The Lorenz curve in economics and econometrics *

Christian Kleiber †

11th July 2005

Abstract

This paper surveys selected applications of the Lorenz curve and related stochastic orders in economics and econometrics, with a bias towards problems in statistical distribution theory. These include characterizations of income distributions in terms of families of inequality measures, Lorenz ordering of multiparameter distributions in terms of their parameters, probability inequalities for distributions of quadratic forms, and Condorcet jury theorems.

Keywords: Lorenz curve, Lorenz order, majorization, income distribution, income inequality, statistical distributions, characterizations, Condorcet jury theorem.
AMS classification: 60E15, 62P20, 62E15.

1 Introduction

100 years ago, in June 1905, a short article entitled

> Methods of Measuring the Concentration of Wealth

appeared in the *Publications of the American Statistical Association* (the forerunner of the *Journal of the American Statistical Association*), proposing a simple method, subsequently called the Lorenz curve, for visualizing distributions of income or wealth with respect to their inherent "inequality" or "concentration." Its author, Max Otto Lorenz, was about to complete his Ph.D. dissertation at the University of Wisconsin. This article apparently remained his only publication in a scientific journal, and it made him famous. A short biography of M.O. Lorenz is available in Kleiber and Kotz (2003, pp. 263–265).
According to Derobert and Thieriot (2003), the term "Lorenz curve" occurs for the first time in King (1912), a statistics textbook written for economists and social scientists. However, it was not until the early 1970s that interest in the Lorenz curve increased

*Invited paper, Gini-Lorenz Centennial Conference, Siena, May 23–26, 2005.

†Work partially supported by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich 475. Correspondence to: Christian Kleiber, Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund, D-44221 Dortmund, Germany. E-mail: kleiber@statistik.uni-dortmund.de

substantially, at least in the English-language statistical and economic literature, triggered by the seminal papers of Atkinson (1970) and Gastwirth (1971). They presented the welfare-economic implications of Lorenz-curve comparisons (Atkinson) and a simple definition of the Lorenz curve for fairly general distributions (Gastwirth). It did not hurt that both were published in highly regarded journals. Among the first contributions of this new wave were Sen's (1973) Radcliffe lectures at the University of Warwick, Fellman's (1976) analysis of transformations, and Jakobsson's (1976) and Kakwani's (1977) studies of progressive taxation.

In the statistical literature, an important paper is due to Goldie (1977) who studied the asymptotics of the Lorenz curve in what nowadays would be called an empirical-process framework. At about the same time, the Lorenz ordering found a multitude of applications in theoretical statistics, often in the form of the more restrictive majorization ordering, among them inequalities for power functions in multivariate analysis. See Marshall and Olkin (1979) and Tong (1988, 1994). The monographs of Arnold (1987) and Csörgő, Csörgő and Horvath (1986) further popularized the concept among statisticians, leading to numerous applications, notably in reliability theory.

The *Current Index of Statistics*, for the year 2004, provides some 140 papers with the keywords "Lorenz curve" or "Lorenz order," while the 2004 version of *EconLit*, the American Economic Association's electronic database, provides some 200 hits just for the years 1969–present. Presumably more than 500 methodological papers have been written in the last 50 years in statistical and econometric journals, not to mention numerous publications of an applied nature that do not list "Lorenz curve" as a keyword.

In view of this large number of publications it appears impossible to provide a comprehensive view in a short article such as the present one. Instead, this paper tries to survey selected applications of the Lorenz curve and of the closely connected Lorenz and majorization orderings in economics and econometrics. My survey is somewhat biased towards statistical distribution theory. In particular, the two classical topics related to the Lorenz curve are not covered at all: economic disparity measures and taxation problems. For surveys of these I refer to Mosler (1994) and to Arnold (1990) and Lambert (2001), respectively.

2 Lorenz curves and the Lorenz order

To draw the Lorenz curve of an n-point empirical distribution $\mathbf{x} = (x_1, \ldots, x_n)$, $x_i \geq 0$, $\sum_{i=1}^n x_i > 0$, say of household income, one plots the share $L(k/n)$ of total income received by the $k/n \cdot 100\%$ of the poorest households, $k = 0, 1, 2, \ldots, n$, and interpolates linearly.

In the discrete (or empirical) case the Lorenz curve is therefore defined in terms of the $n + 1$ points

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

where $x_{i:n}$ denotes the ith smallest income, and a continuous curve $L(u), u \in [0, 1]$, is given

by

$$L(u) = \frac{1}{n\bar{x}} \left\{ \sum_{i=1}^{\lfloor un \rfloor} x_{i:n} + (un - \lfloor un \rfloor) x_{\lfloor un \rfloor + 1:n} \right\}, \quad 0 \leq u \leq 1,$$

where $\lfloor un \rfloor$ denotes the largest integer not exceeding un.
The appropriate definition of the Lorenz curve for a general distribution follows easily by recognizing the expression (1) as a sequence of standardized empirical incomplete first moments. In view of $E(X) = \int_0^1 F_X^{-1}(t)\, dt$, where the quantile function F_X^{-1} is defined as the pseudoinverse of the cumulative distribution function (CDF), F_X,

$$F_X^{-1}(t) \;=\; \sup\{x \,|\, F_X(x) \leq t\}\;, \quad t \in [0,1], \tag{2}$$

equation (1) may be rewritten in the form

$$L_X(u) = \frac{1}{E(X)} \int_0^u F_X^{-1}(t)\; dt\;, \quad u \in [0,1]. \tag{3}$$

Hence any distribution supported on the non-negative halfline with a finite and positive first moment admits a Lorenz curve. Following Arnold (1987), I shall occasionally denote the set of all random variables with distributions satisfying these conditions by $\mathcal{L}$.
It is a direct consequence of (3) that the Lorenz curve has the following properties:

- L is continuous on $[0,1]$, with $L(0) = 0$ and $L(1) = 1$,
- L is increasing, and
- L is convex.

