A NOTE ON THE INTERPRETATION AND APPLICATION OF THE GINI COEFFICIENT

by

Kwang Soo Cheong

Working Paper No. 99-1R
September 1999

A NOTE ON THE INTERPRETATION AND APPLICATION OF THE GINI COEFFICIENT

by

KWANG SOO CHEONG

Department of Economics

University of Hawaii at Manoa

Abstract

We show that the Gini coefficient is a simple linear transformation of the center of gravity of income distribution. The new derivation and inequality decomposition methods are applied to income data for Korea in order to analyze the distributional impact of the recent economic crisis. We also discuss the potential benefits of using additional higher moments of the relative income rankings.

Keywords: Income Inequality, the Gini coefficient, and the Lorenz curve.

JEL Codes: D31, D63.

Address for correspondence: Department of Economics, University of Hawaii, 2424 Maile Way, SSB #528, Honolulu, HI 96822. Phone: 808-956-7653. Fax: 808-956-4347. E-mail: kscheong@hawaii.edu

I. INTRODUCTION

The Gini coefficient is widely used as a measure of income inequality, and there have been many attempts to find an intuitive meaning to it. To mention a few examples, Yitzhaki (1979), Hey and Lambert (1980) and Berrebi and Silber (1985) showed that the Gini coefficient represents the degree of relative deprivation in a society, Lerman and Yitzhaki (1984) and Shalit (1985) related the Gini coefficient to the covariance between a household's income and its income rank, and Milanovic (1994) expressed the Gini coefficient as the weighted average of differences between each household's importance as a member of a society and its importance as an income-receiving unit. In this note, we provide a more satisfying intuitive interpretation of the Gini coefficient using the statistical properties of the Lorenz curve.

In Section II, we derive the Gini coefficient as a linear transformation of the first moment of the distribution function underlying the Lorenz curve. More specifically, the Gini coefficient is linearly related to the mean of households' relative income rankings, and thus identifies the ranking of the household on which the distribution of income is centered. In other words, the center of gravity of an income distribution is obtained as a linear transformation of the Gini coefficient. Furthermore, this new interpretation allows for not only an easy way of computing the Gini coefficient but also a useful decomposition of overall inequality into between-group and within-group components.

Section III illustrates an application of the new derivation and inequality decomposition methods, using income data for Korea. It is found that the recent economic crisis in the country has caused sharp increase in overall income inequality in parallel with a distinct process of income stratification.

In the last section, we discuss the potential benefit of using additional higher moments, especially for situations in which the Lorenz curves cross each other.

II. A NEW INTERPRETATION AND DERIVATION

The Gini coefficient is defined as the ratio of the area between the Lorenz curve and the equality line (or the 45-degree line) to the area below the equality line. Defining the Lorenz curve, $L(p)$, as a function of p where p denotes the cumulative population frequency, so that $0 \leq p \leq 1$, the Gini coefficient, G, is expressed as follows: [1]

$$G = 1 - 2\int_0^1 L(p)\, dp \ . \tag{1}$$

Since $L(p)$ is continuous (from the right), increasing in p and ranges between 0 and 1, it can be considered as a cumulative distribution function of a random variable p. The variable p now indicates the relative income ranking, 0 being the poorest and 1 being the richest, and $L'(p)$ is the corresponding probability density function. The (unconditional) mean of p, which is denoted by $E(p)$, is then obtained as follows: [2]

$$E(p) = 1 - \int_0^1 L(p)\, dp \ . \tag{2}$$

From Equations (1) and (2), we have the following relationship between the Gini Coefficient and the mean of p:

$$E(p) = \frac{1}{2}(1 + G) \ . \tag{3}$$

Equation (3) expresses the mean of the relative income rankings as a simple linear transformation of the Gini coefficient, thereby enabling us to offer an intuitive interpretation of the Gini coefficient. Although the Gini coefficient measures a geometric area

[1] The Gini coefficient can be expressed in many different ways. Yitzhaki (1998) provides a useful summary of alternative formulae.

[2] See Appendix for proof.

according to its original definition, it also finds the mean of the relative income rankings in an income distribution. The mean of the relative income rankings is simply the sum of households' relative income rankings (p) weighted by their income shares ($L'(p)$). For example, when an income distribution is completely equal, all households are equally in the middle of income rankings, and hence the mean of the relative income rankings is $\frac{1}{2}$, which corresponds to the Gini coefficient being 0 according to Equation (3). When an income distribution is completely concentrated, the richest household with the relative ranking of 1 has the total income share while all the other households have no income shares, and hence the mean is 1, which corresponds to the Gini coefficient being 1. Therefore, the mean of the relative income rankings is bounded by $\frac{1}{2}$ from below and by 1 from above with its lower value meaning a lower degree of income inequality as is the case with the Gini coefficient.

