149	0.144	0.6	0.114	0.143	0.136	0.6	0.116	0.137	0.132
0.7	0.126	0.146	0.142	0.7	0.111	0.140	0.133	0.7	0.119	0.140	0.135
0.8	0.124	0.144	0.140	0.8	0.108	0.136	0.131	0.8	0.121	0.143	0.139
0.9	0.121	0.141	0.139	0.9	0.106	0.133	0.130	0.9	0.123	0.146	0.144
1.0	0.119	0.139	0.139	1.0	0.103	0.130	0.130	1.0	0.126	0.149	0.149

Table 17.4: Inequality bounds and mid points for alternative linear combinations of indices (cont.)

1-α	*Real GDP and non-Real GDP indices* Upper bound	Lower bound	Exchan-geability	1-β	*Life expectancy and educational attainment indices* Upper bound	Lower bound	Exchan-geability	1-γ	*Adult literacy and combined gross enrolment indices* Upper bound	Lower bound	Exchan-geability
Weighted											
0.0	0.127	0.137	0.137	0.0	0.092	0.116	0.116	0.0	0.104	0.116	0.116
0.1	0.124	0.135	0.134	0.1	0.093	0.117	0.115	0.1	0.104	0.117	0.116
0.2	0.122	0.132	0.130	0.2	0.094	0.117	0.113	0.2	0.104	0.117	0.115
0.3	0.119	0.129	0.127	0.3	0.096	0.117	0.113	0.3	0.104	0.118	0.115
0.4	0.117	0.127	0.124	0.4	0.097	0.118	0.113	0.4	0.103	0.118	0.114
0.5	0.115	0.124	0.122	0.5	0.098	0.118	0.113	0.5	0.103	0.119	0.115
0.6	0.113	0.122	0.119	0.6	0.099	0.118	0.113	0.6	0.103	0.119	0.115
0.7	0.110	0.119	0.117	0.7	0.100	0.119	0.114	0.7	0.103	0.119	0.116
0.8	0.108	0.117	0.115	0.8	0.101	0.119	0.116	0.8	0.103	0.120	0.117
0.9	0.106	0.115	0.114	0.9	0.102	0.119	0.118	0.9	0.103	0.120	0.119
1.0	0.104	0.112	0.112	1.0	0.103	0.120	0.120	1.0	0.103	0.121	0.121

Source: Authors' estimation from UNDP (2006) data.

References

Anand, S. and A. Sen (1999) The Income Component in the HDI - Alternative Formulations, Occasional Paper, United Nations Development Programme, Human Development Report Office, New York.

Doessel, D. P. and R. Gounder (1994) Theory and Measurement of Living Levels: Some Empirical Results for the Human Development Index, *Journal of International Development*, **6**(4), 415–435.

Hicks, D. A. (1997) The Inequality-Adjusted Human Development Index: A Constructive Proposal, *World Development*, **25**(8), 1283–1298.

Lüchters, G. and L. Menkhoff (1996) Human Development as Statistical Artifact, *World Development*, **24**(8), 1385–1392.

McGillivray, M. (1991) The Human Development Index: Yet Another Redundant Composite Development Indicator?, *World Development*, **19**(10), 183–192.

McGillivray, M. and F. Noorbakhsh (2004) Composite Indices of Human Well-being: Past, Present and Future, Research Paper no. 2004/63, (November), United Nations University: WIDER.

McGillivray, M. and H. White (1993) Measuring Development? The UNDP's Human Development Index, *Journal of International Development*, **5**, 183–192.

McGranahan, D. V., C. Richard-Proust, N. V. Sovani and M. Subramanian (1972) Contents and Measurement of Socioeconomic Development, staff Study of the United Nations Research Institute for Social Development, New York: Praeger.

Noorbakhsh, F. (1998) The Human Development Index: Some Technical Issues and Alternative Indices, *Journal of International Development*, **10**, 589–605.

Noorbakhsh, F. (2007) International Convergence or Higher Inequality in Human Development? Evidence for 1975-2002, in G. Mavrotas and A. Shorrocks (eds.) *Advancing Development*, pp. 149–167, Palgrave Macmillan.

Ogwang, T. (1994) The Choice of Principle Variable for Computing the Human Development Index, *Worlk Development*, **22**(12), 2011–2014.

Palazzi, P. and A. Lauri (1998) The Human Development Index: Suggested Corrections, *Banca Nazionale del Lavoro Quarterly Review*, **51**(205), 193–221.

Ram, R. (1992) International Inequalities in Human Development and Real Income, *Economics Letters*, **38**, 351–354.

