Working Papers Series No.35

THE BETA-LORENZ CURVE

-Rudolf Teekens

Publications Office
Institute of Social Studies
The Hague, The Netherlands

April 1987

THE BETA-LORENZ CURVE

by

Rudolf Teekens

Table of Contents

1. Introduction
2. Inequality Measures and the Lorenz Curve
3. The Incomplete Beta Function as Lorenz Curve
4. Symmetry and Skewness of the Beta-Lorenz Curve
5. The Implied Income Distribution
6. A Reproductive Property
7. Estimation

Appendices

The Hague, March 1987.

THE BETA-LORENZ CURVE

1. Introduction

The Lorenz curve depicts the relation between the proportion of income units (persons, households, etc.) with an income equal to or less than x and the proportional share in total income received by these units. In formal terms:

$$(1.1)\qquad q = L(p) \qquad 0 \leqq p \leqq 1, \quad 0 \leqq q \leqq 1$$

where p is the proportion of income units and q their share in total income. The quantity p is the value of the distribution function in point x and q is the value of the first moment distribution function. In case of a continuous income distribution p and q are defined as

$$(1.2)\qquad p = F(x) = \int_0^x f(t)\,dt$$

and

$$(1.3)\qquad q = F_1(x) = \frac{1}{\mu}\int_0^x t f(t)\,dt$$

with f(.) the density of the income distribution and μ its mean.

Kakwani (1980) derived the following properties of the function L(.):

$$(1.4)\qquad L(0) = 0 \quad \text{and} \quad L(1) = 1$$

$$(1.5)\qquad L'(p) = \frac{x}{\mu} \geqq 0$$

$$(1.6)\qquad L''(p) = \frac{1}{\mu f(x)} > 0$$

$$(1.7)\qquad L(p) \leqq p$$

Traditionally, income distributions have been characterised by theoretical densities, like the Pareto, the Lognormal, the Gamma, etc. These densities are also frequently used as interpolation devices for the estimation of inequality coefficients. Each of these and other theoretical models has shown in empirical applications certain shortcomings. In many cases a better empirical fit could only be obtained at the cost of increasing mathematical complexity.

Several authors have opted for an alternative approach, where the income distribution is characterised by the Lorenz curve instead of the density. This curve may equally well serve as an interpolation device for inequality coefficients, provided that they are mean-independent. Kakwani and Podder (1976) propose a specification of L(p), which does not have all the above mentioned properties, Rasche, Gaffney, Koo and Obst (1980) put forward a specification which cannot be estimated by linear methods and Gupta (1984) proposes a form which allows for very little flexibility since it depends on one parameter only. In general it can be stated that it appears rather difficult to specify a form of L(p), which (a) has the properties (1.4) to (1.7), and (b) is easy to estimate by a linear method.

This paper sets out to state that one may enlarge the class of manageable theoretical Lorenz curves by replacing the requirement (b), i.e. L(p) estimable by linear method, by the requirement that L'(p), the first derivative of L(p) with respect to p, is estimable by linear method. Following this line of reasoning the Incomplete Beta Function turns out to be a good candidate for the specification of a theoretical Lorenz curve (see section 3).

A possible disadvantage of the characterization of an income distribution by a theoretical Lorenz curve is the fact that, if L'(p) is not invertible, one cannot determine analytically the implied density function of the income distribution. This drawback is, however, not very serious, since this density can be generated numerically quite easily and the density is not needed for the determination of most mean-independent inequality measures, in particular the Generalized Entropy Family (see section 2), as well as the median and the mode.

2. Inequality Measures and the Lorenz Curve

As is shown by Pereirinha (1987) most inequality measures which are mean-independent can be written in the form

$$(2.1) \qquad E[H(\tfrac{x}{\mu})] = \int_{-\infty}^{\infty} H(\tfrac{x}{\mu})\, f(x)\, dx$$

with H(.) a convex real-valued function.[1)]

Making use of (1.5), (2.1) may be rewritten as

$$(2.2) \qquad E[H(\tfrac{x}{\mu})] = \int_{0}^{1} H[L'(p)]\, dp$$

hence an analytical expression of f(x) is not required.

More specifically, the Generalized Entropy Family may be expressed as:

$$I_c(x) = \frac{1}{c(c-1)} \left[\int_0^1 [L'(p)]^c\, dp - 1\right] \qquad c \neq 0,1$$

$$(2.3) \qquad I_0(x) = -\int_0^1 \ln L'(p)\, dp \qquad c = 0$$

$$I_1(x) = \int_0^1 L'(p) \ln L'(p)\, dp \qquad c = 1$$

hence, also for this class of inequality measures only the knowledge of L'(p) is required.

Finally, we consider the Gini-coefficient, which is not included in the above two classes of inequality measures. Kakwani shows that the Gini can be written as

$$(2.4) \qquad G = \frac{2}{\mu} \int_0^{\infty} x[F(x) - \tfrac{1}{2}]\, f(x)\, dx = \frac{2}{\mu} \int_0^{\infty} xF(x)f(x)\, dx - 1$$

Substitution of (1.6) and change of variable of integration yields:

(2.5) $$G = 2 \int_0^1 pL'(p)\, dp-1$$

hence knowledge of $L'(p)$ is sufficient for the determination of the Gini coefficient.

