Measuring the Asymmetry of the Lorenz Curve (*)

Agostino Tarsitano
Università degli studi della Calabria
Dipartimento di Economia e Statistica
87030 Arcavacata di Rende (Cs)
agotar@unical.it

Summary
The Gini concentration ratio as well as any other scalar measure of inequality is clearly inadequate as a description of the Lorenz curve. In this paper, a "dual" description, based on a measure of concentration and one of asymmetry (the Zanardi index) is advocated. Two applications are presented to show that the Gini and Zanardi indices capture the most significant aspects of a Lorenz curve.

.

keywords: income indequality, concentration indices,Zanardi index, economic development, principal component analysis.

(*) Ricerche Economiche, Vol. XLII,n.3, 1988 pp. 507-319

1. Introduction

The Gini concentration ratio is a scalar measures of average deviations in the income distribution; thus it gives only a general idea of income inequality. In fact, the same Gini ratio might be associated with very different distributions(1)

Many of the authors who have used the concentration ratio noted such a drawback and several attempts have been made to supplement the index with additional measures: Hagerbaumer (1977), Dagum (1980), Koo *et al.* (1981), Cortes and Rubalcava (1983).

As Chipman (1985) puts it "For too long it has been the bane of social science to seek for a single number that can measure a complicated and emotionally charged concept. The time has come to move on to a multivariate approach" and further *"I would imagine that if Gini were alive today, he would be pioneer in the development of such multivariate approach."*

This article is based on the idea that fundamental complementary information on a size distribution of incomes can be obtained by studying the asymmetry of the Lorenz curve.

After a general discussion on the symmetry of the curve (2nd section), the Zanardi index of asymmetry is presented (3rd section). Finally, in the fourth section, I report two applications which show that the Gini and Zanardi indices capture the most significant aspects of the Lorenz curve.

2. The asymmetry of the Lorenz curve

The asymmetry of the Lorenz curve has been studied by several Italian statisticians: Gini (1932), Panizzon (1955), Giurovich (1959), Zanardi (1964) and (1965). After being in isolation for over fifteen years it has recently been considered to deal with the problem of comparing income distributions whose Lorenz curves intersect: Kakwani (1980), Gagliani and Tarsitano (1987).

In this section I discuss the asymmetry of the Lorenz curve. To do this, the curve is expressed in terms of the Gini's coordinate system(2) instead of the usual, but less direct *(p,q)* plane(3).

The Gini's coordinate are given by the transformation

$$\begin{bmatrix} \pi \\ \theta \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} p+q \\ p-q \end{bmatrix} \tag{1}$$

Clearly, (1) is a one-to-one mapping(4) of (p,q) onto (π,θ).

The equation of the Lorenz curve in terms of π and θ is

$$\theta = L(\pi) \qquad for \;\; 0 \le \pi \le 2 \tag{2}$$

Since (2) is concave(5) the concept of asymmetry of a Lorenz curves becomes very similar to the concept of skewness of a unimodal density function. The Lorenz curve $\theta=L(\pi)$ is symmetric around the pole $\pi=1$ if

$$L(1+\pi) = L(1-\pi) \text{ for } 0 \le \pi \le 1 \tag{3}$$

this entails that a symmetric Lorenz curve has the same ordinate for any two abscissae equidistant from the median point. In addition, if $\hat{\theta}$ denotes the maximum height[(6)] of the Lorenz curve and $\hat{\pi}$ the abscissa at which it is obtained, $\hat{\pi} = 1$.

In figure a, curve A is symmetric; consequently the rising portion of the curve $\pi \le \hat{\pi}$ has the same slope (regardless of the sign) of the declining portion $\pi \ge \hat{\pi}$.

