The Lamé class of Lorenz curves

José María Sarabia, Vanesa Jordá, and Carmen Trueba
Department of Economics, University of Cantabria, Santander, Spain

ABSTRACT
In this paper, the class of Lamé Lorenz curves is studied. This family has the advantage of modeling inequality with a single parameter. The family has a double motivation: it can be obtained from an economic model and from simple transformations of classical Lorenz curves. The underlying cumulative distribution functions have a simple closed form, and correspond to the Singh–Maddala and Dagum distributions, which are well known in the economic literature. The Lorenz order is studied and several inequality and polarization measures are obtained, including Gini, Donaldson–Weymark–Kakwani, Pietra, and Wolfson indices. Some extensions of the Lamé family are obtained. Fitting and estimation methods under two different data configurations are proposed. Empirical applications with real data are given. Finally, some relationships with other curves are included.

1. Introduction
The Lorenz curve (LC) is an important and convenient statistical instrument used in the analysis of economical data. The setting of obtaining an adequate functional form for this curve continues being an active research field. Different parametric models for LCs have been proposed by Kakwani and Poddar (1976), Rasche et al. (1980), Aggarwal (1984), Aggarwal and Singh (1984), Arnold (1986), Pakes (1986), Arnold et al. (1987), Villaseñor and Arnold (1989), Basmann et al. (1990), Ortega et al. (1991), Chotikapanich (1993), Holm (1993), Ryu and Slottje (1996), Sarabia (1997), Sarabia et al. (1999, 2001), Ogwang and Rao (1996, 2000), Sarabia and Pascual (2002), Sarabia et al. (2005, 2010), Rhode (2009), and Helene (2010).

In a recent paper, Henle et al. (2008) introduced a family of LCs, the so-called Lamé family, which is characterized by a single inequality parameter. This family presents several practical and theoretical advantages. The authors obtain the family from an economic model (see Sec. 3). On the other hand, the idea of describing the inequality by a single parameter is quite interesting from an economic point of view, and has been supported by Slottje (2010). The family of Lamé LCs is defined by two simple functional forms, which depend on one single parameter. One of the main advantages of this family is that we can obtain estimates of the Gini index and others inequality measures very similar to those obtained with other more complicated functional forms.

In this paper, the class of Lamé LCs is studied. This family has the advantage of modeling inequality with a single parameter, and many of the most important statistical and economic...
measures for studying the inequality can be obtained in a closed form. The family has a double motivation: it can be obtained from an economic model and from a simple transformation of the potential and classical Pareto LCs. The underlying cumulative distribution functions (CDFs) have a simple closed form, and correspond to the Singh–Maddala and Dagum distributions, which are very popular in the economic literature. The Lorenz order is studied and several inequality and polarization measures are obtained, including Gini, Donaldson–Weymark–Kakwani, Pietra, and Wolfson indices. Two direct and simple extensions of the Lamé family are obtained. Fitting and estimation methods under two different data configurations are proposed. Two empirical applications with real data are presented. Finally, some relationships with other curves are included.

The rest of this paper is organized as follows. In Sec. 2 the basic results, which will be used along the following paper, will be introduced. The family of Lamé LCs is introduced in Sec. 3 and the underlying CDF is obtained in Sec. 4. The Lorenz ordering and the inequality and polarization measures are studied in Sec. 5. Some extensions of the Lamé LCs are discussed in Sec. 6. Estimation methods are proposed in Sec. 7. Section 8 includes two empirical applications. Finally, some relationships with other curves are included in Sec. 9.

2. Previous results

In this paper, we consider the LC as defined by Gastwirth (1971). Let \( \mathcal{L} \) be the class of all nonnegative random variables with positive finite expectation \( \mu \). For a random variable \( X \) in \( \mathcal{L} \) with distribution function \( F_X \) we define its inverse distribution function \( F_X^{-1} \) by

\[
F_X^{-1}(y) = \sup\{x : F_X(x) \leq y\}.
\]

Thus, the LC associated with \( X \) is defined by

\[
L_X(p) = \left[ \int_0^p F_X^{-1}(y)dy \right] / \left[ \int_0^1 F_X^{-1}(y)dy \right], \quad 0 \leq p \leq 1.
\]  

(2.1)

Note that \( \mu_X = \int_0^1 F_X^{-1}(y)dy \) is the expectation of the random variable \( X \). A detailed discussion about these concepts can be found in Arnold (1987), Johnson, Kotz and Balakrishnan (1994, Chapter 12), and Sarabia (2008a).

A characterization of an LC attributed to Gaffney and Anstis by Pakes (1986) is given by the following theorem.

Theorem 2.1. Suppose that \( L(p) \) is defined and continuous on \( [0, 1] \) with second derivative \( L''(p) \). The function \( L(p) \) is an LC if and only if

\[
L(0) = 0, \quad L(1) = 1, \quad L'(0^+) \geq 0, \quad L''(p) \geq 0 \text{ for } p \in (0, 1).
\]  

(2.2)

Note that \( L'(0^+) \) means the value of first derivative of function \( L(p) \) for \( p = 0 \) to the right.

The following lemma considers the composition of LCs.

Lemma 2.1. Let \( L_i(p), i = 1, 2 \) two genuine LCs. Then, the functional form:

\[
L_{12}(p) = L_1[L_2(p)], \quad 0 \leq p \leq 1,
\]

defines a new genuine LC.

Proof. The proof is direct by checking conditions (2.2) in Theorem 2.1. \( \Box \)
Using (2.1) it is clear that 
\[ \frac{F^{-1}}{X} (y) = \mu X L_X' (y). \]
In relation to the probability density function (PDF) \( f(x) \) associated with an LC \( L(p) \) we have the following (Arnold, 1987) theorem.