3. The Incomplete Beta Function as Lorenz Curve

The incomplete Beta function is defined as

(3.1) $$B(x;\ \alpha,\ \beta) = \frac{1}{B(\alpha,\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\, dt \qquad 0 \leqq x \leqq 1, \quad \alpha > 0, \quad \beta > 0$$

where B(.,.), the Beta function is defined as

(3.2) $$B(\alpha,\ \beta) = \int_0^1 t^{\alpha-1}\ (1-t)^{\beta-1}\ \alpha t \qquad \alpha > 0, \quad \beta > 0$$

The Beta function is related to the Gamma function according to the following formula:

(3.3) $$B(\alpha,\ \beta) = \frac{\Gamma(\alpha)\ \Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

The incomplete Beta function as specified in (3.1) is, given certain restrictions on its parameters α and β a suitable candidate to describe the Lorenz curve. We propose:

(3.4) $$L(p) = \frac{1}{B(\alpha,\ \beta)} \int_0^p t^{\alpha-1}\ (1-t)^{\beta-1}\ \alpha t \qquad \alpha \geqq 1, \quad 0 < \beta \leqq 1$$

This specification will be called the Beta Lorenz Curve. In order to verify whether this specification satisfies the properties (1.4) to (1.7), the first and the second derivative of L(p) with respect to P are determined:

$$(3.5) \qquad L'(p) = \frac{1}{B(\alpha,\ \beta)}\, p^{\alpha-1}\, (1-p)^{\beta-1}$$

and

$$(3.6) \qquad L''(p) = \frac{(2-\alpha-\beta)p+\alpha-1}{p(1-p)}\, L'(P)$$

It is easy to verify that L(0) = 0 and L(1) = 1, moreover $L'(p) \geqq 0$ for $0 \leqq p \leqq 1$ and $L''(p) > 0$ if $\beta < 1$. Furthermore, it can be seen that when $\alpha = \beta = 1$ the Beta-Lorenz Curve reduces to L(p) = p, i.e. it coincides with the egalitarian line.

This section is concluded with the derivation of the Gini coefficient and the Generalized Entropy Family of Inequality measures in terms of the parameters α and β.

Making use of the property

$$(3.7) \qquad {}_0\int^1 p^k\, L'(p)\, \alpha p = \frac{\Gamma(\alpha+k)\ \Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\alpha+\beta+k)}$$

For k = 1, the expression (2.5) for the Gini coefficient yields the following result:

$$(3.8) \qquad G = \frac{2\alpha}{\alpha+\beta} - 1 = \frac{\alpha-\beta}{\alpha+\beta}$$

Property (3.7) and equation (A.2) in the appendix also allow us to write the Generalised Entropy Family in terms of the parameters α and β:[3)]

$$I_c(x) = \frac{1}{c(c-1)} \left[\frac{B(c\alpha-c+1,\ c\beta-c+1)}{[B(\alpha,\ \beta)]^c} - 1\right] \qquad \text{for } c \neq 0,\ 1$$

(3.9) $$c < 1/1(1-\beta)$$

$$I_0(x) = \ln B(\alpha,\ \beta) + \alpha + \beta - 2 \qquad \text{for } c = 0$$

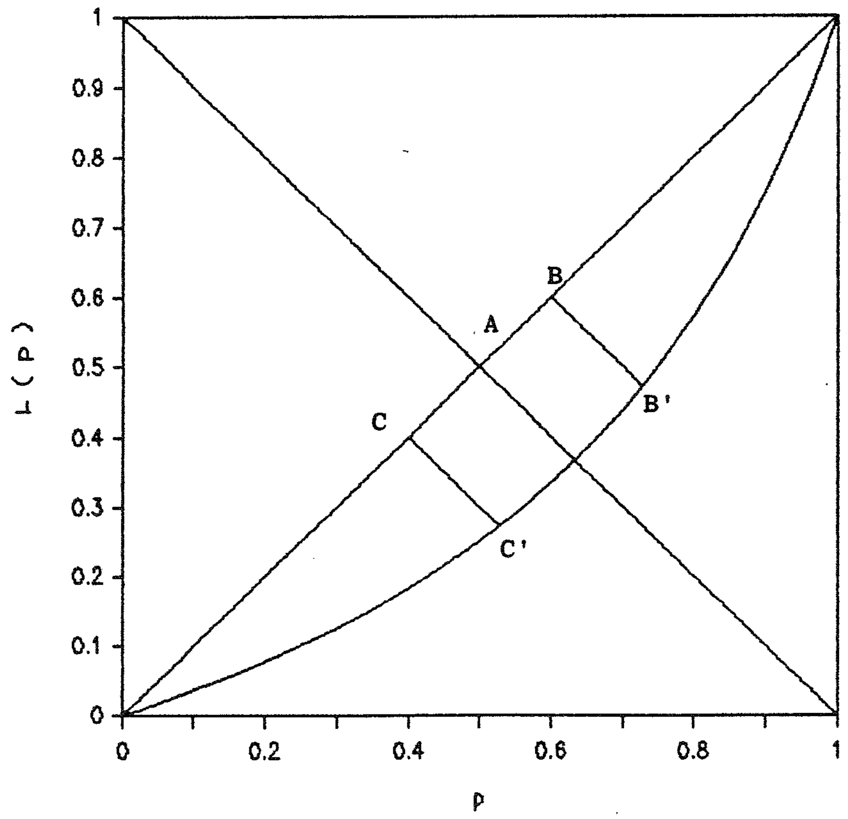
$$I_1(x) = (\alpha-1)[\Psi(\alpha)-\Psi(\alpha+\beta)] + (\beta-1)[\Psi(\beta)-\Psi(\alpha+\beta)] - \ln B(\alpha,\beta) \qquad \text{for } c = 1$$