By the definition of the first moment, the mean of the relative income rankings and (hence the Gini coefficient) locates the center of gravity of an income distribution. Intuitively speaking, it finds the relative ranking of the household on which the distribution of income is centered when the households are lined up in order of income size. For example, the mean of the relative income rankings being 0.7 (or equivalently, the Gini coefficient being 0.4) means that the distribution of income is centered on the seventieth poorest household in percentile income rankings. In other words, the central tendency of the income distribution is toward the seventieth poorest household, and this household then *represents* the income distribution. Along the same line of thought, a completely concentrated distribution of income is represented by the richest household while a completely equal distribution is represented by the middle-ranked household.

The new interpretation of the Gini coefficient proposed here is not only intuitive but also important in that it provides an economically meaningful rationale for extending the use of the Gini coefficient to the cases of intersecting Lorenz curves. Whether the curves intersect or not, income distributions are evaluated by their representative income rankings (or equivalently, their centers of gravity), not by the particular geometric areas which seemingly have no economic meaning.

In practice, the relative income rankings, p, are not continuous, and the Gini coefficient can be calculated using the discrete version of Equation (3). Let y_i denote the income of the i-th poorest household, N the total population, Y the total income (such that $Y = \sum_{i=1}^{N} y_i$). Then the mean of the relative income rankings (henceforth, *the center of gravity* of the income distribution), **E** and the Gini coefficient, **G** are obtained as follows:

$$\mathbf{E} = \sum_{i=1}^{N} \frac{i}{N} \frac{y_i}{Y} \ . \tag{4}$$

$$\mathbf{G} = -1 + 2\mathbf{E} \ . \tag{5}$$

Computing the Gini coefficient using the above formulas requires only a sorted income vector and is much simpler than existing methods such as the matrix algorithm in Milanovic (1994), or the covariance method in Lerman and Yitzhaki (1984) and Shalit (1985).

An additional advantage of using the center of gravity of income distribution is that it allows for a useful decomposition of inequality changes into two parts: one due to within-group inequality and the other due to between-group inequality. [3] Suppose that an ordered income distribution is partitioned into income groups (or strata) with equal group sizes, such as income deciles. The center of gravity can be computed within each group by re-ranking the households in the group, ignoring the income rank assigned in the total

[3] The Gini coefficient can be decomposed too since it is merely a linear transformation of the center of gravity; however, the direct decomposition of the Gini coefficient is more complicated. See, for example, Lambert and Aronson (1993)and Sastry and Kelkar (1994) for different ways of decomposing the Gini coefficient.

population. [4] Denoting the center of gravity within the j-th (poorest) group by E_j, the number of income groups by K, the total income of group j by Y_j, we obtain the following relationship: [5]

$$E = \sum_{j=1}^{K} \frac{1}{K} \frac{Y_j}{Y} E_j + \sum_{j=1}^{K} \frac{j}{K} \frac{Y_j}{Y} - \frac{1}{K} \ . \tag{6}$$

The first term measures the contribution of within-group inequality to overall income inequality as the weighted sum of each group's center of gravity (E_j) with the weight for each group being the product of the group's population share ($\frac{1}{K}$) and income share ($\frac{Y_j}{Y}$). The second term is simply the center of gravity of the distribution of group income, $(Y_1, Y_2, \cdots, Y_K)$; that is, it captures the contribution of the between-group inequality to overall income inequality. The last term is constant, depending upon only the number of income groups. Therefore, as long as the same number of income groups is maintained, one can precisely traced what fraction of inequality changes are attributed to inequality within groups or inequality between groups.

III. AN EMPIRICAL EXAMPLE

To illustrate the use of the center of gravity and the decomposition method, we used income data from the Urban Household Income and Expenditure Survey (UHIES) conducted by the National Statistical Office in Korea. The UHIES collects monthly income data from over three thousand representative worker households living in the seventy-two cities in Korea, and their income data are publicly available on quarterly basis. We chose the

[4] For example, the poorest household in each group is assigned the income rank 1 no matter which income group it belongs to.

[5] See Appendix for proof.

quarters from the first of 1996 to the first of 1999 in order to analyze the distributional impact of the economic crisis erupting in the last quarter of 1997.

