[1] An income inequality measure based on the symmetric properties of both the income distribution and the Lorenz curve.

Mohamed Yusuf Hassan
Department of Statistics
United Arab Emirates University
Al AIN, UAE
Email: myusuf@uaeu.ac.ae

Abstract. In this paper, a new measure of income inequality is proposed. The measure is based on the symmetric properties of both the income distribution and the generalized Lorenz curve. De Casteljau's algorithm is employed to study the shape of the generalized Lorenz curve.The proposed measure is sensitive to changes in the extremes of the income distribution. Legendre transforms and Newton-Raphson method are used to develop the new measure of inequality.

Key words: Bezier curves, Generalized Lorenz curves, Bernstein polynomials, Legendre transform, Newton-Raphson method, Gini index, Skewness coefficient, De Casteljau's algorithm.

1 Introduction

The most widely used measure of inequality is the Gini index, which is twice the area between the Lorenz curve and the line of perfect equality. The Gini index summarizes an entire income distribution with a single value, it indicates the overall degree of income inequality but have little to say about the income differences between economic classes and the shape of the Lorenz curve. Using the Gini index alone, it is not clear whether increases in income inequality have been due to the rich getting richer, or the poor getting poorer since it is not particularly sensitive to the changes in the extremes, see Park(1998). So it may not be the best way for characterizing changes in income concentration. It is also possible for asymmetric cases that the Gini indices for two different intersecting Lorenz curves are equal (see illustrate example in 4.1). Another very popular approach combines the Gini with one or two measures of the Atkinson family, see (Atkinson 1970) and Gastwirth (1972). This practice, however, suffers from certain weaknesses. First, as demonstrated by Newbery (1970), the Gini Index and members of the Atkinson family have distinct theoretical foundations which make it difficult to evaluate their capacity as complementary measures of inequality. Moreover, in contrast to the Gini index the Atkinson measures cannot be expressed in a simple way by the Lorenz curve. Thus, it is difficult to see how one or a few of the Atkinson measures contribute to characterizing the Lorenz curve. Similiar criticism

[1]Mohamed Yusuf Hassan is an Associate Professor of Statistics, United Arab Emirates University.

can be raised against various alternative selections of inequality measures which appear in empirical applications. One way to address this problem is by using other measures sensitive to changes in the extremes such as the measure proposed in this paper.

In estimation problems, parameter estimates are reported with their standard errors to help better understand the variability of these estimates. The same concept could be applied in inequality measures like the Gini by reporting these measures together with other measures sensitive to changes in the extremes, these measures could be based on the shapes of the income distribution as well as the Lorenz curve.

One of the best candidates for explaining changes in the extremes is the coefficient of skewness. See for example, Dagum (1980) and (Encyclopedia of statistical science (1982) pages 156-161), a Lorenz curve is symmetric if $F(\mu) + L(\mu) = 1$; it is left asymmetric if $F(\mu) + L(\mu) > 1$ and right asymmetric if $F(\mu) + L(\mu) < 1$, where μ, F,and L are the mean income, the cumulative distribution of income and the Lorenz curve respectively. For the beta specification of the income distribution, Kakwani(1980) has proposed a coefficient of skewness for the Lorenz curve. The coefficient can be measured by the ratio of the two parameters, a and b. The Lorenz curve is symmetric if $a = b$, skewed toward the lower left if $a > b$ and skewed toward the upper right if $a < b$. Hassan and Abdalla (2004) proposed $\kappa = \frac{F(\mu)+L(\mu)}{\sqrt{2}}$. The Lorenz curve is symmetric if $\kappa = \frac{1}{\sqrt{2}}$, it is skewed to the right if $\kappa < \frac{1}{\sqrt{2}}$ and skewed to the left if $\kappa > \frac{1}{\sqrt{2}}$.

The main purpose of this paper is to propose a new measure of income inequality that is more sensitive to extremes. The measure improves Pietra's index (1915), and is based on the symmetric properties of both the income distribution and the generalized Lorenz curve. De Casteljau's algorithm, Legendre transforms, and Newton-Raphson method are used to develop the new measure of inequality.

2 Definition and some properties of the Lorenz Curve

definition 1. Given a distribution function $F(y)$ with support in the subset of the positive real numbers and with finite expectation μ, the Lorenz curve is defined as

$$L(p) = \mu^{-1} \int_0^p F^{-1}(y)dy, \quad 0 \le p \le 1, \tag{1}$$

where

$$F^{-1}(y) = sup\{z : F(z) \le y\}$$

The following theorem for the characterization of the Lorenz curve is in the paper by

Sarabia et al. (1999), the theorem is attributed to Gaffney and Anstis by Pakes (1981).