where $\Psi(.)$ is the digamma function (see the appendix for its definition)

4. Symmetry and Skewness of the Beta-Lorenz Curve

A Lorenz curve is defined to be symmetric if for AC = AB the following equality holds true : BB' = CC' for all B and C. Kakawani (1980) proves that if the Lorenz curve is symmetric the following relation

Figure 1: Symmetry of the Lorenz Curve



holds:

(4.1) $L(p_\mu) = 1-p_\mu$

where p_μ is the value of p where $x = \mu$; this value follows from (1.5):

(4.2) $L'(p_\mu) = 1$

The Lorenz curve is skewed towards (0,0) if BB' > CC', or

(4.3) $L(p_\mu) > 1-p_\mu$

and skewed towards (1,1) if BB' < CC' or

(4.4) $L(p_\mu) < 1-p_\mu$

For the Beta-Lorenz curve the symmetry condition follows if (3.4) and (3.5) are substituted into (4.1) and (4.2) respectively:

(4.5)
$$\int_0^{p_\mu} t^{\alpha-1} (1-t)^{\beta-1} \, dt = B(\alpha, \beta)(1-p_\mu)$$
$$p_\mu^{\alpha-1} (1-p_\mu)^{\beta-1} = B(\alpha, \beta)$$

The system (4.5) constitutes two equations with three unknowns: α, β and p_μ, thus leaving one degree of freedom. If e.g. β is fixed, α and p_μ follow. Since (4.5) cannot be solved analytically, figure 2 depicts the values for α and β for which the Beta-Lorenz curve is symmetric[4].

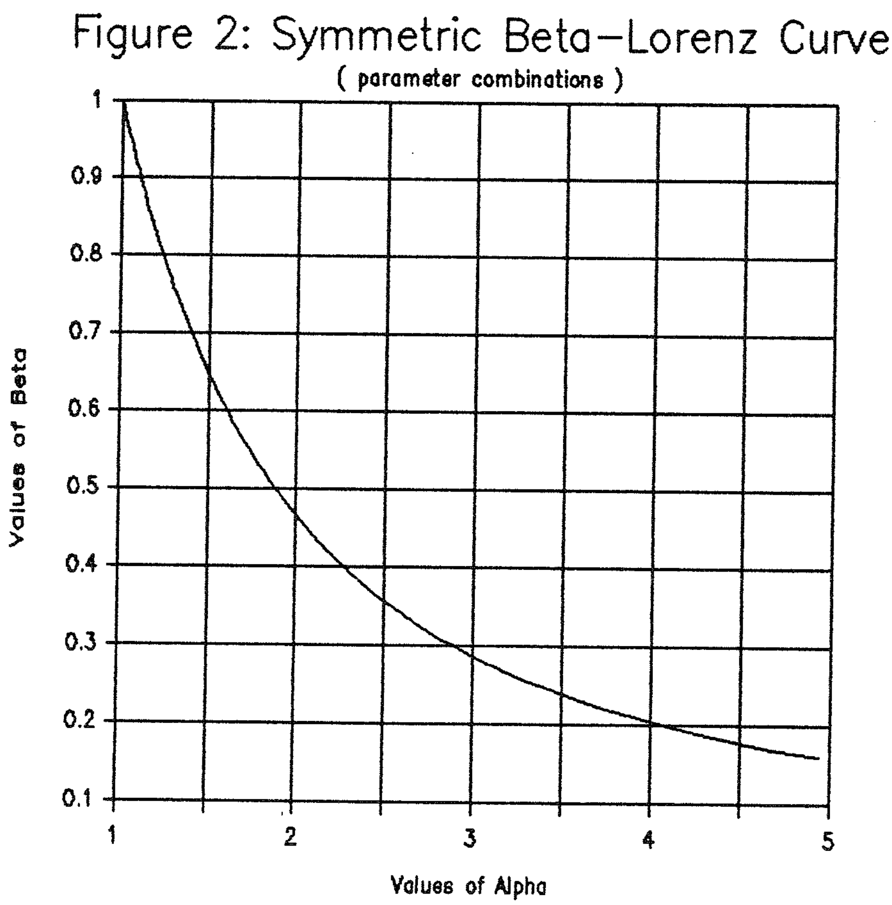


Figure 2: Symmetric Beta–Lorenz Curve
(parameter combinations)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Values of Beta
1
2
3
4
5
Values of Alpha

5. The Implied Density Function and its Characteristics

The density function of the income distribution impled by the Beta-Lorenz curve cannot be derived analytically[5)], but it can be generated easily, making use of equations (1.5), (1.6), (3.5) and (3.6):

$$(5.1) \qquad x = \mu L'(p) = \frac{\mu}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

and

$$(5.2) \qquad f(x) = \frac{}{\mu L''(p)} = \frac{p(1-p)}{(2-\alpha-\beta) p+\alpha-1} \cdot \frac{1}{x}$$

Although the density cannot be derived analytically, two location parameters of the distribution, the median and the mode can be expressed as functions of α, β and μ.