Conversely, any function possessing these properties is the Lorenz curve of a certain statistical distribution (Thompson, 1976). It is also worth noting that the Lorenz curve itself may be considered a CDF on the unit interval. By construction, the quantile function associated with this "Lorenz-curve distribution" is also a CDF. It is sometimes referred to as the Goldie curve, after Goldie (1977) who studied its asymptotic properties.
Among Italian statisticians, the representation (3) in terms of the quantile function was used as early as 1915 by Pietra who was not aware of Lorenz's contribution. It has later been popularized by Piesch (1967, 1971) in the German-language literature. However, it was not until Gastwirth's 1971 *Econometrica* article that interest increased substantially in the English-language statistical literature.
Incidentally, the definition given above is not Lorenz's original definition. Obviously, there are four variants of the basic idea (see Figure 1): Lorenz used the graph $(L(u), u)$, an increasing but concave function. Chatelain (1907), in what would seem to be an independent discovery of the Lorenz curve employed, up to some scaling, $(u, L(1-u))$. King (1912) also used Chatelain's version, while Chatelain (1910) himself soon switched to $(u, 1-L(1-u))$, without mentioning Lorenz's pioneering work. On combinatorial grounds, it would be of some interest to determine a source proposing the form $(u, 1-L(u))$, but I have been

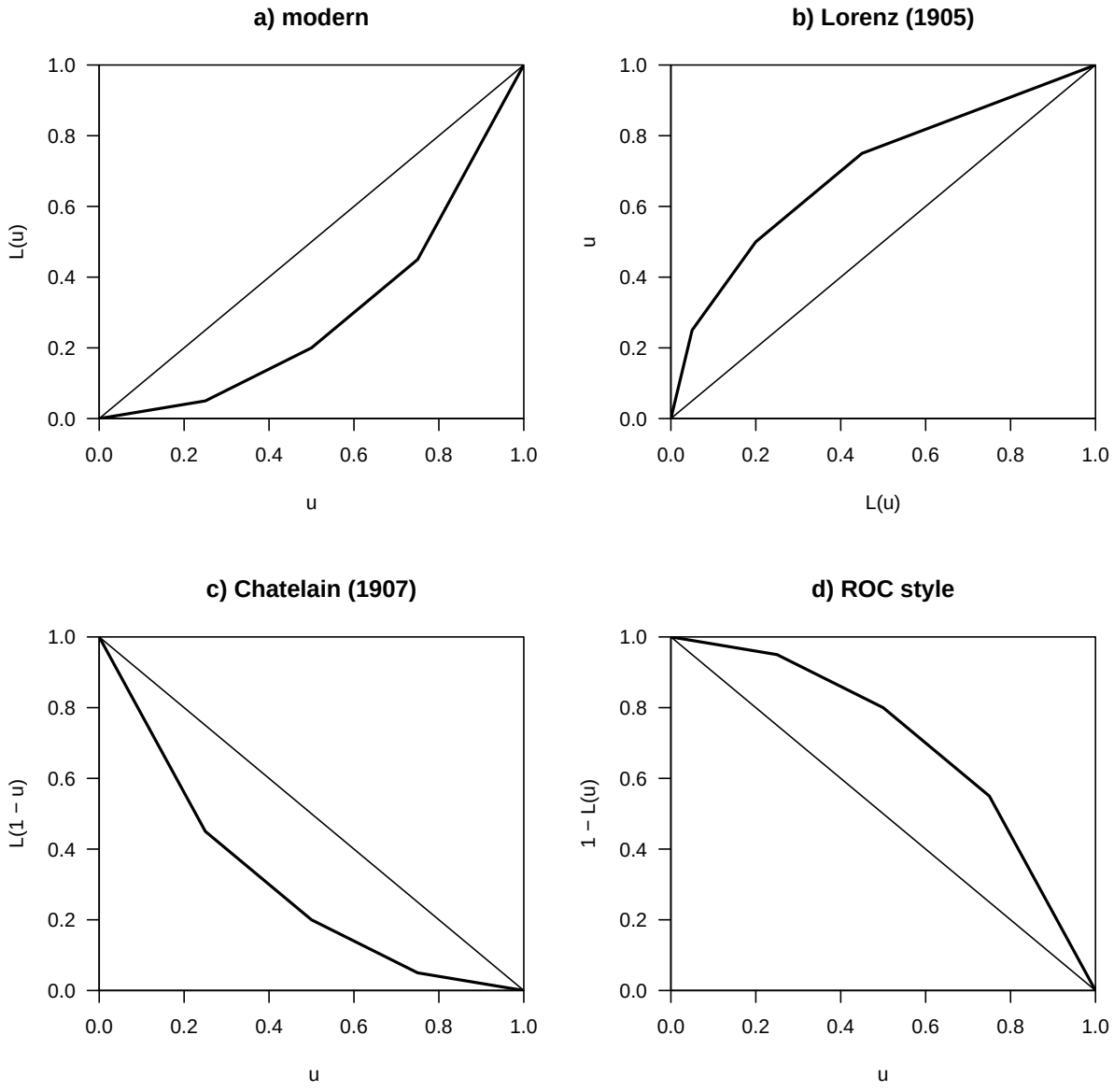

Figure 1: Lorenz curves, historical and modern, for $\mathbf{x} = (1, 3, 5, 11)$.

unable to locate such work. However, this variant coincides, under certain conditions, with the common version of a tool known as the receiver-operating-characteristic (ROC) curve in biostatistics. Further information on the history of the Lorenz curve may be found in Derobert and Thieriot (2003).

The definition of the Lorenz curve suggests to compare entire distributions by comparing the corresponding Lorenz curves. By construction, the diagonal of the unit square corresponds to the Lorenz curve of a society in which everybody receives the same income

and hence serves as a benchmark case against which actual income distributions may be measured. Indeed, virtually every software package that provides this graphical display by default also plots the diagonal of the unit square.
There are several variants of this idea: For two vectors $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ of identical length n satisfying $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, one may define

$$\mathbf{x} \geq_M \mathbf{y} :\Longleftrightarrow \sum_{i=n-j}^{n} x_{i:n} \geq \sum_{i=n-j}^{n} y_{i:n}, \quad j = 0, 1, \ldots, n.$$

This is the majorization ordering introduced by Hardy, Littlewood and Pólya (1929). It has found a multitude of applications in statistics and applied mathematics, see the famous text by Marshall and Olkin (1979) for a comprehensive survey. If $\sum_{i=1}^n x_i \neq \sum_{i=1}^n y_i$ there are several options. The Lorenz curve proceeds via rescaling of the data, i.e. the transformation $(x_1, \ldots, x_n) \mapsto (x_1/\sum_{i=1}^n x_i, \ldots, x_n/\sum_{i=1}^n x_i)$, thereby extending majorization in two directions, permitting (i) scale-free comparisons – in economic terms, currencies or inflation play no role – and (ii) comparisons of populations of different sizes (the "population principle" of economic inequality measurement). Finally, a general definition based on (3) permits comparisons of fairly arbitrary distributions, provided the corresponding random variables are non-negative with positive expectations:

Definition 1 *For $X_1, X_2 \in \mathcal{L}$, the random variable X_1 is said to be at least as unequal (or variable) as X_2 in the Lorenz sense if $L_1(u) \leq L_2(u)$ for all $u \in [0, 1]$. That is,*

$$X_1 \geq_L X_2 \quad :\Longleftrightarrow \quad L_1 \leq L_2. \tag{4}$$

It is convenient to use the notations $X_1 \geq_L X_2$ and $F_1 \geq_L F_2$ simultaneously. Economists usually prefer to denote the situation where $L_1 \leq L_2$ as $X_2 \geq_L X_1$, because the distribution F_2 is, in a certain sense, associated with a higher level of economic welfare (Atkinson, 1970). Here I shall use the form (4) which appears to be the common one in the statistical literature.
It is clear from (4) that the Lorenz order is a partial order, it is scale free in the sense that

$$X_1 \geq_L X_2 \Longleftrightarrow a \cdot X_1 \geq_L b \cdot X_2, \quad \text{for all } a, b > 0.$$

3 Characterizations

Since any distribution is characterized by its quantile function it follows from (3) that the Lorenz curve characterizes a distribution in $\mathcal{L}$ up to a scale parameter (e.g., Iritani and Kuga, 1983). As mentioned above, the Lorenz curve itself may be considered a CDF on the unit interval. This implies, inter alia, that this "Lorenz-curve distribution"—having bounded support—can be characterized by the sequence of its moments. Furthermore, these "Lorenz-curve moments" characterize the underlying distribution up to a scale parameter. This characterization is due to Aaberge (2000). Below I present a slightly different account,

following Kleiber and Kotz (2002), which relates the problem to the moment problem of order statistics.
What are the moments of the "Lorenz-curve distribution"? Denote by X_L a random variable supported on $[0, 1]$ with CDF L. Then

$$E(X_L^k) = k \int_0^1 u^{k-1}\{1 - L(u)\}du.$$

It is not difficult to see that

$$E(X_L) = \frac{G}{2} + \frac{1}{2},$$

where

$$G = 2\int_0^1 (u - L(u))du = 1 - 2\int_0^1 L(u)du \tag{5}$$

is the Gini coefficient, perhaps the most widely used measure of income inequality (Gini, 1914). The Gini index is a relative measure of income inequality since it depends only on income shares. A sizable number of alternative representations are available. For the characterizations of interest here, the expression

$$G = 1 - \frac{E(X_{1:2})}{E(X)} = 1 - \frac{\int_0^\infty \{1 - F(x)\}^2\, dx}{E(X)}, \tag{6}$$

presumably due to Arnold and Laguna (1977), is the most appropriate.
Kakwani (1980) proposed a one-parameter family of generalized Gini indices by introducing different weighting functions for the area under the Lorenz curve,

$$G_n = 1 - n(n-1)\int_0^1 L(u)(1-u)^{n-2}\, du,$$

here n is a non-negative integer. The traditional Gini coefficient is obtained for $n = 2$. Donaldson and Weymark (1980, 1983) and Yitzhaki (1983) have arrived at the same family from different considerations. These authors also defined a family of 'equally-distributed-equivalent-income functions' of the form

$$\Xi_n = -\int_0^\infty x\, d\{(1 - F(x))^n\},$$

which may be rewritten as $\Xi_n = \int_0^\infty \{1 - F(x)\}^n\, dx$. Muliere and Scarsini (1989) observed that Ξ_n equals $E(X_{1:n})$ and that

$$G_n = 1 - \frac{\Xi_n}{E(X)} = 1 - \frac{E(X_{1:n})}{E(X)}. \tag{7}$$

Equation (7) is a direct generalization of (6).

This shows that the Lorenz-curve moments are closely related to moments of order statistics. Furthermore, it suggests to reduce the characterization in terms of Lorenz-curve moments to the well-known moment problem of order statistics. Specifically, let $X_1, \ldots, X_n$ be a sample of size n from a distribution with the CDF F and define the order statistics $X_{i:n}$ in the ascending order by

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}.$$

The moment problem of order statistics inquires to what extent the CDF F is uniquely determined by (a subset of) the first moments of all of its order statistics

$$\{E(X_{i:n}) \mid i = 1, 2, \ldots, n; n = 1, 2, 3, \ldots\}. \tag{8}$$

It follows from

$$E(X_{i:n}) = i\binom{n}{i} \int_0^1 F^{-1}(u)\, u^{i-1}(1-u)^{n-i}\, du$$

that $E|X_{i:n}| \leq c \cdot E|X|$, for some $c > 0$; thus a finite mean of the parent distribution assures the existence of the first moment of any order statistic. This implies that characterizations in terms of the moments of order statistics are of interest for heavy-tailed distributions of the Pareto type, for which only a few moments exist and, consequently, no characterization in terms of (ordinary) moments is feasible. Many parametric models for the size distribution of personal income are of this type, see Kleiber and Kotz (2003) for a recent survey.
In view of the familiar recurrence relation (David, 1981, p. 46)

$$(n-i)\, E(X_{i:n}) + i\, E(X_{i+1:n}) = n\, E(X_{i:n-1})$$

it is not necessary to know the whole array (8), one merely requires one moment for each sample size, e.g. the sequence of expectations of minima $E(X_{1:n})$ will suffice. The basic characterization result is thus as follows:

Lemma 2 *Let $E|X| < \infty$. For $n = 1, 2, 3, \ldots$, let $i(n)$ be an integer with $1 \leq i(n) \leq n$. Then, F is uniquely determined by the sequence $\{E(X_{i(n):n}) \,|\, n = 1, 2, 3, \ldots\}$.*

The most natural choices for $i(n)$ are either 1 or n. Many refinements of this fundamental result are available in the literature, see e.g. Kamps (1998) for further details.
Lemma 2 yields the following characterization via the moment problem of order statistics:

Theorem 3 *Any $F \in \mathcal{L}$ is characterized, up to a scale, by its sequence of generalized Gini indices, $\{G_n\}$.*

Various extensions of this theorem are discussed by Kleiber and Kotz (2002).
As an example, consider the exponential distribution with scale parameter λ. Its CDF is $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$, $\lambda > 0$, and therefore $E(X_{1:n}) = \lambda/n$, hence the sequence $\{G_n \,|\, G_n = 1 - 1/n\}$ characterizes the family of exponential distributions up to a scale.

4 The Lorenz order within parametric families of income distributions

Parametric models for the size distribution of personal income have been of interest to econometricians and applied statisticians for more than a hundred years. An income distribution has the property that its CDF F is supported on the positive halfline, i.e. $\text{supp}(F) \subseteq [0, \infty)$. Atkinson's (1970) classic paper has created much interest in stochastic orders for the comparison of income distributions such as the Lorenz order. It is therefore quite surprising that only fairly recently attempts have been made to characterize the Lorenz order within common parametric families of income distributions.