Theorem 1. *Assume that $L(p)$ is defined and continuous in the interval [0,1] with second derivative $L''(p)$. The function $L(p)$ is a Lorenz curve iff*

$$L(0) = 0, \quad L(1) = 1, \quad L'(0^+) \geq 0 \ , L''(p) \geq 0, for \ \ p \epsilon (0,1). \tag{2}$$

Some properties of the Lorenz curve for the class of continuously differentiable random variables.

Property 1: $L(p) \leq p$

Property 2: $F(\mu) - L(\mu) = sup_y[F(y) - L(y)]$, where $p = F(y)$

Property 3: $L'(p) = \frac{y}{\mu_y}$

Property 4: $L(F)$ is convex

Property 5: $M(\mu) = E(|Y - \mu|)/(2\mu) = F(\mu) - L(\mu)$ (relative mean deviation with respect to the mean)

Property 6: $M(m) = E(|Y - m|)/(2\mu) = F(m) - L(\frac{1}{2})$,(relative mean deviation with respect to the median), where m is the median income.

Property 7: Endpoint interpolation: The Lorenz curve interpolates $L(0)$ and $L(1)$.

Property 8: Linear precision property: If there is a perfect equality, the convex hull of the Lorenz curve collapses to the equalitarian line.

Property 9: Suppose the data are divided into $k + 1$ income intervals with $z_0, z_1, ..., z_{k+1}$ as the interval endpoints $0 \leq z_1 < z_2 < ... < z_{k+1} < \infty$, see Mehran et al. (1975), and Ogwang (2003). Let μ_i, p_i, and $L(p_i)$ denote the mean income in the i-th income interval (z_{i-1}, z_i), the overall mean income, the cumulative proportion of income receiving units with income less than z_i, and the corresponding cumulative proportion of income, respectively.

If μ is the overall mean income and $\beta_i = \frac{L(p_i)-L(p_{i-1})}{p_i-p_{i-1}}$ for $i = 1, 2, 3, ..., k+1$ are the slopes of the line segments joining the observed points on the Lorenz Curve as defined above, then the mean income in the i-th income interval as defined above is given by

$$\mu_i = \beta_i \mu. \tag{3}$$

3 Proposed measure of income inequality for Lorenz Curve (GLC)

Consider the generalized Lorenz curve proposed by Hassan and Abdalla (2004) which has the following parametric form

$$L(p) = (p^{\delta} - \beta_0 p^{\alpha}(1-p)^{\beta} e^{\beta_1 p})^{\gamma} \tag{4}$$

where

$$\gamma \geq 1, \quad 1 \leq \delta \leq \alpha, \quad 0 \leq \beta_1 < \beta \leq 1, \quad \beta_0 \leq \begin{cases} \frac{\delta(\delta-1)}{\alpha(\alpha-1)} & \text{if } \delta \neq 1; \\ 1 & \text{if } \delta = 1. \end{cases}$$

This parametric form has the following well-known Lorenz curves as special cases:

1. $L_1(p) = p^{\alpha}[1-(1-p)^{\beta} e^{\beta_1 p}]$ with $\delta = \alpha$, $\gamma = 1$, and $0 < \beta \leq 1$, is proposed by Abdalla and Hassan (2004),
2. $L_2(p) = p^{\alpha}[1-(1-p)^{\beta}]^{\gamma}$ if $\delta = \alpha$, $\beta_1 = 0$, $\gamma \geq 1$, and $0 < \beta \leq 1$ introduced by Sarabia et al (1999)
3. $L_3(p) = p^{\alpha}[1-(1-p)^{\beta}]$ putting $\gamma = 1$ in (2.) gives the Lorenz curve investigated by Ortega et al (1991).

The proposed measure of inequality is based on the shapes on both the income distribution and the Lorenz curve. In fact if both are symmetric then there is no income gap and the barycenter (center of mass) for egalitarian line is located on point $((F(\mu), L(\mu)) = (F(m), L(m)) = (\frac{1}{2}, \frac{1}{2})$ and $M(m) = M(\mu) = 0$. If either the income distribution or the Lorenz curve moves away from symmetry then the income gab increases, and $M(m) < M(\mu)$, the barycenter for the Lorenz curve also moves away from midpoint of the equalitarian line but stays within the convex hull of the Lorenz curve See Dagum (1980). This relationship between the symmetric properties and the income gap could shed light on how to measure income inequality, but we need to identify points that control the shape of the Lorenz curve and the income distribution. These control points could be based on the observed statistics or the parameters of the income distribution like the mean and median.