The median of the distribution, ξ, is equal to

$$(5.3) \qquad \xi = \frac{\mu}{B(\alpha, \beta)} 2^{2-\alpha-\beta} < \xi \qquad \text{for all } \alpha \neq 1, \ \beta \neq 1$$

The mode of the distribution can be determined by the condition

$$(5.4) \qquad \frac{d \ln f(x)}{dx} = \frac{d f(x)}{dx} / f(x) = \frac{1}{\mu} \frac{L'''(p)}{[L''(p)]^2} = 0$$

It can be demonstrated that the above condition is equivalent to

$$(5.5) \quad \frac{d \ln f(x)}{d\alpha} = - \frac{1}{\mu} \frac{(2-\alpha-\beta)(3-\alpha-\beta)p^2+2(\alpha-1)(3-\alpha-\beta)p+(\alpha-1)(\alpha-2)}{(2-\alpha-\beta)^2 p^2+2(\alpha-1)(2-\alpha-\beta)p+(\alpha-1)^2} \cdot \frac{1}{L'(p)} = 0$$

Analysis of condition (5.5) leads to the following conclusions:

A. If $\alpha < 2$ and $\alpha + \beta \neq 2$ the mode of the distribution corresponds to the value of $p = p_m$ with

(5.6) $$p_m = -\frac{\alpha-1}{2-\alpha-\beta} + \frac{\sqrt{(\alpha-1)(3-\alpha-\beta)(1-\beta)}}{(2-\alpha-\beta)(3-\alpha-\beta)} \qquad 0 < p_m < 1$$

The mode x_m is found by applying (5.1);

B. If $\alpha < 2$ and $\alpha + \beta = 2$, we have

(5.7) $$p_m = \frac{1}{2}(2-\alpha), \text{ with } x_m \text{ from (5.1)}$$

if both $\alpha = \beta = 1$ we have $p_m = 1$ and $x_m = \mu$

C. If $\alpha > 2$ the distribution does not have a mode, and $f(x)$ is a decreasing function.

This section is concluded with an overview of some special cases:

(i) If $\alpha = 1$, the Beta-Lorenz curve becomes

(5.8) $$L(p) = 1-(1-p)^{\beta}$$

and from (4.1) we have

$$\frac{x}{\mu} = \beta(1-p)^{\beta-1}$$

hence

$$\frac{x}{\mu} > \beta$$

and $$p = F(x) = 1-\left(\frac{x}{\beta\mu}\right)^{\frac{1}{\beta-1}}$$

so that

(5.9) $$f(x) = \frac{1}{\beta(1-\beta)\mu} \left(\frac{1}{\beta}\frac{x}{\mu}\right)^{-1-\frac{1}{1-\beta}} \quad \text{for } x > \beta\mu$$

which is the density of a Pareto distribution

(ii) If $\beta = 1$ we have

(5.10) $$L(p) = p^{\alpha} \text{ and } L'(p) = \alpha\, p^{\alpha-1}$$

hence $p = \left(\frac{x}{\alpha\mu}\right)^{\frac{1}{\alpha-1}}$ with $x < \alpha\mu$

and

(5.11) $$f(x) = \alpha \frac{1}{(\alpha-1)\mu} \left(\frac{x}{\alpha\mu}\right)^{\frac{1}{\alpha-1}-1} \quad \text{for } 0 < x < \alpha\mu$$

which is a rectangular density for $\alpha = 2$

Figure 3 shows some examples of densities for different values of α and β.