For one- and two-parameter models this is straightforward, and it is well-known that the Lorenz order is linear within the Pareto and log-normal families. Indeed, for the Pareto distribution, with CDF $F(x) = 1 - (x/x_0)^{-\alpha}$, $x \geq x_0 > 0$, quantile function $F^{-1}(u) = x_0(1-u)^{-1/\alpha}$, $0 < u < 1$, and mean $E(X) = \alpha x_0/(\alpha - 1)$ (which exists if and only if $\alpha > 1$), it follows that

$$L(u) = 1 - (1-u)^{1-1/\alpha}, \quad 0 < u < 1. \tag{9}$$

Hence Lorenz curves from Pareto distributions with a different α never intersect. Specifically,

$$X_1 \geq_L X_2 \Longleftrightarrow \alpha_1 \leq \alpha_2.$$

For the lognormal distribution, with CDF $\Phi((\log x - \mu)/\sigma)$, where $x > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$, and Φ is the CDF of the standard normal distribution, the Lorenz curve is given by

$$L(u) = \Phi(\Phi^{-1}(u) - \sigma^2), \quad 0 < u < 1.$$

It follows that $X_1 \geq_L X_2$ if and only if $\sigma_1^2 \geq \sigma_2^2$.

Within three- and four-parameter families the Lorenz order is no longer linear, however. The first results for a three-parameter family are due to Taillie (1981), who studied the generalized gamma distribution. More than a decade later, Wilfling and Krämer (1993) obtained results for the popular Singh-Maddala (1976) family, and Kleiber (1996) considered the even closer fitting Dagum (1977) distributions. These distributions are special cases of a four-parameter distribution, the generalized beta distribution of the second kind (hereafter: GB2) introduced by McDonald (1984) and Venter (1983) in econometrics and actuarial sciences, respectively. In empirical applications, the GB2 distribution has been found to outperform its competitors, sometimes by wide margins, see Bordley, McDonald and Mantrala (1996) for a comparative study including some 15 distributions.

The GB2 distribution has the density

$$f_{GB2}(x) \quad = \quad \frac{a\, x^{ap-1}}{b^{ap}\, B(p,q)\, [1 + (x/b)^a]^{p+q}}\,, \quad x > 0, \tag{10}$$

where $B(\cdot,\cdot)$ is the beta function and all four parameters a, b, p, q are positive. The parameter b is a scale parameter, the others are shape parameters. Note that a GB2 distribution has finite mean — and, therefore, admits a Lorenz curve — if and only if $aq > 1$. For further details and more than twenty other distributions related to the GB2, see Kleiber and Kotz (2003).

As of early 2005, a complete characterization of the Lorenz ordering within the GB2 family of distributions is still unavailable. The most general result is as follows (Kleiber, 1999):

Theorem 4 *Let X_1, X_2 be in $\mathcal{L}$, with $X_i \sim GB2(a_i, b_i, p_i, q_i)$, i=1,2. Then*

(a) $a_1 \le a_2$, $a_1 p_1 \le a_2 p_2$, and $a_1 q_1 \le a_2 q_2$ imply $X_1 \ge_L X_2$.

(b) $X_1 \ge_L X_2$ implies $a_1 p_1 \le a_2 p_2$ and $a_1 q_1 \le a_2 q_2$.

This leaves open constellations of the type $a_1 \le a_2$, $p_1 \ge p_2$, and $q_1 \ge q_2$, but $a_1 p_1 \ge a_2 p_2$ and $a_1 q_1 \ge a_2 q_2$. However, Theorem 4 encompasses complete characterizations for all subfamilies of the GB2, thereby providing a unified approach to most commonly considered income distribution functions. Specifically, for the Singh-Maddala distribution, with $SM(a, b, q) \equiv GB2(a, b, 1, q)$, Theorem 4 yields

$$X_1 \ge_L X_2 \Longleftrightarrow a_1 \le a_2 \text{ and } a_1 q_1 \le a_2 q_2,$$

for the Dagum distribution, with $D(a, b, p) \equiv GB2(a, b, p, 1)$, it implies

$$X_1 \ge_L X_2 \Longleftrightarrow a_1 \le a_2 \text{ and } a_1 p_1 \le a_2 p_2,$$

for the beta distribution of the second kind, with $B2(b, p, q) \equiv GB2(1, b, p, q)$, we have

$$X_1 \ge_L X_2 \Longleftrightarrow p_1 \le p_2 \text{ and } q_1 \le q_2,$$

whereas for the log-logistic distribution, with $LL(a, b) \equiv GB2(a, b, 1, 1)$ the condition is

$$X_1 \ge_L X_2 \Longleftrightarrow a_1 \le a_2.$$

The proof of part (a) of Theorem 4 utilizes a representation of the GB2 distribution as the distribution of the ratio of two independent generalized gamma variates and Taillie's (1981) Lorenz ordering results for that distribution. Necessity is proved via properties of regularly varying functions, see Kleiber (2000, 2002) for further details. Indeed, a comparison of (10) and Theorem 4 (b) reveals that aq determines the rate of decrease of the density in the upper tail, while ap does likewise for the lower tail. Figure 2 provides an illustration for two Dagum distributions: the more unequal distribution is associated with heavier tails, as is to be expected from Theorem 4 (b).

Incidentally, Theorem 4 also has applications in a reliability context: the GB2 distribution is the distribution of the order statistics from a log-logistic parent distribution, hence Theorem 4 provides conditions for the Lorenz ordering of order statistics from that distribution, see Kleiber (2004) for further details.