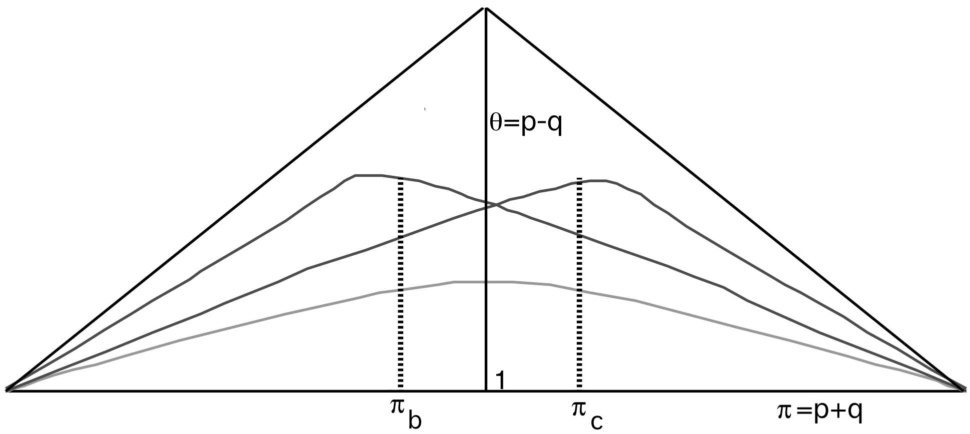

Figure a. -Types of asymmetry of the Lorenz curve

Symmetric curves result from income distributions in which there is a perfect compensation between statistical quantities (e.g. concentration) in the lower and in the upper part of the distribution[(7)]. Curve B is positively asymmetric (or asymmetric to the right) because $\hat{\pi}_b < 1$, and because its rising portion has a steeper slope than its declining portion. Such a curve represents a distribution characterized by a large fraction of relatively poor IRU's whereas the rich IRU's share almost equally. Conversely, curve C is negatively asymmetric (or asymmetric to the left) because $\hat{\pi}_c > 1$, and because its declining portion has the steepest slope. This type of curve is the expression of a distribution in which a majority of relatively poor IRU's shares a given quota of total income while a few rich IRU's possess the remaining income.

Definitions given above[(8)] do not quantify the asymmetry of the Lorenz curve, rather they constitute a framework in which such measurement can be realized.

3. The Zanardi index of asymmetry of a Lorenz curve

The purpose of this section is to present an index of asymmetry of the Lorenz curve that complements the traditional indices of inequality.

In Figure b, point $W \equiv (\pi_w, \theta_w)$ is the intersection point of $L(\pi)$ with the symmetry pole. Let y_w indicate the income level which generates (p_w, q_w) and define "poor" a unit whose income is less than y_w[(9)] and "rich" a unit whose income is at least y_w.

On the basis of yw the income distribution is divided into two subdistributions: the poorest p_w% which determines the portion of the curve to the left of the median *M*, and the richest $(1-p_w)$% which determines the portion to the right of *M*.

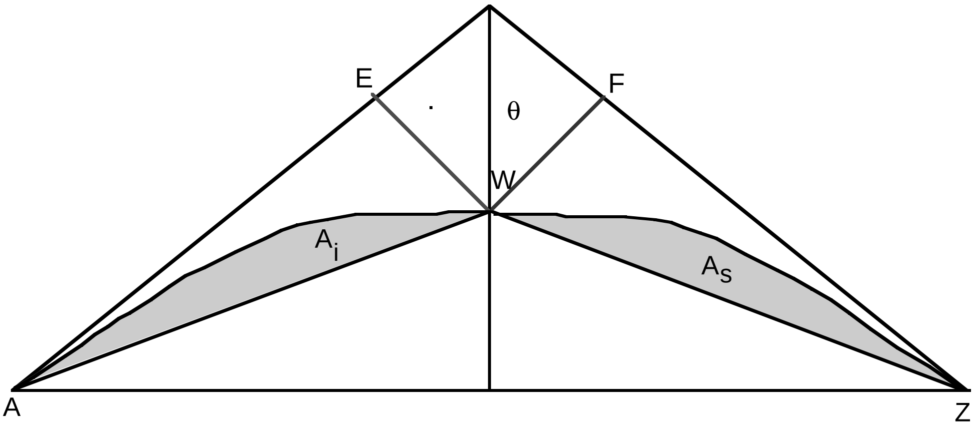

Figure b. -Definition of the Zanardi index

Triangles AWE and WFZ have the same area $H=p_w q_w/2$ and form *(p,q)* planes for the poor and for the rich income-receiving units (IRU's) respectively. The Gini ratios for these subdistributions are $G_i=A_i/H$ and $G_s=A_s/H$. According to Zanardi (1965), the asymmetry of the Lorenz curve reflects unbalances between inequality among poor and inequality among rich IRU's

To quantify the asymmetry of the Lorenz curve, Zanardi introduced the following measure(10):

$$Z = 2H\left[\frac{G_s - G_i}{G}\right] = 2\frac{(A_s - A_i)}{G} = 4\frac{1 - A(1)}{A(2)} \qquad (4)$$

where G=A(2) is the overall Gini ratio and A(π) represents the area below the curve $L(\pi)$ in the interval $[0,\pi]$.