In this paper, we consider the four control points $P_0 = (0,0), P_1 = (F(m), L(m)), P_2 = (F(\mu), L(\mu))$ and $P_3 = (1,1)$ and find the barycenter of the Lorenz curve based on these control points. The new measure of income inequality is then defined using the barycenter and the skewness coefficient of the Lorenz curve. Given the control points from the Lorenz curve, we use the De Casteljau's algorithm which is well suited for the shape analysis of the

Lorenz curve to locate the barycenter of the Lorenz curve. De Casteljau's algorithm, named after its inventer Paul de Casteljau, is a recursive method to evaluate Bezier curves for shape analysis in geometric modelling.

Given control points $P_0, P_1, ..., P_n$ and $t\epsilon[0,1]$, the algorithm can be expressed by the recurrence relation

$$P_{0,j}(t) = P_j$$
$$P_{i,j}(t) = (1-t)P_{i-1,j}(t) + tP_{i-1,j+1}(t)$$
$$i = 1, 2, ..., n,$$
$$j = 0, 1, ..., n-i.$$

Given the four control points $P_0 = (0,0), P_1 = (F(m), L(m)), P_2 = (F(\mu), L(\mu))$ and $P_3 = (1,1)$ and using De Casteljau's algorithm with $t = \frac{1}{2}$, we have

$$P_{1,j}(\frac{1}{2}) = \frac{1}{2}P_j + \frac{1}{2}P_{j+1}$$
$$j = 0, 1, 2, 3.$$

which are the barycenters of the lines connecting the points P_j and P_{j+1}. For example

$$P_{1,0}(\frac{1}{2}) = \frac{1}{2}P_0 + \frac{1}{2}P_1$$

is the midpoint between the control points P_0 and P_1.

Applying De Casteljau's algorithm recursively, the barycenter of the Lorenz curve is calculated and

$$Barycenter = (\frac{1}{2})^3(P_0 + 3P_1 + 3P_2 + P_3)$$
$$= \frac{1}{8}((0,0) + 3(F(m), L(m)) + 3(F(\mu), L(\mu)) + (1,1))$$

So the coordinates of the barycenter are

$$F_b = \frac{1}{8}(3F(m) + 3F(\mu) + 1)$$

and

$$L_b = \frac{1}{8}(3L(m) + 3L(\mu) + 1).$$

One could find the parametric equations F_b and L_b by using a Bezier curve with Bernstein basis evaluated at $t = \frac{1}{2}$. If the income distribution and the Lorenz curve are both symmetric then $F_b = \frac{1}{2}$ and $L_b = \frac{1}{2}$. clearly

$$F_b - L_b = \frac{3}{8}(M(m) + M(\mu))$$

which is the income gap at the barycenter of the Lorenz curve.

The Income inequality measure ψ is defined as

$$\psi = \kappa(F_b - L_b) = \frac{3}{8\sqrt{2}}[(M(m) + M(\mu))(F(\mu) + L(\mu))] \qquad (5)$$

where $\kappa = \frac{F(\mu)+L(\mu)}{\sqrt{2}}$ is the skewness coefficient of the Lorenz curve.

ψ is a function of the relative mean deviation about the mean, relative mean deviation about the median and the skewness coefficient of the Lorenz curve. Thus, it is not only a measure of income gap but it is also sensitive to the changes in the extremes.

Theorem 2. *If ψ is the income inequality measure stated above then*

$$0 \leq \psi \leq \frac{3\sqrt{2}}{8}$$

. Proof:

To see that zero is the lower bound of ψ is trivial. To show ψ is less than or equal to $\frac{3\sqrt{2}}{8}$, we have

$$\begin{aligned}\psi &= \frac{3}{8\sqrt{2}}[(M(m) + M(\mu))(F(\mu) + L(\mu))] \leq \frac{3}{8\sqrt{2}}(2M(\mu))(F(\mu) + L(\mu)) \\ &= \frac{3\sqrt{2}}{8}(F(\mu) - L(\mu))(F(\mu) + L(\mu))) \leq \frac{3\sqrt{2}}{8}(F(\mu))^2 \leq \frac{3\sqrt{2}}{8}.\end{aligned}$$

definition 3. Let β be the slope of a line passing through the origin. If f is any convex function, the Legendre transform, which is also a convex function is defined as

$$(Lf)(\beta) = sup_x(\beta x - f(x))$$

Given a Lorenz curve $L(p)$, the maximal distance between the Lorenz curve and the equalitarian line, see (Figure 1), is given by

$$(Lf)(1) = sup_p(p - L(p)).$$

The distance is maximal at the point $(F(\mu), L(\mu))$, which is unique by the monotonicity of $L^{'}(p)$, see (Encyclopedia of statistical science (1982) pages 156-161). Newton-Raphson method is used to find

$$(Lf)(1)$$

.