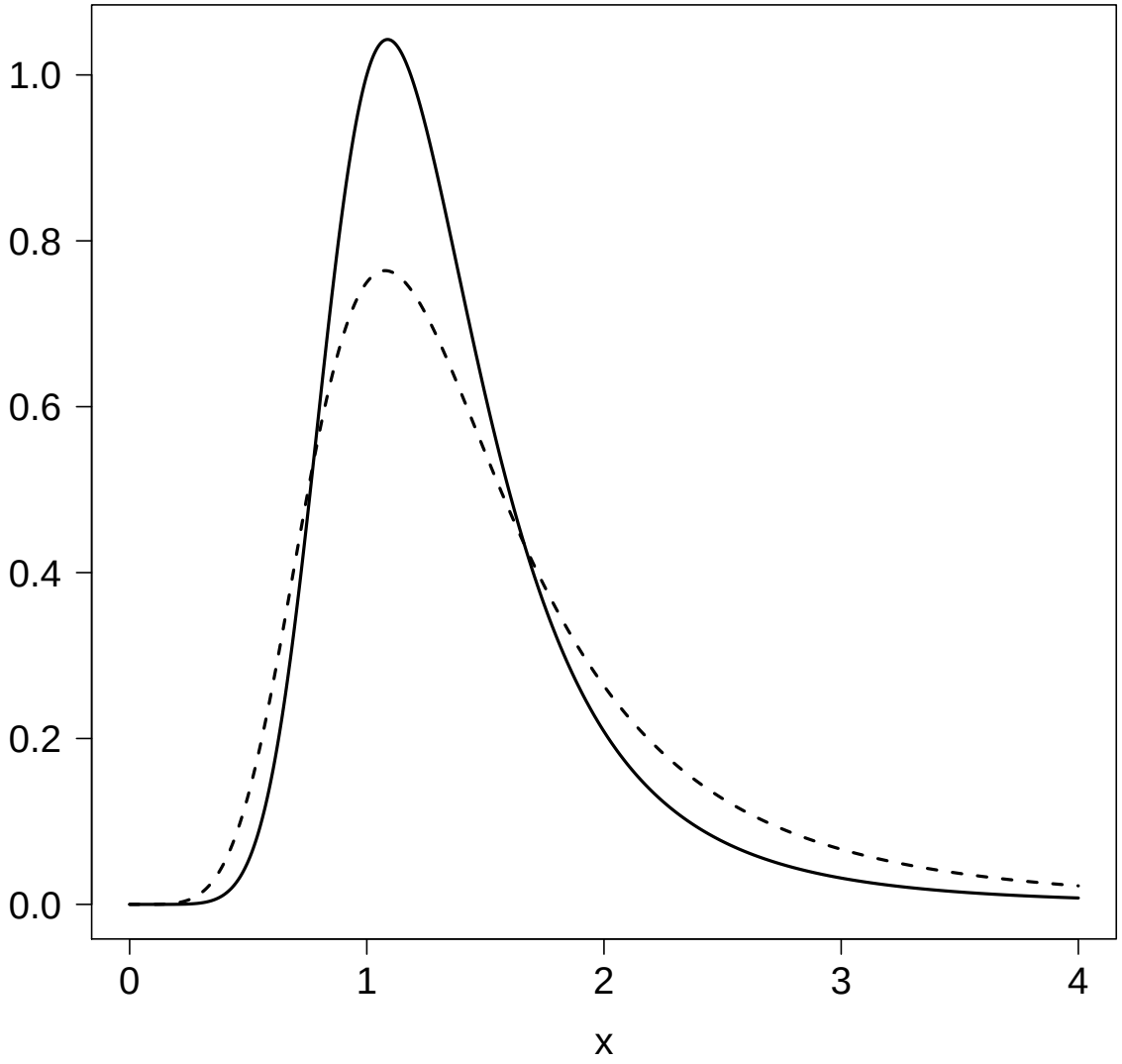

Figure 2: Two Dagum distributions: $X_1 \sim D(3,1,2)$ (dashed), $X_2 \sim D(4,1,2)$ (solid), hence $X_1 \geq_L X_2$.

5 Some probability inequalities in econometrics

A problem in statistical distribution theory that is of considerable interest in econometrics is the distribution of quadratic forms in normal random variables. Suppose $x \sim \mathcal{N}(0, I_n)$, $A \in \mathbb{R}^{n \times n}$ and consider $Q(x) := x^\top Ax$. What is the distribution of Q? From the theory of linear models it is well known that

$$A^\top = A \text{ and } A^2 = A \quad \Longleftrightarrow \quad Q(x) \sim \chi^2(rk(A)).$$

What if A is not an idempotent matrix? This occurs, inter alia, in connection with the Durbin-Watson test, where A is the difference of two positive semidefinite matrices.
The problem may be rewritten in the form

$$Q(x) = \sum_{j=1}^{n} \lambda_j x_j^2$$

where the $x_j \sim \mathcal{N}(0,1)$ are i.i.d.and the λ_j are the eigenvalues of A. If these eigenvalues are distinct, the distribution of Q is not available in closed form and must be determined numerically. However, it is possible to obtain qualitative results using the Lorenz curve, or rather the majorization ordering.
Suppose now that A is p.s.d., hence $\lambda_j \geq 0$, $j = 1, \ldots, n$. A classical inequality due to Okamoto (1960) states that

Theorem 5 *Suppose $X_j \sim \mathcal{N}(0,1)$ are i.i.d. and $\lambda_j \geq 0$. Then*

$$P\left(\sum_{j=1}^{n} \lambda_j X_j^2 \leq x\right) \leq P\left(Y \leq x/\tilde{\lambda}\right), \tag{11}$$

where $Y \sim \chi_n^2$ and $\tilde{\lambda} := (\prod_{j=1}^n \lambda_j)^{1/n}$ is the geometric mean of the λ_j.

Note that the RHS of (11) is a chi-square probability and therefore easily computed.
In view of $\tilde{\lambda} = \exp(\frac{1}{n}\sum_{j=1}^n \log \lambda_j)$ and the basic majorization inequality $(\log \lambda_1, \ldots, \log \lambda_n) \geq_M (\log \tilde{\lambda}, \ldots, \log \tilde{\lambda})$, the inequality (11) suggests that a generalization of Okamoto's inequality in terms of majorization might be available. The following result is due to Marshall and Olkin (1979, p. 303):

Theorem 6 *Suppose $X_j \sim \mathcal{N}(0,1)$, i.i.d., and $a_j, b_j > 0$. Suppose further $(\log a_1, \ldots, \log a_n) \geq_M (\log b_1, \ldots, \log b_n)$. Then*

$$P\left(\sum_{j=1}^{n} a_j X_j^2 \leq x\right) \leq P\left(\sum_{j=1}^{n} b_j X_j^2 \leq x\right).$$

This Theorem says that the probability $P\left(\sum_{j=1}^n a_j X_j^2 \leq x\right)$ is decreasing in $(\log a_1, \ldots, \log a_n)$, in the sense of majorization, hence the more variable the vector of logarithms of the eigenvalues of A is, the more likely is Q to take on extreme values. Further majorization inequalities and bounds in terms of the harmonic mean of the λ_j are discussed by Tong (1988).