Index (4) satisfies certain desirable properties:

1) *Z* does not change when all incomes change proportionally(11).
2) $Z=0$ if the Lorenz curve is symmetric.
3) *Z* lies between -1 and 1.
4) If $L^*(\pi)$ is selfsymmetric(12) to $L^+(\pi)$, then $Z^* = -Z^+$.

Verification of the first three properties is straightforward(13). The last property, which was not mentioned by Zanardi, becomes apparent by observing that L^* is the mirror image of L^+ (with respect to the pole of symmetry $\pi=1$).

It is important to note that *Z* is a normalized index (with respect of G) and, consequently, it can be used to compare the asymmetry of Lorenz curves having a different Gini ratio.

For a better understanding of the bounds on *Z* let us consider Figure c.

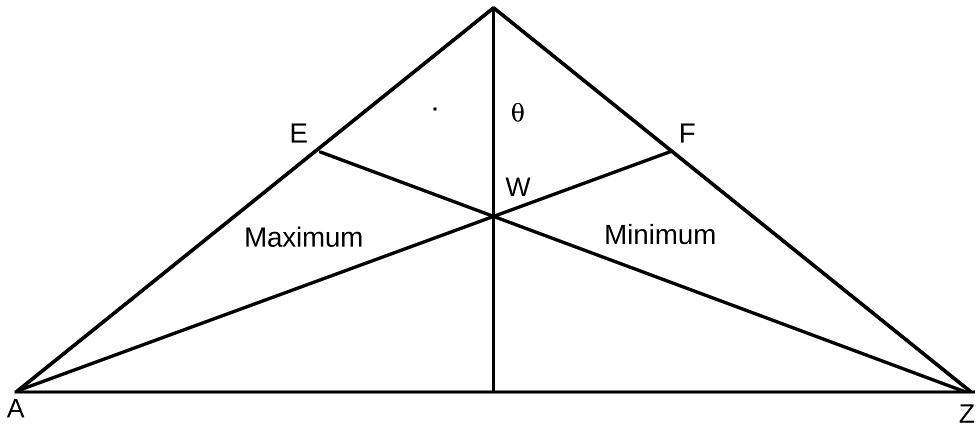

Figure c. - Extreme Lorenz curves

For a fixed value *G* of the Gini ratio (segments AE) curve AEZ represents the distribution whereby the first *G%* receives zero income and each unit in the remaining fractile receives the same amount L_s. Hagerbaumer (1977) called AEZ the "minimum" Lorenz curve. The curve is obviously asymmetric to the left $(A_s=0)$; as *G* tends to zero, that is, as the fractile reaching the upper income level L_s increases, the Zanardi index converges to the lower limit -1.

Similarly, curve AFZ (the "maximum" Lorenz curve) represents the distribution whereby all but the richest unit receive the same amount Li of income. Since Ai=0, the curve is asymmetric to the right. As G (segment FZ) tends to zero, that is, as the "survival" income Li increases, the Zanardi index converges to the upper limit 1.

Computation of the Zanardi index for smooth curves is relatively simple. For the curve[14] introduced by Kakwani and Podder (1976)

$$L(\pi) = R\pi^{a}\left(2-\pi^{b}\right), \qquad 0 < a \le 1; \quad 0 < b \le 1 \tag{5}$$

where *R* denotes the relative mean deviation of incomes[15], *Z* is given by

$$Z(a,b) = 2\left[1 - 2IB(0.5, a+1, b+1)\right] \tag{6}$$

where *IB(x;r,s)* is the incomplete Beta function ratio. It is easy[16] to show that if *a=b*, i.e. if the curve is symmetric, then *Z(a,b)=0*. Moreover, *Z(a,b)= -Z(b,a)*.