**Theorem 2.2.** If the second derivative \( L''(p) \) exists and is positive everywhere in an interval \((x_1, x_2)\), the corresponding CDF \( F \) has a finite positive PDF in the interval \((\mu L'(x_1^+), \mu L'(x_2^-))\), which is given by

\[ f(x) = \frac{1}{\mu L''[F(x)]}. \]

**3. The family**

The Lamé class of LCs is given by the curves:

\[ L_1(p; a) = \left[ 1 - (1 - p)^a \right]^{1/a}, \quad 0 \leq p \leq 1, \quad (3.1) \]

\[ L_2(p; a) = 1 - (1 - p)^{1/a}, \quad 0 \leq p \leq 1. \quad (3.2) \]

It can be proved using Theorem 2.1 that (3.1) is a genuine LC if \( 0 < a \leq 1 \) and (3.2) is a genuine LC if \( a \geq 1 \). Note that for both the curves the case \( a = 1 \) corresponds to the egalitarian line.

**3.1. Motivation**

The family has a double motivation: it can be obtained from an economic model and from simple transformations of classical LCs.

In relation with the economic motivation, Henle et al. (2008) obtained the curves (3.1) and (3.2) making use of two different economic theories, the so-called trickle-up and trickle-down effects. The first approach supposes that an increase in income of the lower-middle class would be more advantageous for the economy since they spend their wealth faster than the upper class. In contrast, trickle-down theory assumes that an increase in income of the upper class would stimulate the investment thus resulting in a fall of unemployment. This situation would benefit the general welfare. Then, the trickle-up effect is related with the following expression:

\[ \frac{\partial I}{\partial t} = \frac{\Delta L}{N(1 - r)}, \quad (3.3) \]

and the trickle-down theory with

\[ \frac{\partial I}{\partial t} = \frac{\Delta (1 - L)}{N r}, \quad (3.4) \]

where \( I \) is the income of a family at rank \( r \), \( \Delta L \) is the aggregate income of poorer citizens, \( N(1 - r) \) is the number of wealthy individuals, \( \Delta (1 - L) \) is the income of the wealthiest citizens, and \( N r \) is the number of individuals of lower rank. Note that following Equation (3.3), the increase in income of any individual of the society is directly associated with an improvement of the economic situation of low rank citizens. Conversely, Equation (3.4) shows an increment of the upper class income would lead an economic progress of the society as a whole. Finally, the curves (3.1) and (3.2) are the solutions of Equations (3.3) and (3.4).

The second motivation is based on Lemma 2.1. If \( L_1(p) = p^{1/a} \) is the potential LC and \( L_2(p) = 1 - (1 - p)^a \) with \( 0 < a \leq 1 \) is the LC corresponding to the classical Pareto distribution, the family (3.1) is \( L_1(L_2(p)) \), which is a genuine LC according to Lemma 2.1. The
second curve (3.2) is of the form \( \tilde{L}_1(\tilde{L}_2(p)) \), where \( \tilde{L}_1(p) = 1 - (1 - p)^{1/a} \) and \( \tilde{L}_2(p) = p^a \), where \( a \geq 1 \).

4. The underlying CDF

The underlying CDFs are obtained in a closed form in the following theorem.

**Theorem 4.1.** The underlying CDF associated to the LCs (3.1) and (3.2) are given by

\[
F_1(x; a, \mu) = 1 - \frac{1}{\left[1 + (x/\mu)^{a/(1-a)}\right]^{1/a}}, \quad 0 \leq x < \infty,
\]

and \( F_1(x; a, \mu) = 0 \) if \( x < 0 \) and

\[
F_2(x; a, \mu) = \frac{1}{\left[1 + (\mu/x)^{a/(a-1)}\right]^{1/a}}, \quad 0 \leq x < \infty,
\]

and \( F_2(x; a, \mu) = 0 \) if \( x < 0 \), respectively, where \( \mu \) represents the mathematical expectation of the population.

**Proof.** For the LC (3.1) we have

\[
F_1^{-1}(x) = \mu(1 - x)^{a-1}[1 - (1 - x)^{a}]^{(1-a)/a},
\]

with support \((0, \infty)\). The change of variable \( z = (1 - x)^{a} \) leads to \( F_1^{-1}(z) = \mu \left( \frac{1-z}{z} \right)^{(1-a)/a} \). Finally, solving previous equation for \( z \) and then for \( x \) we obtain (4.1). Using a similar reasoning with (3.2), we obtain (4.2). \( \square \)

Both distributions correspond to well-known distributions used in the income and wealth literature. The family (4.1) is a Singh–Maddala distribution (Singh and Maddala, 1976) with shape parameters \( \frac{a}{1-a} \) and \( \frac{1}{a} \) and scale parameter \( \mu \). The raw moments of (4.1) are

\[
E(X_1^k) = \mu^k \Gamma \left( \frac{1 + k(1-a)}{a} \right) \Gamma \left( \frac{1 - k(1-a)}{a} \right) \Gamma \left( \frac{1}{a} \right),
\]

if \( a > 1 - 1/k \), where \( \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \) denotes the gamma function. Note that \( E(X_1) = \mu \). The mode is at

\[
\mu \left( \frac{2a - 1}{2 - a} \right)^{(1-a)/a},
\]

if \( a > 1/2 \).

The family (4.2) is a Dagum (1977) distribution with shape parameters \( \frac{a}{a-1} \) and \( \frac{1}{a} \) and scale parameter \( \mu \). In this case, the raw moments are

\[
E(X_2^k) = \mu^k \Gamma \left( \frac{1 + k(a-1)}{a} \right) \Gamma \left( \frac{1 - k(a-1)}{a} \right) \Gamma \left( \frac{1}{a} \right),
\]
if \( a > k/(k - 1) \). The mode is at
\[
\mu \left( \frac{2 - a}{2a - 1} \right)^{(a-1)/a},
\]
if \( a < 2 \).

### 5. Lorenz ordering, inequality, and polarization measures

#### 5.1. Lorenz ordering

The study of Lorenz ordering is a crucial aspect in the analysis of income and wealth distributions. Let \( \mathcal{L} \) be the class of all non negative random variables with positive finite expectation. The Lorenz partial order \( \leq_L \) on the class \( \mathcal{L} \) is defined by
\[
X \leq_L Y \iff L_X(p) \geq L_Y(p), \quad \forall p \in [0, 1].
\]
If \( X \leq_L Y \), then \( X \) exhibits less inequality than \( Y \) in the Lorenz sense. We shall show that families (3.1) and (3.2) are ordered with respect to \( a \) parameters.