The computed results are summarized in Table 1. The table shows that the center of gravity for the first quarter of 1999 is higher than that for any quarters of 1998, which is in turn higher than that of any previous quarters. For example, the household representing the sixty-fourth poorest percentile was at the center of income distribution just prior to the crisis in the third quarter of 1997. However, the income distribution in the first quarter of 1999 was centered on the sixty-eighth poorest household. In terms of the Gini coefficient, this amounts to an increase of 26.2% from 0.2772 to 0.3599. Clearly, income inequality among worker households has sharply increased as a result of the economic crisis.

While the deterioration of income distribution is hardly surprising given the nature of the socio-economic changes brought by the economic crisis, the result from decomposition analysis reveals an interesting phenomenon. In Table 1, inequality changes are decomposed for income deciles (that is, $K = 10$). First, the table shows that the center of gravity of overall income distribution, E, and the center of gravity of group income distribution, E_K, move in the same direction for all quarters, implying that an increase (decrease) in overall income inequality always accompanies an increase (decrease) in between-group income inequality. The table also shows that that is not necessarily true for the within-group component of overall inequality. For example, the increase in E_K exceeds the increase in E between the first and second quarters of 1998, which means that there was a decrease in within-group inequality. In fact, such *overshooting* of between-group inequality almost forms a pattern after the onset of the crisis as it is found for four quarters out of a total of six quarters. The average contribution of between-group inequality during the six quarters is about 122% of the changes in overall inequality. Therefore, it is believed that severe deterioration in between-group inequality has more than offset minor improvement in within-group inequality, resulting in deterioration of overall income inequality. In other words, the worker households in Korea are undergoing a distinct process of income stratification parallel with the concentration of income.

IV. AN EXTENDED IDEA

Although the literature provides normative principles that can be used when the Lorenz curves intersect, [6] conservative researchers have limited the usage of the Gini coefficient to the cases of non-intersecting Lorenz curves; that is, the cases in which one Lorenz curve dominates the other in the sense of first degree stochastic dominance. It seems partly due to not being able to directly relate the geometrical definition of the Gini coefficient with the underlying aspects of an income distribution. The new interpretation offered in this note, however, validates the use of the Gini coefficient regardless of whether the Lorenz curves cross each other, so long as we intend to rank income distributions according to their centers of gravity. [7]

Making further use of the properties of the Lorenz curve as a cumulative distribution function may yield important information about a given income distribution, which can

[6] Considering the principle of diminishing transfers, Kolm (1976) and Shorrocks and Foster (1987) derived a sufficient and necessary condition under which all inequality indices based upon this principle leads to unanimous rankings of income distributions as long as their Lorenz curves intersect only once. Recently, Davies and Hoy (1995) extended this condition to the case in which the Lorenz curves intersect a finite number of times and Beach, Davidson and Slotsve (1994) provided the statistical basis for empirical application of the condition.

[7] One may define a social welfare function using the Gini coefficient, such as the one proposed by Sheshinski (1972), and consider maximizing this social welfare function. However, this type of social welfare function is subject to criticism. Among other things, it is not compatible with a strictly quasiconcave social welfare function, and it gives more weight to transfers near the mode of an income distribution than at the tails. See, for example, Bishop, Chakraborti and Thistle (1991) and Ch.5 in Kakwani (1980).

be overlooked by simply using the Gini coefficient. In general, a cumulative distribution function can be uniquely determined by the entire set of moments. [8] Therefore, we get more information about the structure of an income distribution as we compute additional moments based upon the corresponding Lorenz curve. For example, the second and third moments about the mean will measure the dispersion and skewness, respectively, of the relative income rankings.

Consider the two Lorenz curves shown in Figure 1. Lorenz curve A shows a relatively more equal distribution among the low-income households while Lorenz curve B shows a relatively more equal distribution among the high-income households. The two curves are, though, geometrically symmetric and their Gini coefficients are identical. In this case, the Gini coefficient based upon the first moment of the Lorenz curve fails to capture the critical differences in the two income distributions. One would conclude, solely on the basis of their Gini coefficients, that the two income distributions are equally unequal. This conclusion is hardly satisfactory to those who have seen the Lorenz curves, [9] and here we suggest the use of the second moment in addition to the Gini coefficient.