Income Data sets from Norway and United Kingdom are fitted to the Generalized Lorenz curve (4) by the method of the least squares. A Legendre transform is used to find the the control point $(F(\mu), L(\mu))$. The computation of $(F(m), L(m))$ is similiar

4 Illustrative Examples

4.1 Intersecting Lorenz Curves With Equal Gini indices

In this example two intersecting Lorenz curves with equal Gini coefficients are investigated using income data from the UK and Norway in 1967 and 1963 respectively. The generalized

Lorenz parametric form in (4) with $\gamma = 1$ is fitted to each data set and the parameters are estimated. Gini, skewness coefficient and ψ are calculated from the fitted model. Despite the fact that the Gini indices of the two countries are the same, both 0.36, it is found that the ψ of the fitted GLC for UK is 0.13 and the one for Norway is 0.12, which favors Norway than UK. The location of the overall mean income of the UK is higher than that of Norway. In the UK The income group with the average income was in the lowest 64th percentile of the income groups with cumulative income proportion of 39%, whereas that of Norway was in the lowest 58th percentile with cumulative income proportion of 39%. Figure 2 displays graphs of the fitted Lorenz curves. Estimated parameters, the Gini indices and ψ are given in Table 1. Total income shares are displayed in Table 2.

Estimates	δ	α	β_0	β_1	β	Gini	skewness	ψ
UK	1.64	2.29	0.40	0.0	0.41	0.36	0.18	0.13
Norway	1.88	3.91	0.45	0.0	0.63	0.36	0.19	0.12

Table 1: UK and Norway results based on model (4).

Income Category	Share of Total Income (Norway)	Share of Total Income (UK)
Lowest 20%	5%	6%
Lowest 40%	17%	18%
Lowest 60%	35%	35%
Lowest 80%	59%	57%
Lowest 90%	75%	72%
Lowest 99%	96%	93%

Table 2: UK and Norway Income Distributions

4.2 Intersecting Lorenz Curves With Unequal Gini Coefficients

Intersecting Lorenz curves with unequal Gini coefficients are presented in this example. Two generalized Lorenz curves were fitted to the data of each country, the UAE and Urugway. Parameter estimates and the Gini, skewness coefficient and ψ are shown in Table 3 and the graphs of the fitted models are displayed in Figure 3. The Gini and (ψ) for the UAE, and Urugway are 0.45 (.175) and 0.50 (.177) respectively. The ψ estimate of the fitted GLC for UAE is 0.175 and the one for Urugway is 0.177, which favors UAE than Urugway. These results confirm those of the Gini index and income distribution. Reporting Gini coefficients alone would indicate that, inequality in Urugway is greater than that in UAE. The Gini coefficients did not explain which income groups were contributing the increase in income inequality. Table 4 reports in detail income shares of the different income groups in each country. Results in the table show that, the top income bracket in UAE and the lower income bracket in Urugway contributed much of the disparity in the incomes. The lowest 20% income group in Urugway has only 3% of the total income compared to 5.4% in UAE. On the other hand, The highest one percent income group in UAE has 10% of the total income whereas it is 6% in Urugway. So analysis of inequality based solely on the Gini coefficients would not provide complete description of the income distribution. Reporting the Gini and ψ together would help to draw conclusions about inequalities both within and across countries. It is interesting to see that, in Table 1 and Table 3, the estimates of β_1 for UAE, Norway, UK, and Urugway are 0.48, 0.0, 0.0, and 0.0 respectively. The income distribution in the UAE is bimodal with one mode for the nationals and another for the expatriates. The non-zero β_1 Coefficient for the UAE seem to have captured this bimodality.