**Lemma 5.1.** If \( L_1(p; a) \) is defined in (3.1) and if \( a_1 \leq a_2 \), then \( L_1(p; a_1) \leq L_1(p; a_2) \), for \( 0 \leq p \leq 1 \). On the other hand, if \( a_1 \leq a_2 \) then \( L_2(p; a_1) \geq L_2(p; a_2) \), for \( 0 \leq p \leq 1 \).

**Proof.** If \( 0 < a \leq 1 \), \( L_1(p; a) \) is a differentiable function with respect to \( a \), in consequence:
\[
\frac{\partial L_1(p; a)}{\partial a} = -L_1(p; a) \left\{ \frac{\log[1 - (1 - p)^a]}{a^2} + \frac{(1 - p)^a \log(1 - p)}{a[1 - (1 - p)^a]} \right\} > 0,
\]
for all \( p \in (0, 1) \). For the \( L_2(p; a) \) curve:
\[
\frac{\partial L_2(p; a)}{\partial a} = (1 - p^a)^{1/a} \left\{ \frac{p^a \log p}{a(1 - p^a)} + \frac{\log(1 - p^a)}{a^2} \right\} < 0,
\]
for all \( p \in (0, 1) \). \( \square \)

#### 5.2. Gini index

**Lemma 5.2.** The Gini indices of the curves \( L_1(p; a) \) and \( L_2(p; a) \) are given by
\[
G_1(a) = 1 - \frac{\Gamma(1/a)^2}{a \Gamma(2/a)}, \quad 0 < a \leq 1,
\]
\[
G_2(a) = \frac{\Gamma(1/a)^2}{a \Gamma(2/a)} - 1, \quad a \geq 1,
\]
respectively.

**Proof.** The Gini index is given by \( G = 2 \int_0^1 [p - L(p)]dp = 1 - 2 \int_0^1 L(p)dp \), and substituting (3.1) and (3.2) in previous equation, we obtain easily (5.1) and (5.2), respectively. \( \square \)

Some configurations of the Gini index deserve our attention. If \( a = \frac{1}{n} \), with \( n = 1, 2, \ldots \) the Gini index (5.1) can be written in the simple form:
\[
G_1(n) = 1 - \frac{(n!)^2}{n(2n - 1)!},
\]
which is of the form \( \frac{m}{m+1} \), where \( m \) is an integer. For example, for \( n = 1, 2, \ldots, 10 \) the list of the Gini indices are

\[
0, \frac{2}{3}, \frac{9}{10}, 34, 125, 461, 1715, 6434, 24309, 92377
\]

5.3. Donaldson–Weymark–Kakwani index

An important generalization of the Gini index was proposed by Donaldson and Weymark (1980) and Kakwani (1980) and studied by Yitzhaki (1983). These authors proposed the generalized Gini index defined as

\[
G_X(v) = 1 - v(v + 1) \int_0^1 (1 - p)^{-1} L_X(p) dp,
\]

(5.3)

where \( v > 1 \) and \( L_X(\cdot) \) is the LC. If we set \( v = 1 \) in (5.3) we obtain the Gini index. When \( v \) increases, higher weights are attached to small incomes. The limit case when \( v \) goes to infinity depends on the lowest income, expressing the judgement introduced by Rawls, that social welfare depends only on the poorest society member.

Lemma 5.3. The Donaldson–Weymark–Kakwani indices of the curves \( L_1(p, a) \) and \( L_2(p, a) \) are given by

\[
GL_1(v; a) = 1 - v(v + 1) \sum_{k=0}^{\infty} \frac{\Gamma(k - 1/a)}{\Gamma(k + 1) \Gamma(-1/a)} \frac{1}{v + ak},
\]

(5.4)

\[
GL_2(v; a) = 1 - v(v + 1) \left( \frac{1}{v} - \sum_{k=0}^{\infty} \frac{\Gamma(k - 1/a)}{\Gamma(k + 1) \Gamma(-1/a)} B(ak + 1, v) \right),
\]

(5.5)

respectively, where \( B(x, y) = \int_0^1 t^{x-1} (1 - t)^{y-1} dt \) denotes the beta function.

Proof. Using general expression (5.3) and the series expansion of the binomial expression.

In the case of the curve (3.1), if \( v \) is an integer, expression (5.4) can be written in a simple form. It can be proved (Muliere and Scarsini, 1989) that Equation (5.3) is equivalent to

\[
G_X(v) = 1 - \frac{E(X_{1:v})}{E(X)},
\]

(5.6)

where \( X_{1:v} \) represents the minimum random variable in a random sample of size \( v \). In our case, the distribution of the minimum of a Singh–Maddala distribution is again a Singh–Maddala distribution with parameters \( \frac{a}{1-a}, \frac{v}{a} \), and \( \mu \). Using (5.6), we obtain the expression:

\[
GL_1(v; a) = 1 - \frac{\Gamma(\frac{1}{a}) \Gamma\left(\frac{v+1+a}{a}\right)}{\Gamma\left(\frac{1}{a}\right)}.
\]

Other simple expressions can also be obtained in the case of \( L_2 \). For example, if \( v = 2 \), Equation (5.5) can be written of the form:

\[
GL_2(2; a) = 1 - 3 \left[ 1 - \Gamma\left(1 + \frac{1}{a}\right) \left( \frac{2(a-2)/\sqrt{\pi}}{\Gamma(\frac{2+a}{2a})} - \frac{\Gamma\left(\frac{2+a}{a}\right)}{\Gamma\left(\frac{3+a}{a}\right)} \right) \right].
\]
and if \( \nu = 3 \), Equation (5.5) becomes

\[
G_{L_2}(3; a) = 1 - 4 \left[ 1 - 3\Gamma \left( 1 + \frac{1}{a} \right) \left( \frac{\Gamma \left( \frac{1+a}{a} \right)}{\Gamma \left( \frac{2+a}{a} \right)} - \frac{\Gamma \left( \frac{2+a}{a} \right)}{\Gamma \left( \frac{3+a}{a} \right)} + \frac{\Gamma \left( \frac{3+a}{a} \right)}{3\Gamma \left( \frac{4+a}{a} \right)} \right) \right].
\]