6 Condorcet jury theorems

A further problem in statistical distribution theory involving the Lorenz curve, or rather the Lorenz order, is concerned with bounds for the CDF of the sum of heterogeneous Bernoulli variables. This has an interesting application in the theory of social choice.
Consider a panel of jurors facing a binary choice. One of the alternatives is assumed to be correct. Being experts, the jurors are able to do better than a fair coin, that is, they are able to identify the correct alternative with a probability exceeding 1/2.
In his *Essai sur l'application de l' analyse à la probabilité des décisions rendues à la pluralité des voix*, Condorcet (1785) expressed the belief that these jurors, utilizing a simple majority rule, would be likely to make the correct decision. Mathematical formulations substantiating this belief have become known as Condorcet jury theorems (CJTs) in social choice and (theoretical) political science, see Grofman and Owen (1989) and Boland (1989) for surveys and further references.
Suppose the random variable X_i indicates whether the ith expert makes the correct decision, where $p_i = P(X_i = 1)$, for $i = 1, \ldots, n$. Define $S := \sum_{i=1}^{n} X_i$, the random variable indicating the number of correct decisions. Clearly, if $p_i \equiv p$ for all i and the experts decide independently,

$$h_n(p) := P(S \geq k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

In order to avoid ties, it is convenient to suppose that the jury size is odd, $n = 2m + 1$. Hence the majority rule corresponds to $k = m + 1$, and the quantity of interest is

$$h_{2m+1}(p) := P(S \geq m+1) = \sum_{i=m+1}^{2m+1} \binom{2m+1}{i} p^i (1-p)^{2m+1-i}.$$

The classical form of the CJT is as follows:

- $h_n(p) > p$, i.e., with a majority voting system, we are more likely to arrive at the correct decision with a panel of experts of equal competence p than with a single individual of competence p.

This is the simplest and most popular form of the CJT. At least two generalizations would seem to be of interest: the first substitutes homogeneity with varying competence, the second allows for correlation. I shall confine myself to the first.
The experts now have different abilities to identify the correct alternative, that is, p_i does not necessarily equal p_j, $i \neq j$. The p_is are collected in a vector, $\mathbf{p} := (p_1, \ldots, p_n)$. A convenient reference point is provided by the average expert competence, $\bar{p}$.
Perhaps surprisingly, this setting invokes the following classical inequality due to Hoeffding (1956):

Lemma 7 *Let $k > 0$ be an integer and suppose $\bar{p} \geq k/n$. Then*

$$P(S \geq k) \geq \sum_{i=k}^{n} \binom{n}{i} \bar{p}^i (1-\bar{p})^{n-i}.$$

With $k = m + 1$ this yields Boland's (1989) generalization of the CJT:

Theorem 8 *Suppose $n \geq 3$, $\bar{p} \geq 1/2 + 1/(2n)$. Then*

$$h_n(\mathbf{p}) := h_n(p_1, \ldots, p_n) > \bar{p}.$$

A panel of experts with average competence $\bar{p}$ will therefore do better than a single expert with competence $\bar{p}$.
How is all this related to the Lorenz order? The preceding theorem suggests that, in view of the basic majorization inequality $\mathbf{p} = (p_1, \ldots, p_n) \geq_M (\bar{p}, \ldots, \bar{p})$, it might be true that $\mathbf{p}_1 \geq_M \mathbf{p}_2$ implies $h_n(\mathbf{p}_1) \geq h_n(\mathbf{p}_2)$. Indeed there is some such result, but the details are more involved. A more general version depends on a refinement of Hoeffding's inequality due to Gleser (1975):

Lemma 9 *Suppose $\mathbf{p}_1 \geq_M \mathbf{p}_2$. Then*

$$P(S \leq k|\mathbf{p}_1) \leq P(S \leq k|\mathbf{p}_2), \quad k \leq \lfloor n\bar{p} - 2 \rfloor,$$

Note that, for $\mathbf{p}_2 = (\bar{p}, \ldots, \bar{p})$, the Lemma yields Hoeffding's result. As discussed by Gleser, the more stringent condition $k \leq \lfloor n\bar{p} - 2 \rfloor$ cannot be removed in the general case.
Lemma 9 yields

Theorem 10 *Suppose $n \geq 7$ and $\bar{p} \geq 1/2 + 5/(2n)$. Then $\mathbf{p}_1 \geq_M \mathbf{p}_2$ implies*

$$h_n(\mathbf{p}_1) \geq h_n(\mathbf{p}_2).$$

The majorization approach yields further insights for CJTs, among them results for super-majority voting rules, see Kleiber (2005) for further discussion.

7 Generalized Lorenz curves

The Lorenz curve is only a partial order, so what does one do if two Lorenz curves intersect? The most widely used alternative to the Lorenz order is the generalized Lorenz order, due to Shorrocks (1983) and Kakwani (1984). It is defined in terms of the generalized Lorenz curve, GL_X, where

$$GL_X(u) = E(X) \cdot L_X(u) = \int_0^u F_X^{-1}(t)dt, \quad 0 \leq u \leq 1, \tag{12}$$

and suggests to prefer a distribution F over another distribution G if its generalized Lorenz curve is nowhere below the generalized Lorenz curve of G. This is denoted as $F \geq_{GL} G$. (Note that this definition represents a reversal of the inequality defining the Lorenz order in section 2. For the purposes of this section, this is more convenient than the standard version.) Generalized Lorenz curves are nondecreasing, continuous and convex, with $GL_F(0) = 0$ and $GL_F(1) = \mu_F < \infty$. Thistle (1989a) shows that a distribution is uniquely determined by its generalized Lorenz curve. Also, from e.g. Thistle (1989b), generalized Lorenz dominance is equivalent to second-order stochastic dominance, denoted here as $SD(2)$, where $F \geq_{SD(2)} G$ if and only if $\int_0^x F(t)\,dt \leq \int_0^x G(t)\,dt$ for all $x \in \mathbb{R}_+$.
Being defined in terms of the Lorenz curve, the generalized Lorenz order encompasses inequality (equity) aspects, being scaled by $E(X)$, it also encompasses size (efficiency) aspects. It is therefore of interest to pursue decompositions of the generalized Lorenz order into these equity and efficiency aspects. The Lorenz curve provides a natural tool for measuring equity. One way to study efficiency, in a global sense, is to consider the classical stochastic order (more familiar as first-order stochastic dominance, or SD(1), in economics), where $F \geq_{SD(1)} G$ if $F(x) \leq G(x)$ for all x. In the terminology of welfare economics, $F \geq_{SD(1)} G$ means that the distribution F is ranked higher than G by all social welfare functions with increasing utility, while $F \geq_{SD(2)} G$ means that F is ranked higher than G by all social welfare functions with increasing and concave utility.
Kleiber and Krämer (2003) provide the following result:

Theorem 11 *Suppose F, G are income distributions supported on the positive halfline with finite expectations. Then the following are equivalent:*

(a) $F \geq_{GL} G$.

(b) There exists an income distribution H_1, with $\mu_{H_1} = \mu_G$, such that $F \geq_{SD(1)} H_1 \geq_L G$.