The second moment about the mean, the variance, measures the degree of dispersion of households' relative income rankings. Denoted by $Var(p)$, the variance can be written as [10]

$$Var(p) = 2\int_0^1 L(p)\,dp - \{\int_0^1 L(p)\,dp\}^2 - 2\int_0^1 pL(p)\,dp\ . \tag{7}$$

Since the two income distributions in consideration have the same Gini coefficients, the first two terms in Equation (7) do not make any difference. It is the last term, particularly the expression $\int_0^1 pL(p)\,dp$, that can distinguish the income distributions by their variances.

[8] One sufficient condition frequently noted is the existence of the moment generating function.

[9] See Wolff (1997) for expository discussion of a similar example.

[10] See Appendix for proof.

It is obvious from the figure that that expression has a higher value for Lorenz curve B; consequently, Lorenz curve A generates a larger variance than Lorenz curve B. In other words, the income distribution showing more equality among low-income households shows a higher degree of dispersion of the relative income rankings.

In practice, the variance of the relative income rankings, denoted by **VAR**, can be computed from the following formula: [11]

$$\mathbf{VAR} = \sum_{i=1}^{N} \left[\frac{i}{N} \right]^2 \frac{y_i}{Y} - \left[\mathbf{E} \right]^2 , \tag{8}$$

where **E** is already obtained from Equation (4), making it such that little additional effort is required to compute the variances along with the Gini coefficients. As simple as it is, computing the variances seems an economic, if not indispensable, procedure to take, especially when the Lorenz curves are not drawn or the Gini coefficients show little difference.

Although the variance of the relative income rankings may reveal differences that the Gini coefficient fails to demonstrate, it remains a value judgement as to which of the two income distributions is more desirable. In a sense, it can be thought of as a decision of how to distribute weights of importance over households. If, for example, we give more weight to low-income households relative to high-income households, then we can conclude that the income distribution generating Lorenz curve A is more desirable than the one generating Lorenz curve B, despite their identical Gini coefficients. However, it should be noted that there is already a distribution of weights implicitly built in the Gini coefficient [12] and these weights are necessarily inconsistent with the weights newly assigned for comparing variances.

The above discussion clearly demonstrates the potential usefulness of computing the

[11] This formula is a discrete version of Equation (13) in Appendix.

[12] As is well known, each household's weight implicit in the Gini coefficient is determined by its income ranking. See, for example, Ch.2 in Sen (1997).

variance of the relative income rankings along with the Gini coefficient. A comparison based upon the Gini coefficients alone may disregard potentially important differences. In the same vein, one should not focus only on the variance and ignore the implication of the Gini coefficient. In general, we will get a better *description* of an income distribution as we compute additional moments. On the other hand, we need a convenient means of *evaluating* income distributions in terms of income inequality. This trade off between details and convenience basically comes out of our lack of consensus with regards to equity criteria and is, hence, an inevitable issue in income studies. As a consequence, we have yet to answer questions such as up to which higher moment we should compute and how the conflicting equity implications can be balanced.

Table 1.   Changes in Overall Inequality and Between-Group Inequality

Quarter	1996-1	1996-2	1996-3	1996-4	1997-1	1997-2	1997-3	1997-4	1998-1	1998-2	1998-3	1998-4	1999-1
G	0.2942	0.2670	0.2756	0.2932	0.2904	0.2653	0.2772	0.2761	0.3188	0.3352	0.3300	0.3232	0.3499
E	0.6471	0.6335	0.6378	0.6466	0.6452	0.6327	0.6386	0.6381	0.6594	0.6676	0.6650	0.6616	0.6750
E_K	0.4986	0.4918	0.4929	0.5009	0.4972	0.4887	0.4915	0.4905	0.5061	0.5150	0.5111	0.5070	0.5197
ΔE		-0.0136	0.0043	0.0088	-0.0014	-0.0126	0.0060	-0.0006	0.0214	0.0082	-0.0026	-0.0034	0.0134
ΔE_K		-0.0067	0.0011	0.0080	-0.0038	-0.0084	0.0028	-0.0010	0.0156	0.0090	-0.0040	-0.0040	0.0127
$\Delta E_K / \Delta E$		49.57%	25.43%	91.31%	270.33%	67.22%	46.25%	181.07%	73.04%	109.35%	153.12%	118.64%	95.14%

<u>Note</u>

G: The Gini coefficient
E: The Center of Gravity of overall income distribution (overall inequality)
E_K: The Center of Gravity of group income distribution (between-group inequality)
ΔE: change in E.
ΔE_K: change in E_K.
ΔE_K / ΔE: percentage ratio of E_K to E.