### 5.4. Pietra index

The Pietra index is defined as the maximal vertical deviation between the LC and the egalitarian line:

\[
P_L = \max_{0 \leq p \leq 1} \{ p - L_X(p) \}.
\]

If we assume that \( F_X \) 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 \( 1 - F_X^{-1}(x)/\mu_X \) is zero, that is, when \( x = F_X(\mu_X) \). The value of \( p - L_X(p) \) in this point is given by

\[
P_L = F_X(\mu_X) - L_X(F_X(\mu_X)).
\] (5.7)

This lemma provides a simple expression for the Pietra's indices of the curves \( L_1 \) and \( L_2 \).

**Lemma 5.4.** *The Pietra indices of the curves \( L_1(p, a) \) and \( L_2(p, a) \) are given by*

\[
P_1(a) = 1 - \frac{1}{2^{1/a-1}}, \quad 0 < a \leq 1,
\] (5.8)

\[
P_2(a) = \frac{1}{2^{1/a-1}} - 1, \quad a \geq 1,
\] (5.9)

*respectively.*

**Proof.** The proof is direct using formula (5.7) and the expressions (4.1) and (4.2). \( \square \)

### 5.5. Polarization index

The polarization measurement has been recently proposed as an important variable to characterize income and wealth distributions. It is well known, that polarization is widely accepted as a different concept from inequality. A polarization measure concentrates the income distribution on several polar modes. On the other hand, inequality relates to the overall dispersion of the distribution. A more bipolarized income and wealth distribution is one that is more spread out from the middle, thus implying that the middle class is made up of fewer individuals (Wolfson, 1994).

The Wolfson's index of bipolarization was originally proposed for a population divided in two groups by the median value. This measure is given by

\[
W_X = \frac{2(1/2 - L_X(1/2)) - G_X}{m_X/2\mu_X},
\] (5.10)

where \( G_X, m_X, \) and \( \mu_X \) represent the Gini index, median, and mean associated to the LC \( L_X \), respectively.

We have the following lemma.
Lemma 5.5. The Wolfson polarization measures corresponding to the Lamé LCs (3.1) and (3.2) are given by

\[ W_{X_1} = 2 \frac{\Gamma(1/a)^2}{\Gamma(2/a)} - \frac{2(1 - 1/2^a)^{1/a}}{(2^a - 1)^{1/(1-a)/a}}, \]  
(5.11)

\[ W_{X_2} = 2 \frac{(1 - 1/2^a)^{1/a}}{(2^a - 1)^{1/(1-a)/a}} - \frac{\Gamma(1/a)^2}{2 \Gamma(1/a)^2}, \]  
(5.12)

respectively.

Proof. Previous expressions are obtained from (5.10), using (3.1) and (3.2) together with (4.1) and (4.2) for obtaining the median. \□

6. Extensions

Recent research about the LC (Basmann et al., 1990; Ryu and Slottje, 1996; Sarabia et al., 2005, Sarabia, 2008a) has shown that some families of LCs approximate only some segments of the income distribution but not others. This fact justifies the consideration of more complex models for the LC beginning with an initial LC.

The following two direct extensions of (3.1) and (3.2) can be considered:

\[ L_3(p; a, b) = \left[ 1 - (1 - p)^a \right]^b, \quad 0 \leq p \leq 1, \]  
(6.1)

\[ L_4(p; a, b) = 1 - (1 - p^b)^a, \quad 0 \leq p \leq 1. \]  
(6.2)

Curve (6.1) is genuine if \( 0 < a \leq 1 \leq b \) and (6.2) is genuine if \( a \geq 1 \) and \( 0 < b \leq 1 \). Curve (6.1) was proposed by Rasche et al. (1980) and (6.2) was considered (without to be studied) by Arnold (1987). The Gini index of the LC (6.2) is

\[ G_{L_4}(a, b) = \frac{2}{a} \frac{\Gamma\left(\frac{1}{a}\right)}{\Gamma\left(\frac{1}{a} + b + 1\right)} \Gamma\left(\frac{1}{a} + b + 1\right) - 1, \]

which will be used in the next section. Other kind of extensions can also be considered using the methodologies proposed by Sarabia et al. (1999, 2005).

7. Estimation

In this section we consider two different estimation methods for two different data configurations.

7.1. Estimation from data of the LC

Let us suppose that we wish to estimate the parameter \( a \) from the curves \( L_1 \) and \( L_2 \). Let \( X_1, \ldots, X_n \) be a sample of size \( n \) of income data. The observations consist of \( n \) pairs of points \((p_i, q_i), \ldots, (p_n, q_n)\), where \( p_i = i/n, q_i = s_i/s_n \) and \( s_i = x_{i:n} + \cdots + x_{i:n} \) for \( i = 1, 2, \ldots, n \), being \( x_{i:n} \), \( i = 1, 2, \ldots, n \) the \( i \)th order statistic.

The simplest way of estimating \( a \) parameter is minimizing the quantity:

\[ u(a) = \sum_{i=1}^{n} \left[ q_i - L_k(p_i; a) \right]^2, \quad k = 1, 2, \]  
(7.1)
or alternatively,

\[ u_1(a) = \sum_{i=1}^{n} \left[ (1 - q_i^a) - (1 - p_i)^a \right]^2, \tag{7.2} \]

for \( L_1 \) and

\[ u_2(a) = \sum_{i=1}^{n} \left[ (1 - q_i)^a - 1 + p_i^a \right]^2, \tag{7.3} \]

for \( L_2 \), using a non-linear optimization standard algorithm, which is available in any of the existing statistical and econometric software packages, including SAS, SPSS, SHAZAM, EViews, and Mathematica. We can take \( a_0 = 1 \) as initial value. An alternative robust method of estimation is given by Castillo et al. (1998).