Figure 3: Beta-Lorenz Densities

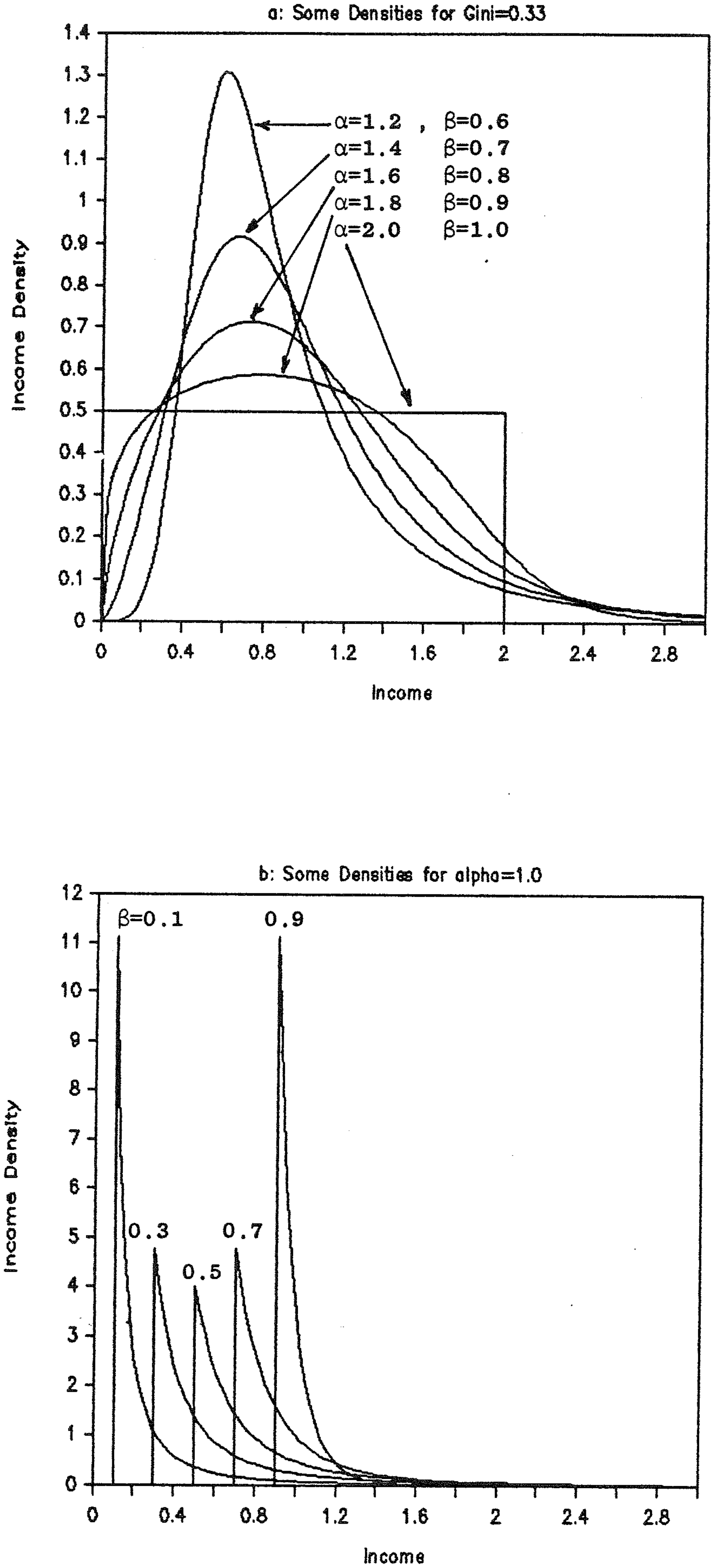
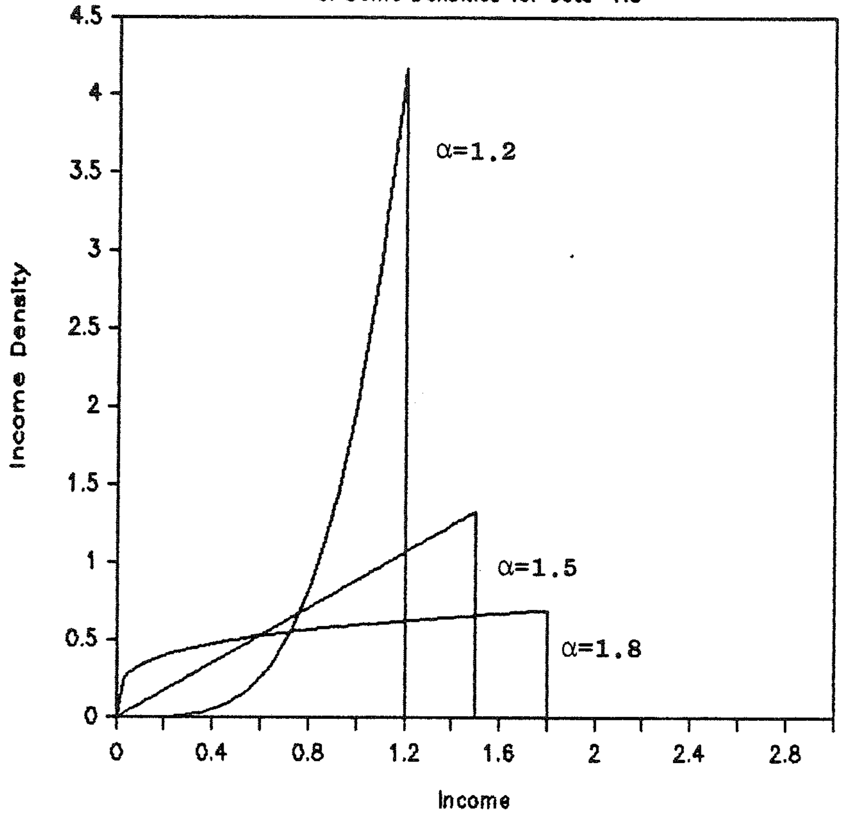