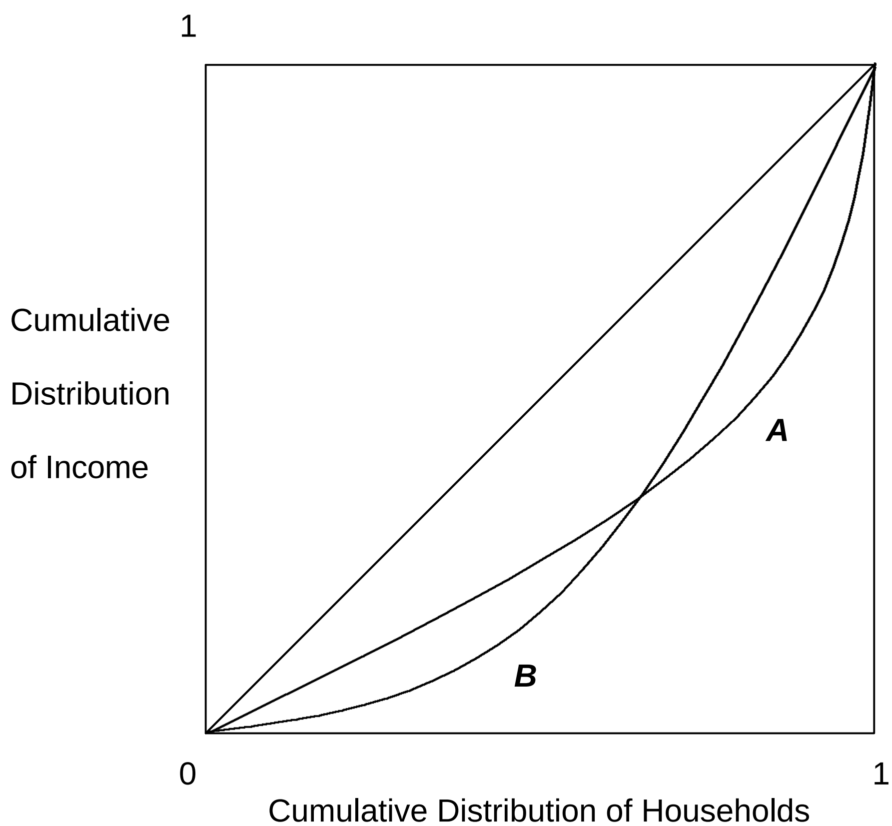

Figure 1. Two Lorenz Curves ***A*** and ***B*** are based upon considerably different income distributions. However, their Gini coefficients are identical since the curves are geometrically symmetric. In this case, the differences are captured by the variances of relative income rankings.

APPENDIX

Proof of Equation (2):

The mean of p is given by

$$E(P) = \int_0^1 p\, dL \ . \tag{9}$$

Substituting $L'(p)dp$ for dL and using integration by parts, we obtain

$$E(P) = \Big[\ pL(p)\ \Big]_0^1 - \int_0^1 L(p)\, dp \ . \tag{10}$$

The first term of Equation (10) is 1, and therefore Equation (2) follows. *Q.E.D.*

Proof of Equation (6):

Denote the income rankings of the households in the j-th poorest income group by $(G_{j-1}+1, G_{j-1}+2, \cdots, G_j)$, where $G_0 = 0$ and $G_K = N$. Also denote the number of household in each group by n, that is, $n = \frac{N}{K}$. Then E can be written as

$$E = \sum_{j=1}^{K} \left[\sum_{i=G_{j-1}+1}^{G_j} \frac{i}{N} \frac{y_i}{Y} \right] \ . \tag{11}$$

Since

$$\begin{aligned}
\sum_{i=G_{j-1}+1}^{G_j} \frac{i}{N}\frac{y_i}{Y} &= \sum_{i=1}^{G_j - G_{j-1}} \frac{i+G_{j-1}}{N} \frac{y_{i+G_{j-1}}}{Y} \\
&= \sum_{i=1}^{G_j - G_{j-1}} \frac{i}{N} \frac{y_{G_{j-1}+i}}{Y} + \sum_{i=1}^{G_j - G_{j-1}} \frac{G_{j-1}}{N} \frac{y_{G_{j-1}+i}}{Y} \\
&= \frac{n}{N}\frac{Y_j}{Y} \sum_{i=1}^{G_j - G_{j-1}} \frac{i}{n} \frac{y_{G_{j-1}+i}}{Y_j} + \frac{G_{j-1}}{N} \sum_{i=1}^{G_j - G_{j-1}} \frac{y_{G_{j-1}+i}}{Y} \\
&= \frac{n}{N}\frac{Y_j}{Y} E_j + \frac{G_{j-1}}{N} \frac{Y_j}{Y} \ ,
\end{aligned}$$

we obtain

$$E = \sum_{j=1}^{K} \frac{n}{N}\frac{Y_j}{Y} E_j + \sum_{j=1}^{K} \frac{G_{j-1}}{N} \frac{Y_j}{Y} \ . \tag{12}$$