### 7.2. Estimation with limited information

The practical use of the LC requires information on the income and wealth distribution or at least, the income shares for a number of income classes. Unfortunately, sometimes this information is not available. However, in some economic database the available comparable cross-country information is limited to per capita income and the Gini index (see, for example, Chotikapanich et al., 1997).

Then, assuming that the only available information for the estimation of (3.1) and (3.2) (or (4.1) and (4.2)) is the mean and the Gini index. A plausible estimation method which gives place to consistent estimates, consists of solving the system:

\[
\mu = \bar{X}, \\
G_k(a) = g, \quad k = 1, 2,
\]

where \( \bar{X} \) and \( g \) represent the mean and the Gini sample values, respectively, and \( G_k(a), k = 1, 2 \) the theoretical Gini indices (5.1) and (5.2), respectively. Therefore, the estimate values of \( \mu \) and \( a \) are given by

\[
\hat{\mu} = \bar{X}, \\
\hat{a} = G_k^{-1}(g), \quad k = 1, 2. \tag{7.4}
\]

Note that the right-hand side of (7.4) is a monotonic function of parameter \( g \), and consequently has only one solution.

### 8. Empirical applications

#### 8.1. Estimation when some point of the LC are available

Results of the estimation of the LCs studied and the corresponding Gini index are presented in this section. The source of data used for this purpose is taken from Shorrocks (1983). The data correspond to cumulated income shares for 19 countries derived from Jain (1975). This data set is relevant since the sample considered is characterized by heterogeneity among countries in terms of income inequality, thus allowing us to draw conclusions regarding the adjustment of the LCs for different levels of inequality.

Tables 1–4 present the estimation results for the two Lamé LCs (3.1) and (3.2) and for the two extensions (6.1) and (6.2), respectively. The estimators have been obtained by non
Table 1. Fitted Lorenz Curves, error measures, and Gini index for the Lorenz curve (3.1). Standard errors in parentheses.

| Country      | a       | MSE $\times 10^2$ | MAE   | MAX   | Gini index |
|--------------|---------|-------------------|-------|-------|------------|
| Brazil       | 0.5051  | 0.0824            | 0.0254| 0.0592| 0.6587     |
| (0.0091)     |         |                   |       |       |            |
| Columbia     | 0.5646  | 0.0230            | 0.0125| 0.0343| 0.5658     |
| (0.0051)     |         |                   |       |       |            |
| Denmark      | 0.7013  | 0.0170            | 0.0111| 0.0288| 0.3609     |
| (0.0053)     |         |                   |       |       |            |
| Finland      | 0.6314  | 0.0142            | 0.0109| 0.0191| 0.4634     |
| (0.0044)     |         |                   |       |       |            |
| India        | 0.6246  | 0.1010            | 0.0262| 0.0701| 0.4736     |
| (0.0117)     |         |                   |       |       |            |
| Indonesia    | 0.6264  | 0.2341            | 0.0408| 0.1040| 0.4709     |
| (0.0178)     |         |                   |       |       |            |
| Japan        | 0.7610  | 0.0008            | 0.0023| 0.0050| 0.2785     |
| (0.0012)     |         |                   |       |       |            |
| Kenya        | 0.5072  | 0.1537            | 0.0355| 0.0765| 0.6555     |
| (0.0124)     |         |                   |       |       |            |
| Malaysia     | 0.5980  | 0.0042            | 0.0052| 0.0152| 0.5142     |
| (0.0023)     |         |                   |       |       |            |
| Netherlands  | 0.6426  | 0.0005            | 0.0018| 0.0044| 0.4466     |
| (0.0008)     |         |                   |       |       |            |
| New Zealand  | 0.6997  | 0.0144            | 0.0110| 0.0170| 0.3632     |
| (0.0049)     |         |                   |       |       |            |
| Norway       | 0.7067  | 0.0183            | 0.0125| 0.0186| 0.3533     |
| (0.0056)     |         |                   |       |       |            |
| Panama       | 0.6430  | 0.0006            | 0.0020| 0.0048| 0.4460     |
| (0.0009)     |         |                   |       |       |            |
| Sri Lanka    | 0.6690  | 0.0027            | 0.0041| 0.0116| 0.4075     |
| (0.0020)     |         |                   |       |       |            |
| Sweden       | 0.6859  | 0.0059            | 0.0068| 0.0151| 0.3830     |
| (0.0030)     |         |                   |       |       |            |
| Tanzania     | 0.5652  | 0.1258            | 0.0314| 0.0680| 0.5649     |
| (0.0120)     |         |                   |       |       |            |
| Tunisia      | 0.6036  | 0.0102            | 0.0087| 0.0217| 0.5056     |
| (0.0036)     |         |                   |       |       |            |
| United Kingdom| 0.6996 | 0.0034            | 0.0043| 0.0140| 0.3634     |
| (0.0023)     |         |                   |       |       |            |
| Uruguay      | 0.6138  | 0.0274            | 0.0132| 0.0378| 0.4900     |
| (0.0059)     |         |                   |       |       |            |

Linear least squares, according to the method established in the previous section, minimizing expression (7.1) or (7.2) and (7.3). Using a consistent estimate of the covariance matrix of the coefficients (Amemiya, 1985) we computed the corresponding standard errors. These tables also give several error measures. The mean square error (MSE), which is given by

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left[ q_i - L(p_i; \hat{\theta}) \right]^2,$$

the mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| q_i - L(p_i; \hat{\theta}) \right|,$$

and the maximum absolute error (MAX):

$$\text{MAX} = \max_{i=1,...,n} \left| q_i - L(p_i; \hat{\theta}) \right|,$$
Table 2. Fitted Lorenz Curves, error measures, and Gini index for the Lorenz curve (3.2). Standard errors in parentheses.