Estimates	δ	α	β_0	β_1	β	Gini	Skewness	ψ
UAE	1.56	1.77	0.50	0.48	0.49	0.45	0.76	0.175
Urugway	1.93	2.18	0.52	0.0	0.66	0.50	0.68	0.177

Table 3: UAE and Urugway results based on model (4)

Income Category	Share of Total Income (UAE)	Share of Total Income (Urugway)
Lowest 20%	5.4%	3%
Lowest 40%	15%	11%
Lowest 60%	28%	24%
Lowest 80%	48%	47%
Lowest 86%	56.35%	56.35%
Lowest 90%	64%	66%
Lowest 99%	90%	94%

Table 4: UAE and Urugway Income Distributions

5 Conclusion

In this paper we proposed an income inequality measure based on the symmetric properties of both the income distribution and the Lorenz curve. The measure is sensitive to the changes in the extremes, and helps to see differences between economic classes and the shape of the Lorenz curve. The new measure of income inequality is developed by using the barycenter and the skewness coefficient of the Lorenz curve. De Casteljau's algorithm is used to find the barycenter of the curve. Two illustrative examples were presented using a generalized Lorenz curve. Legendre transform and the Newton-Raphson method were used to get the maximum income disparity.

References

[1] Abdalla, I., Hassan, M. Y., (2004) "Fitting income distribution using a new parametric Lorenz curve," *Pakistan Journal of Statistics,volume 20,2, 277-286.*

[2] ... (1970) "On the measurement of inequality," *J. Econ. Theory 2, 244-263.*

[3] Dagum, C. (1980) "The generation and distribution of income, the Lorenz curve and the Gini ratio.," *Economie Appliquee, 33, 327-367.*

[4] Gastwirth, J. L. (1972) "Estimation of the Lorenz curve and Gini index," *Review of Economics and Statistics, 54, 306-316.*

[5] Kakwani, N. C. (1980) "On a Class of Poverty Measures," *Econometrica, 48, 437-446.*

[6] Kakwani, N. C. (1980) *Income inequality and Poverty Measures*, London, Oxford University press.

[7] Hassan, M. Y., Abdalla, I.,(2004) "Generalized Lorenz Curve and its Coefficient of Skewness" Unpublished manuscript, UAE University, Al Ain.

[8] Mehran, Farhad (1975), "Bounds on the Gini index based on observed points on the Lorenz curve," *Journal of American Statistical Association, 70, 64-66.*

[9] Newbery, D. (1970) "A theorem on the measurement of inequality," *Journal of Economic Theory. 2, 264-266.*

[10] Ogwang, Tomson (2003), "Bounds of the Gini index using sparse information on mean incomes," *Review of Income and Wealth, 49, 415-423.*

[11] Ortega, P., Fernandez, M. A., Lodoux, M. and Garcia, A. (1991), "Functional Form for Estimating Lorenz Curves," *Review of Income and Wealth, 37, 447-452.*

[12] Pakes, A. G. (1981), "On income distribution and their Lorenz curve.," Technical report, Department of Mathematics, University of Westren Australia, Nedlands, WA.

[13] Park, K. H. (1998) "Trend Income Distribution and the Middle Class ," *Proceedings of the American Society of Business and Behavioral Sciences, Vol. 4, pp. 73-79, 1998.*

[14] Pietra, G. (1915) "Delle relazioni tra gli indici di variabilita. Nota I. ," *Atti Regio Istituto Veneto, Tomo 74, Park II, Sciences, pp. 775-792.*

[15] Rasche, R. H.,Gaffney, J., Koo, A. and Obst, N. (1980), "Functional Forms for Estimating Lorenz Curve," *Econometrica, 48, 1061-1062.*

[16] Rolf, A. (2000) "Characterization of Lorenz curves and income distributions," *Social Choice and Welfare, 17, 639-653.*

[17] Sarabia, J. M., Castillo, E., and Slottje, D. J. (1999) "An Ordered Family of Lorenz Curves," *Journal Econometrics, 91, 43-60.*