Figure 3 (continued)

c: Some Densities for beta=1.0



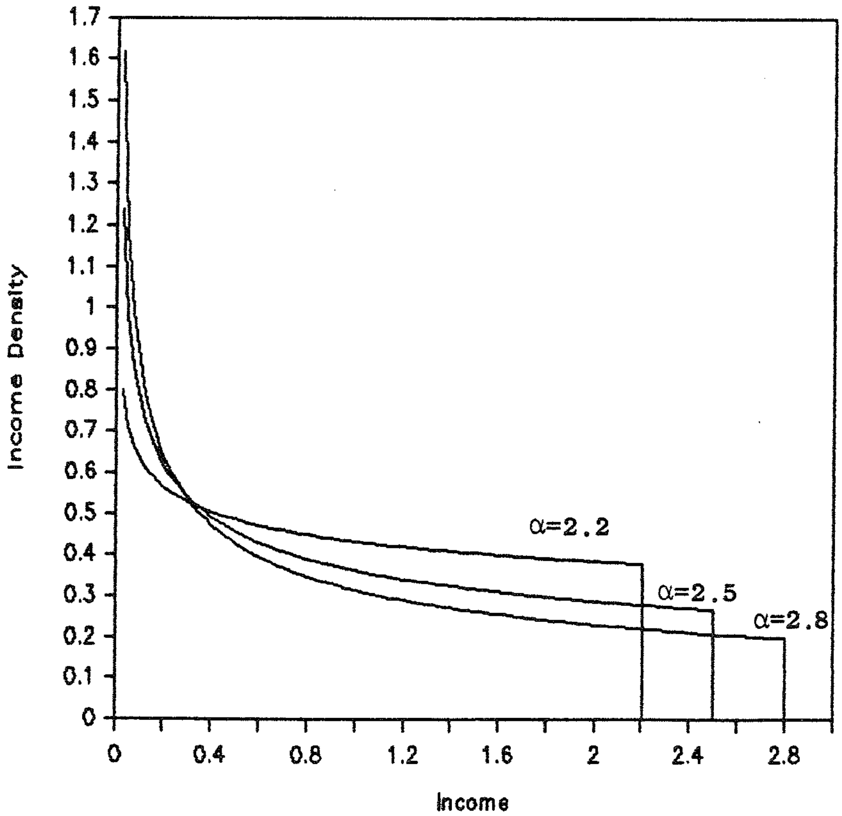
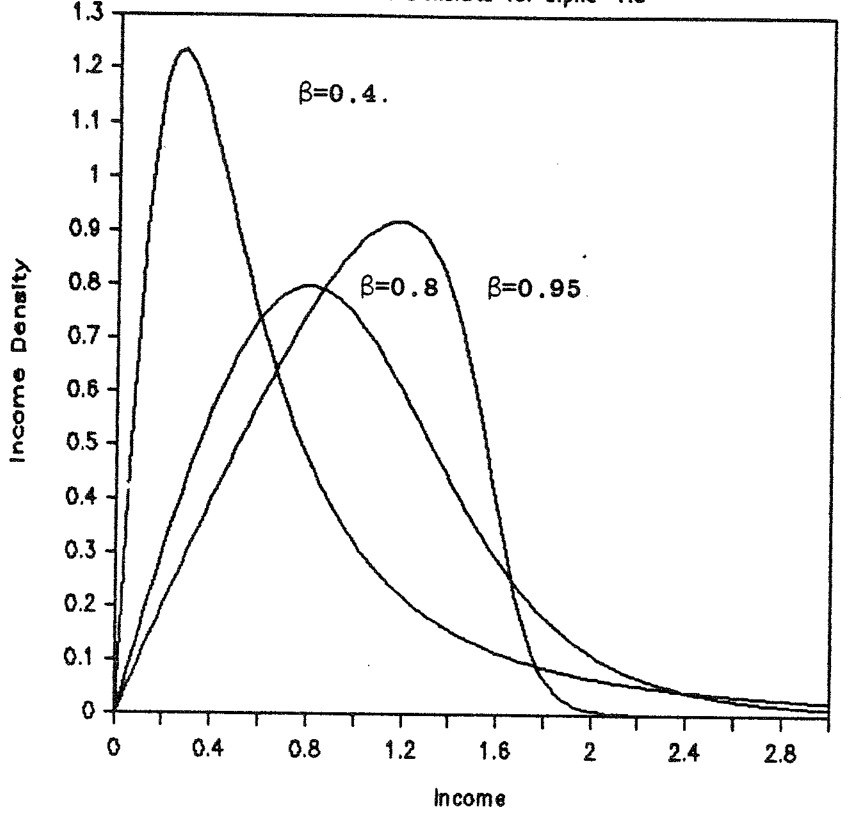
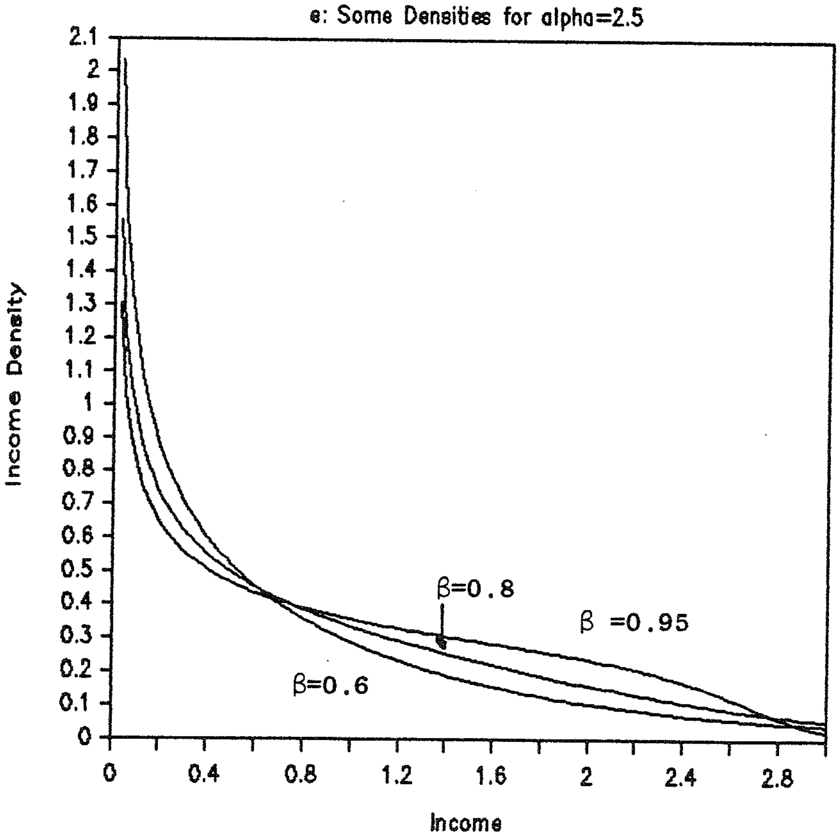