| Country       | $a$     | MSE $\times 10^2$ | MAE  | MAX  | Gini index |
|---------------|---------|-------------------|------|------|------------|
| Brazil        | 2.3455  | 0.0745            | 0.0245 | 0.0528 | 0.6594     |
|               | (0.0627)|                   |      |       |            |
| Columbia      | 1.9856  | 0.0185            | 0.0155 | 0.0284 | 0.5664     |
|               | (0.0234)|                   |      |       |            |
| Denmark       | 1.4853  | 0.0191            | 0.0116 | 0.0312 | 0.3611     |
|               | (0.0149)|                   |      |       |            |
| Finland       | 1.7006  | 0.0161            | 0.0115 | 0.0233 | 0.4637     |
|               | (0.0169)|                   |      |       |            |
| India         | 1.7262  | 0.0932            | 0.0257 | 0.0657 | 0.4743     |
|               | (0.0417)|                   |      |       |            |
| Indonesia     | 1.7197  | 0.2228            | 0.0403 | 0.0994 | 0.4717     |
|               | (0.0640)|                   |      |       |            |
| Japan         | 1.4000  | 0.0011            | 0.0027 | 0.0058 | 0.3125     |
|               | (0.0032)|                   |      |       |            |
| Kenya         | 2.3306  | 0.1442            | 0.0346 | 0.0701 | 0.6562     |
|               | (0.0862)|                   |      |       |            |
| Malaysia      | 1.8310  | 0.0027            | 0.0044 | 0.0101 | 0.5147     |
|               | (0.0078)|                   |      |       |            |
| Netherlands   | 1.6617  | 0.0003            | 0.0016 | 0.0032 | 0.4470     |
|               | (0.0022)|                   |      |       |            |
| New Zealand   | 1.4899  | 0.0155            | 0.0113 | 0.0195 | 0.3635     |
|               | (0.0135)|                   |      |       |            |
| Norway        | 1.4715  | 0.0188            | 0.0127 | 0.0176 | 0.3535     |
|               | (0.0146)|                   |      |       |            |
| Panama        | 1.6601  | 0.0001            | 0.0008 | 0.0032 | 0.4463     |
|               | (0.0016)|                   |      |       |            |
| Sri Lanka     | 1.5766  | 0.0042            | 0.0047 | 0.0148 | 0.4078     |
|               | (0.0077)|                   |      |       |            |
| Sweden        | 1.5773  | 0.0072            | 0.0073 | 0.0180 | 0.4081     |
|               | (0.0096)|                   |      |       |            |
| Tanzania      | 1.9826  | 0.1184            | 0.0308 | 0.0622 | 0.5655     |
|               | (0.0589)|                   |      |       |            |
| Tunisia       | 0.8071  | 0.0151            | 0.0106 | 0.0264 | 0.5059     |
|               | (0.0181)|                   |      |       |            |
| United Kingdom| 1.4904  | 0.0023            | 0.0037 | 0.0113 | 0.3638     |
|               | (0.0052)|                   |      |       |            |
| Uruguay       | 1.7662  | 0.0331            | 0.0142 | 0.0423 | 0.4903     |
|               | (0.0258)|                   |      |       |            |

where $L(p; \hat{\theta})$ represents the LC estimated. It can be concluded that the empirical results reported in this study indicate that the four functions, though simple and easy to estimate, are very satisfactory in fitting data. We also present the estimates of the Gini indices associated with the four LCs under study, which are presented in the last column of each table.

From Tables 1 and 2 we observe that almost 50% of the countries considered are better fitted by the curve $L_1$, whereas the curve $L_2$ outperforms the adjustment in the rest of the countries. As consequence, we cannot conclude the superiority of one curve over the other, when the previous three error criteria are considered.

If we consider the two extensions $L_3$ and $L_4$, the superiority of the curve $L_4$ is concluded, since it adjusts the income distribution more adequately in 13 out of 19 countries included in the sample (Tables 3 and 4). It is important to note that all of these countries are characterized by low levels of inequality (Gini values lower than 0.5). On the contrary, the $L_3$ LC fits better income distributions of countries with high inequality (Gini values greater than 0.5).

Regarding the comparison of the general models $L_3$ and $L_4$ (Equations (3.1) and (3.2)) with its special cases $L_1$ and $L_2$ (Equations (6.1) and (6.2)), some comments are included. As might be expected, the extensions considered outperform single-parameter curves since the three
error measures are substantially lower for the curves $L_4$ and $L_3$. However, Gini index estimates based on Lamé curves and its extensions are practically identical, differing by no more than 0.035 and, in the majority of cases the pairwise difference is lower than 0.01. In line with other studies that compare the performance of more complex functional forms with one-parameter curves (Ogwang and Rao, 1996), we observe that the greater differences correspond to the countries that present higher levels of inequality. Thus, following the parsimony principle, the use of more complicated expressions would not be relevant for the adjustment of the Lamé LCs and the posterior estimation of the corresponding Gini index, at least for the data set considered.

### 8.2. Estimation with limited information

In this section we focus on international inequality at global level, using the methodology proposed in Sec. 7. We use the information provided by Bourguignon and Morrison (2002), which includes the estimates of several world inequality measures. Using the empirical Gini index and Equation (7.4) with $L_1$ and $L_2$, we have obtained the single-parameter estimates
Table 4. Fitted Lorenz Curves, error measures, and Gini index for the Lorenz curve (6.2). Standard errors in parentheses.