Kakwani (1980) proposed [(a/b)-1] as a measure of asymmetry for the Lorenz curve. Such a measure has an infinite range and, therefore, offers no benchmark to which its values can be related[17].

To estimate the Zanardi index for grouped data[18] the construction of the empirical Lorenz curve is required. To this end, a common method is the use of a linear function R_i joining (π_{i-1},θ_{i-1}) and (π_i,θ_i) for $i=1,2,...,k$; with $\pi_0 = \theta_0 = 0$.

If *k* is the number of groups, the Zanardi index is given by

$$Z = 2\left\{\frac{1}{G}\left[\left(\theta_m + \theta_{m-1}\right)\left(1-\pi_{m-1}\right) + \sum_{i=1}^{m-1}\left(\theta_i + \theta_{i-1}\right)\left(\pi_i - \pi_{i-1}\right)\right] - 1\right\} \qquad (7)$$

with

$$G = \frac{1}{2}\sum_{i=1}^{k}\left(\theta_i + \theta_{i-1}\right)\left(\pi_i - \pi_{i-1}\right) \qquad (8)$$

and where

$$\theta_w = \theta_{m-1} + \frac{\left(\theta_m - \theta_{m-1}\right)\left(1-\pi_{m-1}\right)}{\left(\pi_m - \pi_{m-1}\right)} \qquad (9)$$

is the ordinate of the intersection between the straight line R_m and the symmetry pole $\pi_w = 1$. Both (3.7) and (3.8) are based on the trapezoidal rule. As the Lorenz curve is concave, the estimates of *A(1)* and *A(2)* are inferior to their true values. The effect on *Z* will be negligible when the number of groups is large or when the curve is symmetric. In other cases the bias due to grouping depends on the bias affecting *A(1)/A(2)*.

4. Applications.

The Gini and Zanardi indices do not identify completely a Lorenz curve. As shown in Figure d it is possible that different Lorenz curves have the same combination of the two indices.

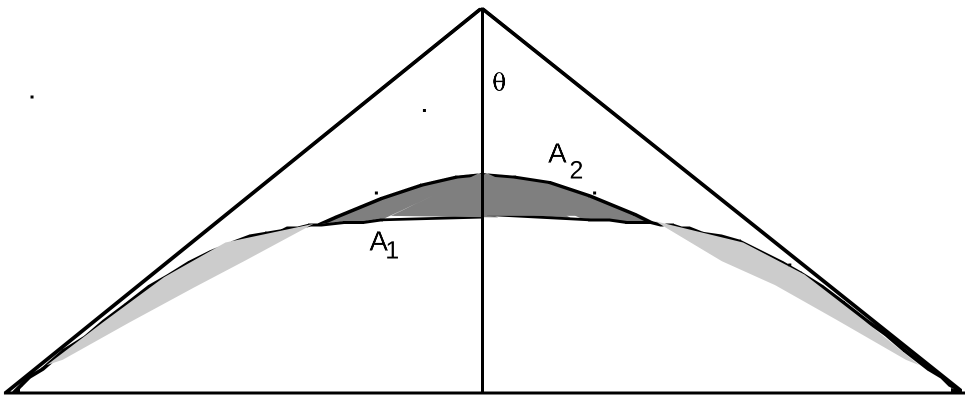

Figure d. - Symmetric Lorenz curves having the same Gini index

However, comparison of Lorenz curves can now be carried out on a two-dimensional basis, and, therefore, the ambiguity of single summary measures is reduced. In this section I report two applications which show that the combination Gini/Zanardi constitutes an efficient tool of income distribution analysis.

4.1 Income distribution and economic development

Kakwani (1980, ch.17) suggested using Lorenz curve's asymmetry to test the Kuznets' hypothesis on income distribution in developed and developing countries. In table 1 below are reported the values of G_i, G_s, G, and Z as computed[19] from the data published in Grosh and Nafzinger (1986).