Figure 3 (continued)

d: Some Densities for alpha=1.5



e: Some Densities for alpha=2.5



6. A Reproductive Property

The distribution associated to the Beta-Lorenz curve has a similar reproductive property as the log-normal distribution:

A. If the variable X has a Beta-Lorenz curve with parameters α_x and β_x then the variable $Y = \gamma_o X^{\gamma_1}$ has also a Beta Lorenz curve with parameters $\alpha_y = \gamma_1 \alpha_x - \gamma_1 + 1$ and

$$\beta_y = \gamma_1 \beta_x - \gamma_1 + 1, \text{ provided that } \gamma_1 < 1 / (1-\beta)$$

<u>Proof</u>: Let $Y = g(X)$, then

$$p_y = P[Y \leqq y] = P[g(X) \leqq y] = P[X \leqq g^{-1}(y)]$$

$$= F[g^{-1}(y)], \text{ hence}$$

$$g^{-1}(y) = F^{-1}(p_y), \text{ or}$$

(6.1) $$y = g\,[F^{-1}(p_y)]$$

Moreover, we may easily verify the property

(6.2) $$F^{-1}(p) = \mu_x\, L'(p)$$

Substitution of (6.2) into (6.1) yields:

(6.3) $$y = g[\mu_x\, L'\,(p_y)]$$

since $y = g(x) = \gamma_o\, x^{\gamma_1}$, equation (6.3) yields

(6.4) $$y = \gamma_o \mu_x^{\gamma_1} \{L'(p_y)\}^{\gamma_1} = \frac{\gamma_o \mu_x^{\gamma_1}}{B^{\gamma_1}(\alpha_x, \beta_x)} p_y^{\gamma_1 \alpha_x - \gamma_1} (1-p_y)^{\gamma_1 \beta_x - \gamma_1}$$

where use has been made of equation (3.5), and since

(6.5) $$\mu_y = E[y] = \gamma_o E[\frac{x}{\mu_x}]^{\gamma_1} \cdot \mu_x^{\gamma_1} = \gamma_o \mu_x^{\gamma_1} E[L(p_x)]^{\gamma_1}$$

$$= \gamma_o \mu_x^{\gamma_1} \cdot \frac{B(\gamma_1 \alpha_x - \gamma_1 + 1, \gamma_1 \beta_x - \gamma_1 + 1)}{B^{\gamma_1}(\alpha_x, \beta_x)} \text{ provided that } \gamma_1 < \frac{1}{1-\beta}$$

combination of (6.4) and (6.5) gives:

(6.6) $$\frac{y}{\mu_y} = \frac{1}{B(\gamma_1 \alpha_x - \gamma_1 + 1, \gamma_1 \beta_x - \gamma_1 + 1)} p_y^{\gamma_1 \alpha_x - \gamma_1} (1-p_y)^{\gamma_1 \beta_x - \gamma_1}$$

This relation is easily identified as a special case of (3.5), which concludes the proof.