| Country     | a       | b       | MSE $\times 10^2$ | MAE     | MAX     | Gini index |
|-------------|---------|---------|-------------------|---------|---------|------------|
| Brazil      | 1.5694  | 0.3193  | 0.0012            | 0.0027  | 0.0063  | 0.6362     |
| Columbia    | 1.6846  | 0.4430  | 0.0006            | 0.0019  | 0.0054  | 0.5567     |
| Denmark     | 1.6858  | 0.7547  | 0.0024            | 0.0038  | 0.0116  | 0.3675     |
| Finland     | 1.9410  | 0.6596  | 0.0003            | 0.0016  | 0.0026  | 0.4711     |
| India       | 1.2429  | 0.4479  | 0.0073            | 0.0071  | 0.0076  | 0.4560     |
| Indonesia   | 1.0223  | 0.3927  | 0.0164            | 0.0108  | 0.0260  | 0.4432     |
| Japan       | 1.3774  | 0.7041  | 0.0008            | 0.0025  | 0.0040  | 0.3117     |
| Kenya       | 1.3159  | 0.2882  | 0.0015            | 0.0034  | 0.0065  | 0.6226     |
| Malaysia    | 1.7280  | 0.5210  | 0.0001            | 0.0008  | 0.0019  | 0.5113     |
| Netherlands | 1.6863  | 0.6093  | 0.0001            | 0.0008  | 0.0026  | 0.4479     |
| New Zealand | 1.6801  | 0.7482  | 0.0003            | 0.0013  | 0.0047  | 0.3694     |
| Norway      | 1.6723  | 0.7637  | 0.0001            | 0.0027  | 0.0083  | 0.3595     |
| Panama      | 1.6644  | 0.6037  | 0.0001            | 0.0008  | 0.0031  | 0.4465     |
| Sri Lanka   | 1.6622  | 0.6641  | 0.0016            | 0.0036  | 0.0066  | 0.4106     |
| Sweden      | 1.6551  | 0.7030  | 0.0076            | 0.0023  | 0.0054  | 0.3874     |
| Tanzania    | 1.2745  | 0.3647  | 0.0010            | 0.0028  | 0.0045  | 0.5386     |
| Tunisia     | 1.9037  | 0.5776  | 0.0131            | 0.0100  | 0.0187  | 0.5091     |
| United Kingdom | 1.4328 | 0.6483  | 0.0005            | 0.0022  | 0.0053  | 0.3691     |
| Uruguay     | 2.1110  | 0.6599  | 0.0066            | 0.0070  | 0.0146  | 0.5010     |

of the inequality indices $a$. The standard errors were computed using parametric bootstrap with $B = 999$ bootstrap replications. Using these estimators, we have obtained the Pietra and Polarization indices given in Equations (5.8)–(5.9) and (5.11)–(5.12), respectively. All these estimators are recorded in Tables 5 and 6.

The results point out that inequality has increased in the last two centuries, conclusion that is congruent with the evolution of the Gini index. Moreover, the polarization has intensified over the study period thus indicating that the differences among wealthiest and poorer countries have increased.

9. Relationships with other curves

Bonferroni (1930) considered a curve for studying inequality, which is more suitable for the analysis of low income groups. The curve, in terms of the quantile function is given by

$$B_X(p) = \frac{1}{p \mu_X} \int_0^p F_X^{-1}(u)du = \frac{L_X(p)}{p}, \quad 0 \leq p \leq 1. \tag{9.1}$$
Table 5. Estimated inequality and polarization measures with limited information using the curve (3.1). Standard errors based on parametric bootstrap in parentheses.

| Year | Gini index | a     | Pietra index | Wolfson index |
|------|------------|-------|--------------|---------------|
| 1820 | 0.5000     | 0.6073| 0.3613       | 0.4734        |
| 1850 | 0.5320     | 0.5864| 0.3867       | 0.5198        |
| 1870 | 0.5600     | 0.5683| 0.4093       | 0.5642        |
| 1890 | 0.5880     | 0.5594| 0.4324       | 0.6128        |
| 1910 | 0.6100     | 0.5363| 0.4508       | 0.6546        |
| 1929 | 0.6160     | 0.5325| 0.4559       | 0.6666        |
| 1950 | 0.6400     | 0.5171| 0.4765       | 0.7176        |
| 1960 | 0.6350     | 0.5203| 0.4722       | 0.7066        |
| 1970 | 0.6500     | 0.5107| 0.4853       | 0.7405        |
| 1992 | 0.6570     | 0.5062| 0.4914       | 0.7572        |

Table 6. Estimated inequality and polarization measures with limited information using the curve (3.2). Standard errors based on parametric bootstrap in parentheses.

| Year | Gini index | a     | Pietra index | Wolfson index |
|------|------------|-------|--------------|---------------|
| 1820 | 0.5000     | 1.7915| 0.3583       | 0.4569        |
| 1850 | 0.5320     | 1.8800| 0.3833       | 0.5003        |
| 1870 | 0.5600     | 1.9651| 0.4055       | 0.5416        |
| 1890 | 0.5880     | 2.0584| 0.4282       | 0.5870        |
| 1910 | 0.6100     | 2.1386| 0.4463       | 0.6259        |
| 1929 | 0.6160     | 2.1616| 0.4513       | 0.6371        |
| 1950 | 0.6400     | 2.2593| 0.4716       | 0.6848        |
| 1960 | 0.6350     | 2.2382| 0.4663       | 0.6744        |
| 1970 | 0.6500     | 2.3029| 0.4802       | 0.7062        |
| 1992 | 0.6570     | 2.3345| 0.4862       | 0.7218        |

In the context of informetric data, Sarabia (2008b) has proposed another curve namely the Leimkuhler curve (see also Balakrishnan et al. (2010) and Sarabia and Sarabia (2008)). This curve is defined as

\[
K_X(p) = \frac{1}{\mu_X} \int_{1-p}^{1} F_X^{-1}(u)du = 1 - L_X(1 - p), \quad 0 \leq p \leq 1. \tag{9.2}
\]

In our case, the Bonferroni and Leimkuhler curves associated to (3.1) and (3.2) can be obtained directly using expressions (9.1) and (9.2).
Acknowledgments

We are grateful for the constructive suggestions provided by the reviewers, which improved the paper.

Funding

The authors thank Ministerio de Economía y Competitividad, project ECO2010-15455, for partial support. The second author thanks Ministerio de Educación (FPU AP-2010-4907) for partial support.