Table 1. - Asymmetry of some observed Lorenz curve

Economic Country Classifications	Indices G_s	G_i	G	Z
Low income countries	0.5433	0.5132	0.4666	0.0141
Middle income countries	0.5334	0.6841	0.5436	-0.0573
Industrialized countries	0.4036	0.5105	0.3900	-0.0630
Centrally planned economies	0.2386	0.2929	0.2362	-0.0558

It is to be noted that although $|Z|$ can vary from zero to one, it is unlikely to find empirical applications for which $|Z|>0.30$. This is the real benchmark against which make comparisons[20]. In this respect the results in table 1 appear consistent with the hypothesis that in the course of economic development a Lorenz curve tends not only to enclose a smaller area but, also, to change shape: from asymmetric to the right (or from quasi-symmetric) to asymmetric to the left.

4.2 Identification of the Lorenz curve by the Gini ratio and the Zanardi index

Deciles, as expressed by the proportion of total income held by a given tenth of IRU's, are a convenient representation of the size distribution of income since each decile summarizes a different part of the distribution. Nonetheless, a tabulation of nine values can be not easy to understand even for relatively small data sets.

In this section I try to attain some degree of parsimony by collapsing the relationships between income deciles to a smaller set of composite indices. This will throw light on one of the possible ways the index of asymmetry can be put to use.

The choice of principal component analysis (see Seber, 1984 for a review of the technique) is dictated by the assumption that the matrix of correlations contains all the information on the common characteristics of deciles.

The data used for the analysis consist of 32 income distributions by deciles of various developed and developing countries[21]published by van Ginneken and Park (1984).

Table 2 reports the lower half of the matrix of correlation coefficients between income deciles. Examination of these coefficients confirms that nearly all the deciles (the only possible exception is D9) are highly correlated and that a linear reduction technique of dimensionality can be usefully applied in this setting. It also appears that the correlation coefficients decrease and become negative as one moves from the poorest decile towards the richest.

Table 2. - Matrix of correlation coefficients between income deciles

	D1	D2	D3	D4	D5	D6	D7	D8
D2	0.9405							
D3	0.8974	0.9776						
D4	0.8639	0.9534	0.9545					
D5	0.8329	0.9090	0.9285	0.9685				
D6	0.7584	0.8406	0.8937	0.9010	0.9502			
D7	0.6530	0.7351	0.8022	0.8243	0.8946	0.9599		
D8	0.4685	0.5145	0.5717	0.6273	0.7075	0.7514	0.8345	
D9	-0.3765	-0.4605	-0.4035	-0.3899	-0.3105	-0.2305	-0.0742	0.3350

Table 3 displays the eigenvalues of the correlation matrix

Table. 3 - Eigenvalues of the correlation matrix

		Percentage of variability	
Component	Eigenvalue	Component	Cumulative
1	6.8625	76.25	76.25
2	1.4941	16.60	92.85
3	0.3633	4.04	96.89
4	0.1017	1.13	98.02
5	0.0803	0.89	98.91
6	0.0539	0.60	99.51
7	0.0225	0.25	99.76
8	0.0135	0.15	99.91
9	0.0080	0.09	100.00

Only the first two components, accounting for about 93% of total variability, are of practical significance. Hence the aim of reducing the dimensionality of the problem has been achieved. The eigenvectors corresponding to the first two components are shown in table 4 (the loadings have been scaled so that the sum of their squares equals one).

Table 4: eigenvectors for first two components

Deciles	C1	C2
D1	0.3378	-0.1735
D2	0.3631	-0.1871
D3	0.3695	-0.1161
D4	0.3722	-0.0758
D5	0.3748	0.0218
D6	0.3654	0.1150
D7	0.3429	0.2612
D8	0.2685	0.5509
D9	-0.1236	0.7282

The first component accounts for about 76% of the total variability and may be considered an overall characteristic of the income distribution because of the high and approximately equal weights given to most of the deciles. The only decile with little weight is D9 which is, in effect, less correlated with the others (the highest entry in row D9 of table 2 is -0.4605 whereas all the other rows have at least one entry greater that 0.8). The second component explains about 17% of the total variability and may be regarded as a bipolar factor because negative loadings for D1-D4 contrasts positive loadings for D5-D9.

To further clarify the meaning of these components, I have computed their correlations with the Gini index (as a summary statistic of overall inequality) and the Zanardi index (as an indicator of disparity of concentration between poor and rich IRU's). The results are included in table 5.

Table 5: correlation coefficients between C1,C2,G, and Z.