(c) There exists an income distribution H_2, with $\mu_{H_2} = \mu_F$, such that $F \geq_L H_2 \geq_{SD(1)} G$.

As an illustration, consider the gamma distribution, with density

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\lambda x}, \quad x > 0,$$

where $\lambda > 0, \alpha > 0$. Suppose $F \sim \text{Ga}(\alpha, \lambda)$ and $G \sim \text{Ga}(\beta, \nu)$, respectively. Taillie (1981) shows that $F \geq_L G$ if and only if $\alpha \geq \beta$. From e.g. Ramos, Ollero and Sordo (2000, pp. 290-291) we moreover have that $\lambda > \nu$ and $\alpha/\lambda \geq \beta/\nu$ imply $F \geq_{GL} G$, whereas $\lambda \leq \nu$ and $\alpha \geq \beta$ imply $F \geq_{SD(1)} G$. Now suppose $F \sim$ Ga(20,5) and $G \sim$ Ga(10,4), hence $F \geq_{GL} G$. Then H_1 may be chosen as Ga(15,6), whereas H_2 could be Ga(12,3), for example. The generalized Lorenz curves of all these distributions are depicted in Figure 3.

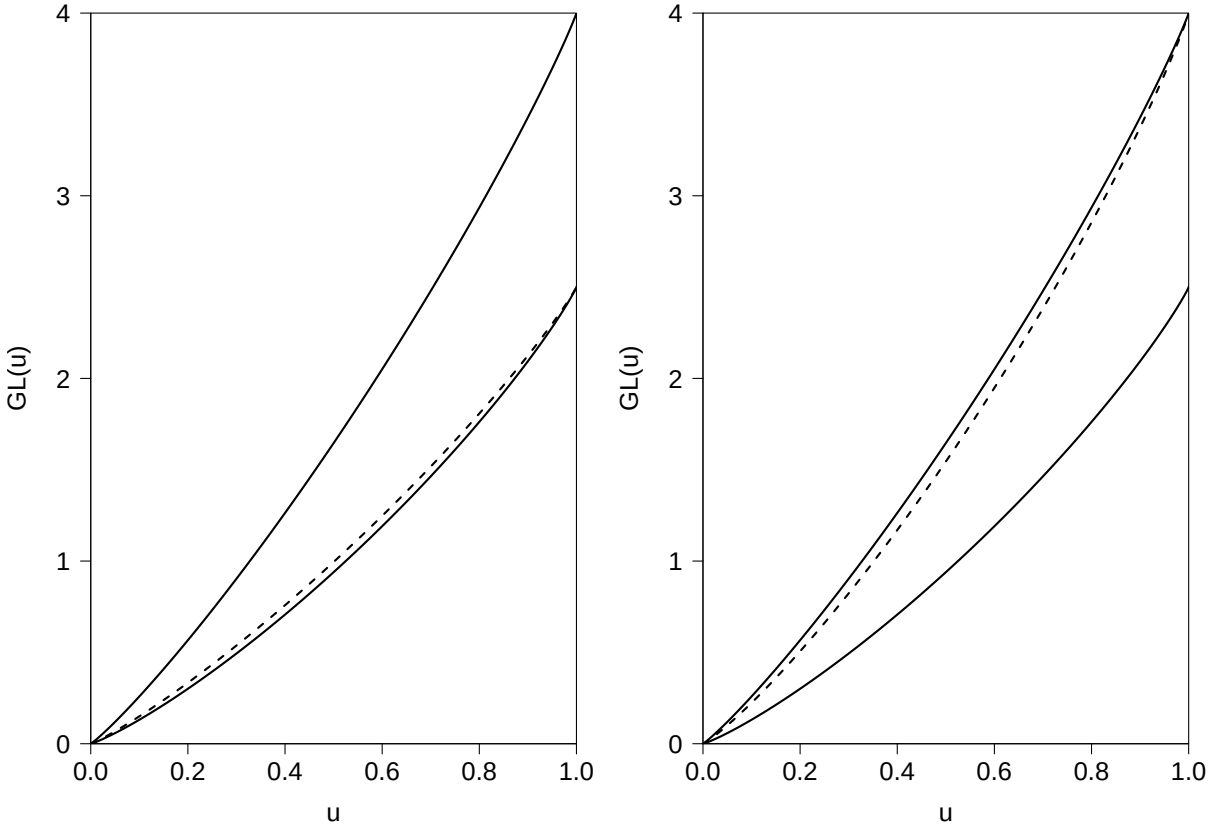

Figure 3: Generalized Lorenz curves for two gamma distributions: Ga(20,5) (top) and Ga(10,4) (bottom). Two intermediate generalized Lorenz curves (dashed) as described in Theorem 11: $H_1 \sim$ Ga(15,6) (left panel), $H_2 \sim$ Ga(12,3) (right panel).

8 Conclusion

Evidently, the Lorenz curve and the associated Lorenz order have considerable further potential in theoretical and applied economics. On the theoretical side, it would be interesting to explore generalizations of majorization inequalities to the Lorenz case, for example in the context of Condorcet jury theorems.

On the practical side, what appears to be lacking is a suite of tools for distributional analysis in a major statistical software package. As of early 2005, there are several standalone tools for distributional analyses that require the use of an additional statistics package for more standard tasks. I am currently working on an add-on package for the R language (R Development Core Team, 2005) that allows for the analysis of Lorenz curves and the fitting of size distributions while at the same time giving access to a large number of standard tools.

References

Aaberge, R. (2000). Characterizations of Lorenz curves and income distributions. *Social Choice and Welfare*, 17, 639–653.

Arnold, B.C. (1983). *Pareto Distributions*. Fairland, MD: International Co-operative Publishing House.

Arnold, B.C. (1987). *Majorization and the Lorenz Order*. Lecture Notes in Statistics 43, Berlin and New York: Springer.

Arnold, B.C. (1990). The Lorenz order and the effects of taxation policies. *Bulletin of Economic Research*, 42, 249–264.

Arnold, B.C., and Laguna, L. (1977). On generalized Pareto distributions with applications to income data. International Studies in Economics No. 10, Dept. of Economics, Iowa State University, Ames, Iowa.

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

Boland, P.J. (1989). Majority systems and the Condorcet jury theorem. *The Statistician*, 38, 181–189.

Bordley, R.F., McDonald, J.B., and Mantrala, A. (1996). Something new, something old: Parametric models for the size distribution of income. *Journal of Income Distribution*, 6, 91–103.

Chatelain. É. (1907). Les successions déclarées en 1905. *Revue d' Économie politique*, 160–170.

Chatelain. É. (1910). Le tracé de la courbe des successions en France. *Journal de la Société Statistique de Paris*, 51, 352–356.