References

Aggarwal, V. (1984). On optimum aggregation of income distribution data. *Sankhya B* 46:343–355.
Aggarwal, V., Singh, R. (1984). On optimum stratification with proportional allocation for a class of Pareto distributions. *Commun. Stat. Theory Methods* 13:3017–3116.
Amemiya, T. (1985). *Advanced Econometrics*. Oxford: Basic Blackwell.
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*. Vol. 43. Lecture Notes in Statistics. Berlin: Springer.
Arnold, B.C., Robertson, C.A., Brockett, P.L., Shu, B.Y. (1987). Generating ordered families of Lorenz curves by strongly unimodal distributions. *J. Bus. Econ. Stat.* 5:305–308.
Balakrishnan, N., Sarabia, J.M., Kolev, N. (2010). A simple relation between the Leimkuhler curve and the mean residual life. *J. Informetrics* 4:602–607.
Basmann, R.L., Hayes, K.L., Slottje, D.J., Johnson, J.D. (1990). A general functional form for approximating the Lorenz curve. *J. Econometrics* 92(4):727–744.
Bonferroni, C.E. (1930). *Elementi di Statistica Generale*. Firenze: Seeber.
Bourguignon, F., Morrisson, C. (2002). Inequality among world citizens: 1820-1992. *Amer. Econ. Rev.* 92:727–744.
Castillo, E., Hadi, A.S., Sarabia, J.M. (1998). A method for estimating Lorenz curves. *Commun. Stat. Theory Methods* 27:2037–2063.
Chotikapanich, D. (1993). A comparison of alternative functional forms for the Lorenz curve. *Econ. Lett.* 41:129–138.
Chotikapanich, D., Valenzuela, R., Rao, D.S.P. (1997). Global and regional inequality in the distribution of income: Estimation with limited and incomplete data. *Empirical Econ.* 22:533–546.
Dagum, C. (1977). A new model a personal income distribution: Specification and estimation. *Economie Appliquée* 30:413–437.
Donaldson, D., Weymark, J.A. (1980). A single-parameter generalization of the Gini index of inequality. *J. Econ. Theory* 22:67–86.
Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica* 39:1037–1039.
Helene, O. (2010). Fitting Lorenz curves. *Econ. Lett.* 108:153–155.
Henle, J. M., Horton, N. J., Jakus, S. J. (2008). Modelling inequality with a single parameter. In: Chotikapanich, D., eds. *Modeling Income Distributions and Lorenz Curves: Vol. 5, Economic Studies in Inequality, Social Exclusion and Well-Being*. New York, NY: Springer.
Holm, J. (1993). Maximum entropy Lorenz curves. *J. Econometrics* 44:377–389.
Jain, S. (1975). *Size Distribution of Income: A Compilation of Data*. Washington, DC: World Bank.
Johnson, N.L., Kotz, S. Balakrishnan, N. (1994). *Continuous Univariate Distributions* (Vol. 1. Second Edition). New York, NY: Wiley.
Kakwani, N. (1980). On a class of poverty measures. *Econometrica* 48:437–446.
Kakwani, N.C., Podder, N. (1976). Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations. *Econometrica* 44:137–148.
Scarsini, M. (1989). A note on stochastic dominance and inequality measures. *J. Econ. Theory* 49:314–323.
Ogwang, T., Rao, U.L.G. (1996). A new functional form for approximating the Lorenz curve. *Econ. Lett.* 52:21–29.
Ogwang, T., Rao, U.L.G. (2000). Hybrid models of the Lorenz curve. *Econ. Lett.* 69:39–44.
Ortega, P., Martín, A., Fernández, A., Ladoux, M., García, A. (1991). A new functional form for estimating Lorenz curves. *Rev. Income Wealth* 37:447–452.

Pakes, A.G. (1986). On income distributions and their Lorenz curves. Technical Report. Department of Mathematics, University of Western Australia, Nedlands, W.A.

Rasche, R.H., Gaffney, J., Koo, A.Y.C., Obst, N. (1980). Functional forms for estimating the Lorenz curve. *Econometrica* 48:1061–1062.

Rhode, N. (2009). An alternative functional form for estimating the Lorenz curve. *Econ. Lett.* 105:61–63.

Ryu, H., Slottje, D.J. (1996). Two flexible functional forms for approximating the Lorenz curve. *J. Econometrics* 72:251–274.

Sarabia, J.M. (1997). A hierarchy of Lorenz curves based on the generalized Tukey’s lambda distribution. *Econometrics Rev.* 16:305–320.

Sarabia, J.M. (2008a). Parametric Lorenz curves: Models and applications. In: Chotikapanich, D., eds. *Modeling Income Distributions and Lorenz Curves*, Economic Studies in Inequality, Social Exclusion and Well-Being, (Vol. 5, pp. 67–190). New York, NY: Springer.

Sarabia, J.M. (2008b). A general definition of Leimkuhler curve. *J. Informetrics* 2:56–163.

Sarabia, J.M., Castillo, E., Pascual, M., Sarabia, M. (2005). Mixture Lorenz curves. *Econ. Lett.* 89:89–94.

Sarabia, J.M., Castillo, E., Slottje, D.J. (1999). An ordered family of Lorenz curves. *J. Econometrics* 91:43–60.

Sarabia, J.M., Castillo, E., Slottje, D.J. (2001). An exponential family of Lorenz curves. *Southern Econ. J.* 67:748–756.

Sarabia, J.M., Pascual, M. (2002). A class of Lorenz curves based on linear exponential loss functions. *Commun. Stat. Theory Methods* 31:925–942.

Sarabia, J.M., Prieto, F., Sarabia, M. (2010). Revisiting a functional form for the Lorenz curve. *Econ. Lett.* 107:249–252.

Sarabia, J.M., Sarabia, M. (2008). Explicit expressions for the Leimkuhler curve in parametric families. *Informat. Process. Manage.* 44:1808–1818.

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

Singh, S.K., Maddala, G.S. (1976). A function for the size distribution of incomes. *Econometrica* 44:963–970.

Slottje, D. (2010). Modeling income distributions and Lorenz curves. *J. Econ. Inequality* 8:525–526.

Villaseñor, J.A., (1989). Elliptical Lorenz curves. *J. Econometrics* 40:327–338.

Wolfson, M.C. (1994). When inequalities diverge. *Amer. Econ. Rev.*: 84:353–358.

Yitzhaki, S. (1983). On an extension of the Gini inequality index. *Int. Econ. Rev.* 24:617–628.