	C1	C2	Gini	Zanardi
C1	1.0000			
C2	0.0000	1.0000		
Gini	-0.9980	0.0084	1.0000	
Zanardi	-0.2180	-0.7960	0.1919	1.0000

Table 5 indicates that the Gini index is an excellent proxy for the first component whereas the Zanardi index is a remarkably good substitute of the second component. These results shows that the Gini and Zanardi measures suffice to characterize the income distribution by deciles.

5 Conclusion.

In this paper I argue that the Zanardi index of asymmetry of the Lorenz curve provides a useful supplementary measure to those usually employed in income distribution analysis This measure, when used in conjunction with the well known Gini index of inequality is a valuable tool for making intertemporal and international comparisons of income distributions.

References

Champernowne D.G. (1956), Discussion on paper by Hart and Prais. J*ournal of the Royal Statistical Society, A,* 119, 181-183.

Chipman J.S. (1985). The theory and measurement of income distribution. *Advances in Econometrics,* 4,135-165.

Cortes F.C., Rubalcava R.M. (1983). Analisis del cambio en la concentration a traves del coeficiente de Gini . *Economia y Demografia,* 17, 458-488.

Dagum C. (1980). Inequality measures between income distributions with applications. *Econometrica*, 48,1791-1803

Gagliani G., A. Tarsitano (1987). Distribuzione personale del reddito, modifiche nella struttura dell'occupazione e sviluppo economico» pp. 246-251 in Strutture economiche e dinamiche dell'occupazione, A cura di C. Cazzola e A. Perrucci, La Nuova Italia Scientifica, Roma

Gini C. (1932. Intorno alle curve di concentrazionee.,*Metron,* 9, pp. 2-76.

Giurovich G. (1959). Su un indice approssimato di asimmetria della curva di concentrazione. *Statistica*,19, 463-482.

Grosh M.E., Nafzinger E.W.(1986). The computation of world income distribution. *Economic Development and Cultural Change*, 34, 347-359
Kakvani N.C. (1980). Income inequality and poverty. Methods of estimation and policy applications, Oxford University Press, Oxford.
Kakwani N.C., Podder N. (1976). Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations. *Econometrica*, 44, 137-148.
Kendall, M.G.(1956), Discussion on paper by Hart and Prais, Journal of the Royal Statistical Society, A, vol. 119, pp. 184-185.
Koo Y.C. Quan N.T., Rasche, R. (1981). Identification of the Lorenz curve by Lorenz coefficient. *Weltwirtschaftliches Archiv*, 117,124-135.
Hagerbaumer J.B. (1977). The Gini concentration ratio and the minor concentration ratio. A two-parameter index of inequality. *The Review of Economics and Statistics*, 59,377-379.
Maddala G.S., Singh S.K. (1977). Estimation problems in size distribution of incomes. *Èconomie Appliquèe*, 30, 461-480.
Musgrove P. (1980). Income distributions and the aggregate consumption function. J*ournal of Political Economy*, 88, 504-525.
Panizzon G. (1955). La concentrazione della proprietà fondiaria in Italia. *Ricerche Economiche*, vol. 9, pp. 251-278.
Patimo A. (1978). Dall'indice specificativo del rapporto di concentrazione ad un indice di asimmetria della curva di concentrazione. Atti della XXIX Riunione Scientifica della SIS, 177-187.
Piesch W. (1975), Statistische Konzentrationsmaße, J.C.B. Mohr, Tübingen.
Seber G.A.F. (1984), Multivariate Observations, John Wiley & Sons,New York
Tagughi T. (1968). Concentration curve methods and structures of skew populations. A methodology for the analysis of economic data. *Annals of the Institute of Statistical Mathematics*, 20,107-141.
Van Ginneken W., Park J. (1984). Generating internationally comparable income distribution estimates, ILO, Geneva.
Zanardi G. (1964). Della asimmetria condizionata delle curve di concentrazione. Lo scentramento. R*ivista Italiana di Economia, Demografia e Statistica*,18, 431-466.
Zanardi G. (1965). L'asimmetria statistica delle curve di concentrazione. *Ricerche Economiche*, 19, 355-396.