Schechtman, E. and S. Yitzhaki (1987) A Measure of Association Based on Gini's Mean Difference, *Communications in Statistics Theory and Methods*, **A16**(1), 207–231.

Schechtman, E. and S. Yitzhaki (1999) On the Proper Bounds of the Gini Correlation, *Economics Letters*, **63**, 133–138.

United Nations Development Programme (2006) *Human Development Report 2006 - Beyond Scarcity: Power, Poverty and the Global Water Crisis*, Oxford University Press, New York.

United Nations Research Institute for Social Development (UNRISD) (1970) Contents and Measuremetns of Socioeconomic Development, United Nationas Research Institute for Social Development, Geneva.

Wodon, Q. and S. Yitzhaki (2003) Inequality and the Accounting Period, *Economics Bulletin*, **4**(36), 1–8.
Yitzhaki, S. and Q. Wodon (2004) Mobility, Inequality, and Horizontal Equity, *Research on Economic Inequality*, **12**, 179–199.

Appendix

Proof of Equation (17.4):
The proof consists of finding upper and lower bound for G_α . The upper bound is

$$\begin{aligned} G_\alpha &= 2\text{cov}[\alpha Y_1 + (1-\alpha)Y_2, F(Y_\alpha)] \\ &= 2\alpha\text{cov}[Y_1, F(Y_\alpha)] + 2(1-\alpha)\text{cov}[Y_2, F(Y_\alpha)] \\ &\leq 2\alpha\text{cov}[Y_1, F(Y_1)] + 2(1-\alpha)\text{cov}[Y_2, F(Y_2)] = \alpha G_1 + (1-\alpha)G2. \end{aligned}$$

The derivation of the upper bound is based on Cauchy-Shwartz inequality, which can be utilized to show that for all Y_j and Y_k, $\text{cov}[Y_j, F(Y_k)] \leq \text{cov}[Y_j, F(Y_j)]$.

The lower bound obtains from the following:

$$\begin{aligned} G_\alpha &= 2\text{cov}[\alpha Y_1 + (1-\alpha)Y_2, F(Y_\alpha)] \\ &= 2\alpha\text{cov}[Y_1, F(Y_\alpha)] + 2(1-\alpha)\text{cov}[Y_2, F(Y_\alpha)] \\ &\geq \text{Max}[0, 2\alpha\text{cov}[(Y_1, F(Y_2)] + 2(1-\alpha)\text{cov}[Y_2, F(Y_1)] \\ &= \text{Max}[0, \alpha G_1 \Gamma_{12} + (1-\alpha)G_2\Gamma_{21}]. \end{aligned}$$

Proof of Equation (17.5):
Equation (17.5) states that when the variables are exchangeable then:

$$G_\alpha^2 = \alpha^2 G_1^2 + (1-\alpha)^2 G_2^2 + 2\alpha(1-\alpha)G_1 G_2 \Gamma$$

As before, using the properties of the covariance we can writee:

$$\begin{aligned} G_\alpha &= 2\text{cov}[\alpha Y_1 + (1-\alpha)Y_2, F(Y_\alpha)] \\ &= 2\alpha\text{cov}[Y_1, F(Y_\alpha)] + 2(1-\alpha)\text{cov}[Y_2, F(Y_\alpha)] \\ &= \alpha\Gamma_{1\alpha}G_1 + (1-\alpha)\Gamma_{2\alpha}G_2 \end{aligned}$$

Under exchangeability between $(Y_1, Y_\alpha), (Y_2, Y_\alpha)$, and (Y_1, Y_2), $\Gamma_{ij} = \Gamma_{ji}$ for $i, j = 1, 2, \alpha$. Substituting $\Gamma_{k\alpha}$ by $\Gamma_{\alpha k}$ $(k = 1, 2)$, we can write $\Gamma_{\alpha k}$ $(k = 1, 2)$ in terms of covariances, and move the denominator to the left hand side of the equation. Rearranging terms, and using $\Gamma_{12} = \Gamma_{21} = \Gamma$, we get the proof of equation (17.5). Equation (17.5) can be adjusted to hold for distributions with different expected value. In this case:

$$G_{\alpha}^{2}=s^{2}G_{1}^{2}+(1-s)^{2}G_{2}^{2}+2s(1-s)G_{1}G_{2}\Gamma,$$

where $s=\alpha\mu_1/(\alpha\mu_1+(1-\alpha)\mu_2)$, where $\mu_i>0$, is the expected value of the appropriate variable.

Index