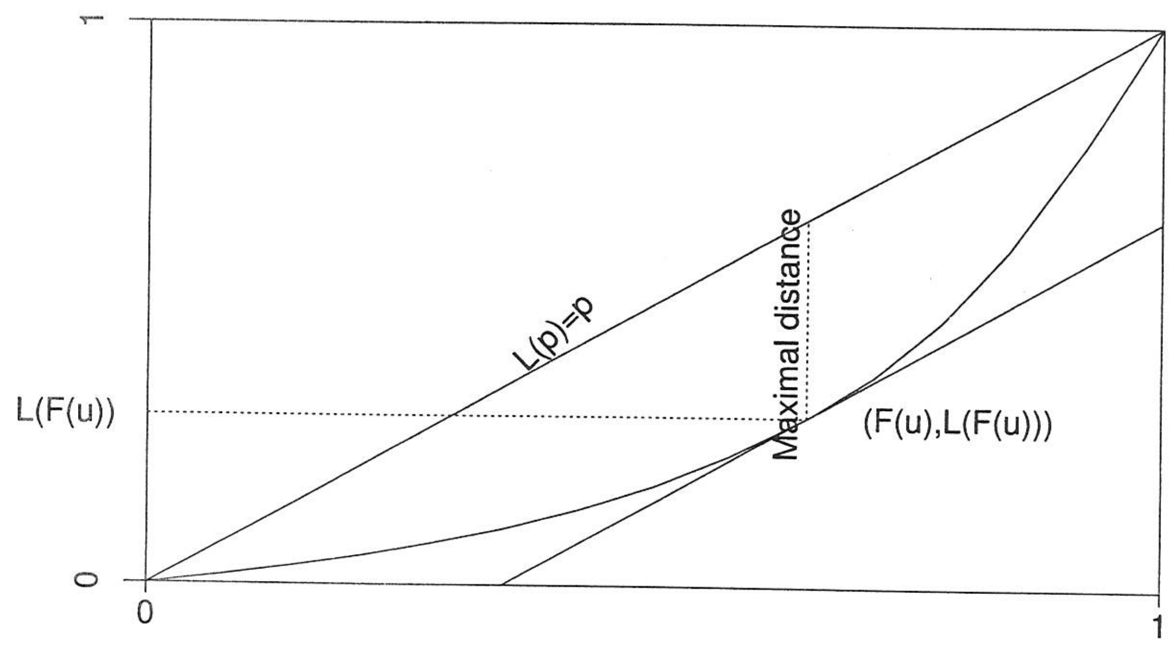

Figure 1: Maximum disparity

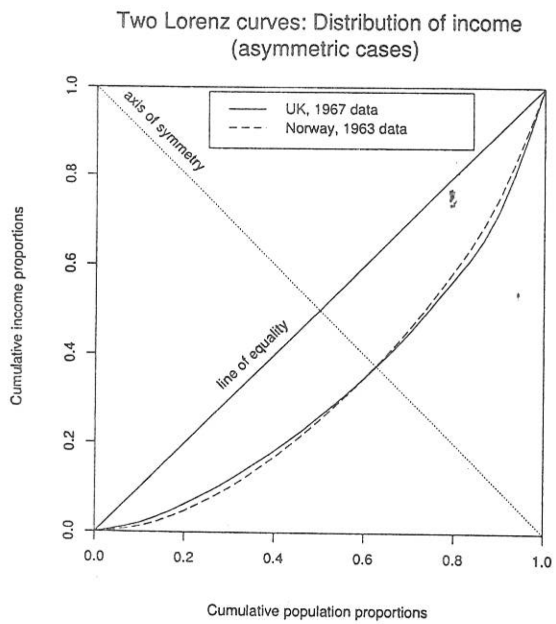

Figure 2: Gini coefficients of the two intersecting curves are both the same 0.36 but for the UK ψ is 0.13, where as Norway has a ψ of 0.12

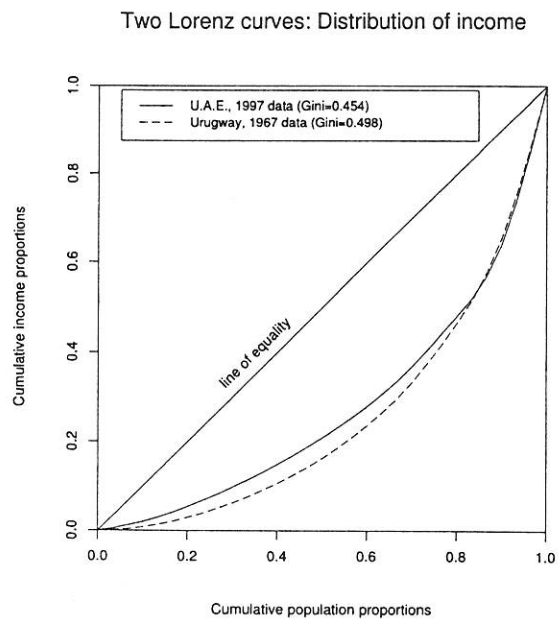

Figure 3: The Gini indices of the UAE and Urugway are 0.45 and 0.50 respectively. ψ for the UAE is 0.175, for Urugway it is 0.177.