Notes

(1)Two populations of the same size have the same Gini ratio if the relative shares of units equidistant from the median differ, in each population, of the same amount. This may be illustrated using a discrete ungrouped income distribution $\{y_1<y_2<...<y_n\}$ with finite mean $\mu>0$. The Gini ratio can be written as

$$G = \sum_{i=1}^{\left[\frac{n+1}{2}\right]} \left(h_{n-i+1} - h_i\right) d_i$$

where $[x]$ is the greatest integer $\leq$x, $d_i=(n+1-2i)/n$, and $h_i=y_i/(n\mu)$ is the share of the i-th unit. For example, the two distributions: {2,4,6,8,10} and {7,7,8,15,23} have the same G.

(2) See Gini (1932).

[3] *p* is the fraction of income receiving units (IRU's) having an income lower than a given level and *q* is their fraction of total income.
[4] Maddala and Singh (1977) claim that *"there is no a priori reason why (p-q) should be explained by (p+q), nor is it clear what the meaning of such a relationship is"*. Since

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \sqrt{2} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

the meaning of (1) is evident: the plane *(p,q)* is first reflected, then rotated about the origin through an angle of 45º and, finally, stretched by the scalar $\sqrt{2}$.
[5] This is a consequence of the convexity of the Lorenz curve in the *(p,q)* plane.
[6] $\hat{\theta}$ is also known as Pietra index and corresponds to the relative mean deviation of incomes.
[7] A symmetric Lorenz curve is obtained (Kendall, 1956) if and only if the density function of income can be expressed as $f(y)=(\mu/y)^{(3/2)}g[Ln(y/\mu)]$ where *g(.)* ia an even function of its argument. A similar condition is given in Champernowne (1956).
[8] See Piesch (1975, p. 87-94) for alternative definitions of the symmetry of the Lorenz curve.
[9] The income y_w is a measure of central tendency which coincides with the arithmetic mean when the Lorenz curve is culminant, that is if the maximum of $L(\pi)$ is obtained at the median abscissa.
[10] Taguchi (1968) defined a measure of asymmetry based on the curvature of the Lorenz curve. Such a measure is not applicable to piece-wise linear Lorenz curves.
[11] A change of location do alter *Z*. In fact, the Gini index of the linear transformation $x=a+by$ is $G_x=(\mu_y/(a+\mu_y))G_y$, and, therefore, since poor and rich IRU's have a different mean income, it follows that $Z_x \neq Z_y$.
[12] See Taguchi (1968) for a definition of selfsymmetry. For example, the ordinary Lorenz curve is selfsymmetric to the Lorenz curve, based on the same income distribution, but cumulated from the highest grade.
[13] In particular, the third property derives from the fact that , in the (π,q) plane, (G/2) is the maximum of $|A_s-A_i|$.
[14] Curve (5) has several drawbacks: first, it is not necessarily a Lorenz curve because some of its ordinates may fall outside the maximum inequality triangle ANZ. Furthermore, it cannot generate a culminant curve. An alternative to (5) is the curve studied by Gini (1932)

$$L(\pi) = \begin{cases} R\left[1-\left(\frac{b-\pi}{b}\right)^c\right] & for \quad 0 \le \pi \le b \\ R\left[1-\left(\frac{\pi-b}{2-b}\right)^c\right] & for \quad b \le \pi \le 2 \end{cases}$$

where $b \in [0,2]$, $c \in [0,b/R]$, $d \in [0,(2-b)/R]$.
[15] The authors failed to note this aspect of the parameter *R*. In fact, in their formulation, *R* is only constrained to be positive.
[16] The recursive formula *IB(x;r,s)=1-IB(1-x;s;r)* can be used.
[17] A more plausible measure is the difference *(a-b)* proposed by Musgrove (1980).
[18] Zanardi (1965) discussed his index only for smooth curves. Patimo (1978) derived a formula of the Zanardi index for a piecewise linear Lorenz curve in the *(p,q)* plane.
[19] The entries in table I were obtained by applying the trapezoidal rule to the empirical Lorenz curve as determined by the deciles. As a result, the *G* values are slightly different from those reported by Grosh and Nafzinger.
[20] I found that the empirical distribution of *Z* (based on 306 decile's income distribution from various sources) is well approximated by a Gaussian law with mean -0.05 and standard deviation 0.09.
[21] The data refer to money income, pre- or post-tax, and either to families or households.