Condorcet, N. (1785). Essai sur l'application de l' analyse à la probabilité des décisions rendues à la pluralité des voix. Paris: Imprimerie Royale.

Csörgő, M., Csörgő, S., and Horvath, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes*. Lecture Notes in Statistics 33, Berlin and New York: Springer.

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

David, H.A. (1981). *Order Statistics*, 2nd ed. New York: John Wiley.

Derobert, L., and Thieriot, G. (2003). The Lorenz curve as an archetype: A historico-epistemological study. *European Journal of the History of Economic Thought*, 10, 573–585.

Donaldson, D. and Weymark, J.A. (1980). A single-parameter generalization of the Gini index of inequality. *Journal of Economic Theory*, 22, 67–86.

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

Fellman, J. (1976). The effect of transformations on Lorenz curves. *Econometrica*, 44, 823–824.

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

Gleser, L.J. (1975). On the distribution of the number of successes in independent trials. *Annals of Probability*, 3, 182–188.

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. English translation (2005) in *Metron*, 63, 3–38.

Goldie, C. (1977). Convergence theorems for empirical Lorenz curves and their inverses. *Advances in Applied Probability*, 9, 765–791.

Grofman, B. and Owen, G. (1989). Condorcet models, avenues for future research. In: B. Grofman and G. Owen: *Information Pooling and Group Decision Making*, Greenwich, CT: JAI Press, 93–102.

Hardy, G.H., Littlewood, J.E., and Pólya, G. (1929). Some simple inequalities satisfied by convex functions. *Messenger of Mathematics*, 58, 145–152.

Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *Annals of Mathematical Statistics*, 27, 713–721.

Iritani, J., and Kuga, K. (1983). Duality between the Lorenz curves and the income distribution functions. *Economic Studies Quarterly*, 34, 9–21.

Jakobsson, U. (1976). On the measurement of the degree of progression. *Journal of Public Economics*, 5, 161–168.

Kakwani, N. (1977). Applications of Lorenz curves in economic analysis. *Econometrica*, 45, 719–727.

Kakwani, N. (1980). On a class of poverty measures. *Econometrica*, 48, 437–446.

Kakwani, N. (1984). Welfare ranking of income distributions. *Advances in Econometrics*, 3, 191–213.

Kamps, U. (1998). Characterizations of distributions by recurrence relations and identities for moments of order statistics. In: Balakrishnan, N., and Rao, C.R. (eds.): *Handbook of Statistics*, Vol. 16. Amsterdam: Elsevier.

King, W.I. (1912). *The Elements of Statistical Method*. New York: Macmillan.

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. (2000). *Halbordnungen von Einkommensverteilungen*. Angewandte Statistik und Ökonometrie, Vol. 47. Göttingen: Vandenhoeck & Ruprecht.

Kleiber, C. (2002). Variability ordering of heavy-tailed distributions with applications to order statistics. *Statistics & Probability Letters*, 58, 381–388.

Kleiber, C. (2004). Lorenz ordering of order statistics from log-logistic and related distributions. *Journal of Statistical Planning and Inference*, 120, 13–19.

Kleiber, C. (2005). A majorization approach to Condorcet jury theorems. Working paper, Universität Dortmund.

Kleiber, C., and Kotz, S. (2002). A characterization of income distributions in terms of generalized Gini coefficients. *Social Choice and Welfare*, 19, 789–794.

Kleiber, C., and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken, NJ: John Wiley.

Kleiber, C., and Krämer, W. (2003). Efficiency, equity, and generalized Lorenz dominance. *Estadística*, 55 (Special Issue on Income Distribution, Inequality and Poverty, ed. C. Dagum), 173–186.

Lambert, P.J. (2001). *The Distribution and Redistribution of Income*, 3rd ed. Manchester: Manchester University Press.

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.

Marshall, A.W., and Olkin, I. (1974). Majorization in multivariate distributions. *Annals of Statistics*, 2, 1189–1200.

Marshall, A.W., and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Orlando, FL: Academic Press.

McDonald, J.B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52, 647–663.

Mosler, K. (1994). Majorization in economic disparity measures. *Linear Algebra and Its Applications*, 199, 91–114.

Muliere, P., and Scarsini, M. (1989). A note on stochastic dominance and inequality measures. *Journal of Economic Theory*, 49, 314–323.

Piesch, W. (1967). Konzentrationsmaße von aggregierten Verteilungen. In: A.E. Ott (ed.): *Theoretische und empirische Beiträge zur Wirtschaftsforschung*. Tübingen: J.C.B. Mohr (Paul Siebeck), pp. 269–280.

Piesch, W. (1971). Lorenzkurve und inverse Verteilungsfunktion. *Jahrbücher für Nationalökonomie und Statistik*, 185, 209–234.

Piesch, W. (1975). *Statistische Konzentrationsmaße*. Tübingen: J.C.B. Mohr (Paul Siebeck).

Pietra, G. (1915). Delle relazioni fra indici di variabilità, note I e II. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, 74, 775–804.

R Development Core Team (2005). R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.r-project.org/`

Ramos, H.M., Ollero, J., and Sordo, M.A. (2000). A sufficient condition for generalized Lorenz order. *Journal of Economic Theory*, 90, 286–292.

Sen, A.K. (1973). *On Economic Inequality*. Oxford: Clarendon Press.

Shorrocks, A.F. (1983). Ranking income distributions. *Economica*, 50, 3–17.

Singh, S.K., and Maddala, G.S. (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. *Statistical Distributions in Scientific Work*, 6, 181–192.

Thistle, P.D. (1989a). Duality between generalized Lorenz curves and distribution functions. *Economic Studies Quarterly*, 40, 183–187.

Thistle, P.D. (1989b). Ranking distributions with generalized Lorenz curves. *Southern Economic Journal*, 56, 1–12.

Thompson, W.A., Jr. (1976). Fisherman's luck. *Biometrics*, 32, 265–271.

Tong, Y.L. (1988). Some majorization inequalities in multivariate statistical analysis. *SIAM Review*, 30, 602–622.

Tong, Y.L. (1994). Some recent developments on majorization inequalities in probability and statistics. *Linear Algebra and Its Applications*, 199, 69–90.

Venter, G. (1983). Transformed beta and gamma distributions and aggregate losses. *Proceedings of the Casualty Actuarial Society*, 70, 156–193.

Wilfling, B., and Krämer, W. (1993). Lorenz ordering of Singh-Maddala income distributions. *Economics Letters*, 43, 53–57.

Yitzhaki, S. (1983). On an extension of the Gini inequality index. *International Economic Review*, 24, 617–628.