Since $G_{j-1} = n(j-1)$, the second term of Equation (12) can be written as follows:

$$\sum_{j=1}^{K} \frac{n(j-1)}{N} \; \frac{Y_j}{Y}$$
$$= \sum_{j=1}^{K} \frac{(j-1)}{K} \; \frac{Y_j}{Y}$$
$$= \sum_{j=1}^{K} \frac{j}{K} \; \frac{Y_j}{Y} - \frac{1}{K} \; \sum_{j=1}^{K} \; \frac{Y_j}{Y}$$

Since $\sum_{j=1}^{K} \; \frac{Y_j}{Y} \; = 1$, we obtain Equation (6) from Equation (12). *Q.E.D.*

Proof of Equation (7):

The variance of p is given by

$$Var(p) = \int_0^1 p^2 \; L'(p) \, dp - \; \Big[\; E(p) \; \Big]^2 \; . \tag{13}$$

Using integration by parts, the first term of Equation (13) can be written as

$$1 - 2 \int_0^1 p \; L(p) \, dp \; . \tag{14}$$

Substituting Equations (2) and (14) into Equation (13) yields Equation (7). *Q.E.D.*

REFERENCES

Charles Beach, Russell Davidson and George Slotsve, 1994, "Distribution-Free Statistical Inference for Inequality Dominance with Crossing Lorenz Curves," Institute for Economic Research Discussion Paper #912, Queen's University, Canada.

Z. M. Berrebi and Jacques Silber, 1985, "Income Inequality Indices and Deprivation: A Generalization," *Quarterly Journal of Economics*, 809-810.

John A. Bishop, Subhabrata Chakraborti, and Paul D. Thistle, 1991, "Relative Deprivation and Economic Welfare: A Statistical Investigation with Gini-Based Welfare Indices," *Scandinavian Journal of Economics* 93, 421-437.

John Hey and Peter Lambert, 1980, "Relative Deprivation and the Gini Coefficient: Comment," *Quarterly Journal of Economics*, 566-573.

Nanak Kakwani, 1980, *Income Inequality and Poverty: Methods of Estimation and Policy Applications*, Oxford University Press.

Serge-Christopher Kolm, 1976, "Unequal Inequalities II," *Journal of Economic Theory* 13, 82-111.

Peter J. Lambert and J. Richard Aronson, 1993, "Inequality decomposition analysis and the Gini coefficient revisited," *The Economic Journal* 103, 1221-1227.

Robert Lerman and Shlomo Yitzhaki, 1984, "A Note on the Calculation and Interpretation of the Gini Index," *Economics Letters* 15, 363-368.

Branko Milanovic, 1994, "The Gini-type Functions: An Alternative Derivation," *Bulletin of Economic Research* 46, 81-90.

D. V. S. Sastry and Ujwala R. Kelkar, "Note on the decomposition of Gini inequality," *The Review of Economics and Statistics* 1994, pp.584-586.

Amartya Sen, 1997, *On Economic Inequality*, (enlarged edition), Clarendon Press, Oxford.

Haim Shalit, 1985, "Calculating the Gini Index of inequality for Individual Data," *Oxford Bulletin of Economics and Statistics* 47, 185-189.

Eytan Sheshinski, 1972, "Relation Between a Social Welfare Function and the Gini Index of Income Inequaliy," *Journal of Economic Theory* 4, 98-100.

Anthony Shorrocks and James Foster, 1987, "Transfer Sensitive Inequality Measures," *Review of Economic Studies* 54, 485-497.

Edward Wolff, 1997, Chapter 3 in *Economics of Poverty, Inequality and Discrimination*, South-Western College Publishing, Cincinnati.

Shlomo Yitzhaki, 1979, "Relative Deprivation and the Gini Coefficient," *Quarterly Journal of Economics*, 321-324.

Shlomo Yitzhaki, 1998, "More than a Dozen Alternative Ways of Spelling Gini," *Research on Economic Inequality* 8, 13-30.