6. Estimation

The estimation of the Beta-Lorenz curve is most easily carried out through the estimation of the logarithm of $L'(p)$, its first derivation with respect to p:

(6.1) $\ln L'(p) = \ln x - \ln \mu = -\ln B(\alpha,\beta) + (\alpha-1) \ln p + (\beta-1) \ln (1-p) + u$

or

(6.2) $y = \gamma_o + \gamma_1 z_1 + \gamma_2 z_2 + u$

where $y = \ln x$ logarithm of income

$$\gamma_o = \ln B(\alpha,\beta) + \ln \mu$$
$$\gamma_1 = \alpha-1$$
$$\gamma_2 = \beta-1$$
$$z_1 = \ln p$$
$$z_2 = \ln (1-p)$$

If we assume a scalar covariance matrix for the disturbances, least squares estimates $\hat{\gamma}_o$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ lead to the following MVU estimates of α and β:

(6.3) $\hat{\alpha} = 1 + \hat{\gamma}_1$

(6.4) $\hat{\beta} = 1 + \hat{\gamma}_2$

and a consistent estimate of μ:

(6.5) $\hat{\mu} = \exp\{\hat{\gamma}_o - \ln B\,(\hat{\alpha},\ \hat{\beta})\}$

An alternative way of estimation of the parameters of the Beta-Lorenz curve can be found by making use of the second derivative L"(p) and its relation to L'(p), x and f(x). Making use of (1.5), (1.6) and (3.6) we obtain

(6.6) $xf(x) = \dfrac{p(1-p)}{(2-\alpha-\beta)\ p+\alpha-1}$

or

(6.7) $\dfrac{p(1-p)}{xf(x)} = \alpha-1 + (2-\alpha-\beta)\ p$

Hence the parameters of the Beta-Lorenz curve can also be estimated by

(6.8) $y = \delta_o + \delta_1\ p + v$

where $y = p(1-p)/(xf(x))$
$\delta_o = \alpha-1$
$\delta_1 = 2-\alpha-\beta$

If we assume alternatively that v has a scalar covariance matrix, the least squares estimates $\hat{\delta}_o$ and $\hat{\delta}_1$ lead to the following MVU estimates of α and β:

(6.9) $$\hat{\alpha} = 1 + \hat{\delta}_o$$

and

(6.10) $$\hat{\beta} = 1-\hat{\delta}_o-\hat{\delta}_1$$

This estimation method does not provide an estimate of μ and make use of the quantity $f(x)$ which is usually not readily available in income distribution data.

APPENDIX

Some Inequality Measures Expressed as Functions of the Parameters of the Beta-Lorenz Curve

The Coefficient of Variation:

$$(A.1)\quad CV = \left\{\frac{B(2\alpha-1,\ 2\beta-1)}{B^2(\alpha,\ \beta)} - 1\right\}^{\frac{1}{2}} \qquad \text{for } \beta > \frac{1}{2}$$

This expression is obtained easily, if it is realised that

$$(A.2)\quad \int_0^1 [L'(p)]^k\, dp = \frac{B(k\alpha-k+1,\ k\beta-k+1)}{B^k(\alpha,\ \beta)} \qquad \text{for } \beta > 1 - \frac{1}{k}$$

The Logvariance:

$$(A.3)\quad \sigma_L^2 = (\alpha + \beta - 2)^2 + \frac{\pi^2}{3}(\alpha-1)(1-\beta)$$

The general expression for the logvariance reads as:

$$\sigma_L^2 = \int_0^1 \{\ln L'(p)]^2\, dp - \{\int_0^1 \ln L'(p) dp\}^2$$

where for the Beta Lorenz curve

$$\ln L'(p) = (\alpha-1) \ln p + (\beta-1) \ln (1-p) - \ln B(\alpha,\ \beta)$$

and integration yields (A.3).

Theil has proposed two inequality measures, firstly

$$(A.4)\quad T_1 = -E[\ln(\frac{x}{\mu})]$$

which is a special case of the Generalised Entropy Family (c = 0) and

(A.5) $T_2 = E[\frac{X}{\mu} \ln (\frac{X}{\mu})]$

which is also a special case of the GEF (c = 1). The expressions for T_1 and T_2 are given in (3.9), where $\Psi(x)$ is a digamma function:

(A.6) $\Psi(x) = \frac{\alpha \ln \Gamma(x)}{\alpha x}$

REFERENCES

Gupta, M.R. (1984): "Functional Form for Estimating the Lorenz Curve", Econometrica, 52, pp. 1313-1314.

Kakwani, N.C. (1980): Income Inequality and Poverty, Oxford University Press.

Kakwani, N.C. and N. Podder (1976): "Efficient estimation of the Lorenz Curve and the Associated Inequality Measures from Grouped Observations", Econometrica, 44, pp. 137-148.

Pereirinha, J. (1987): Economic Inequalities in Portugal, Ph.D. thesis, forthcoming.

Rasche, R.H., Gaffney, J., Koo, A.Y.C. and N. Obst (1980): "Functional Forms for Estimating the Lorenz Curve", Econometrica, 48, pp. 1061-1062.

Shorrocks, A.F. (1980): "The Class of Additively Decomposable Inequality Measures", Econometrica, 48, pp. 613-625.

NOTES

1. See Pereirinha (1987) for details.

2. See Shorrocks (1980).

3. See the Appendix for the expressions of some particular members of this family.

4. An extensive table with the numerical values of α, β and the associated p_{μ} for which Beta-Lorenz is symmetric is available at request from the author.

5. Except for some special cases, which will be considered at the end of section 5.