How Has the Literature on Gini's Index Evolved in the Past 80 Years?*

Kuan Xu

Department of Economics

Dalhousie University

Halifax, Nova Scotia

Canada B3H 3J5

January 2004

Abstract

The Gini coefficient or index is perhaps one of the most used indicators of social and economic conditions. From its first proposal in English in 1921 to the present, a large number of papers on the Gini index has been written and published. Going through these papers represents a demanding task. The aim of this survey paper is to help the reader to navigate through the major developments of the literature and to incorporate recent theoretical research results with a particular focus on different formulations and interpretations of the Gini index, its social welfare implication, and source and subgroup decomposition.

1 Introduction

Since the Gini coefficient or index as a summary statistics bore Gini (1912, 1914, 1921)'s name as we now know it, the theoretical literature has evolved for more than 80 years. During the past 80 years, the Gini index gradually became one of the principal inequality measures in the discipline of economics. This measure is understood by many economists and has been applied in numerous empirical studies and policy research.[1] As many are aware, research on inequality and poverty measurement continues to evolve. Many economists, experienced and newly minted, always wish to have the collection of the theoretical results on the Gini index handy and accessible. Anand (1983) and Chakravarty (1990) provided comprehensive surveys on the measures of inequality including Gini index. But the literature is in a constant state of flux even in the research area which is considered to be well established. Other authors such as Lambert (1989), Silber (1999), and Atkinson and Bourguignon (2000) also provided comprehensive references for income inequality and poverty with the Gini index as one of many inequality measures. This survey paper differs from those references in that it collects the theoretical results only on the Gini index, old and more recent, in one place.[2]

---

[1] Many economists in China such as Mr. Li Shi also worked on the Gini measures using the Chinese data. According to the news conference given by Premier Zhu Rongji of the People's Republic of China in March 15, 2001, the Gini index calculated in 1999 in China reached 0.39 which was considered at an alarming level by the international standard. The Chinese Government vowed to solve the problem during the development process.

[2] Thank Buhong Zheng for pointing out an excellent technical survey paper on the Gini index by Yizhaki (1998), which presents the results only based on the continuous distributions. Hence the current paper may still be justifiable given it focuses on both the history and technical results and considers both discrete and continuous distributions.

This survey paper is a natural continuation of Anand (1983) and Chakravarty (1990) on this specific literature and attempts to incorporate additional research results on the Gini index. It is hoped that this paper will not only provide readers a summary of main theoretical literature but also cover some related issues such as different formulations and interpretations of the Gini index, its social welfare implication, and source and subgroup decomposition.

The Gini index can be used to measure the dispersion of a distribution of income, or consumption, or wealth, or a distribution of any other kinds. But the kind of distributions where the Gini index is used most is the distribution of income. For this reason, and for simplicity, this paper will focus on the Gini index in the context of income distributions although its applications should not be limited to income distributions. An income distribution may be for different incomes: household incomes or individual incomes. The choice of the income unit is often determined by the purpose of research. For simplicity, the discussion in this paper is based on income distributions of the individuals within the population. Even the concept of income distribution can vary from the incomes that are pre tax and other fiscal transfers to those that are after tax and other fiscal transfers. As a convention, this is not the focus of the paper. Instead, the paper focuses on theoretical results and interpretations. Similarly, the transformation from a family income to individual incomes via equivalent scale will not be discussed here.

This paper will show that the Gini index has many different formulations and interesting interpretations. It can be expressed as a ratio of two regions defined by a 45 degree line and a Lorenz curve in a unit box, or a function of Gini's mean difference, or a covariance between incomes and their ranks,

or a matrix form of a special kind. Each formulation has its own appeal in a specific context.

The Gini index was proposed as a summary statistics of dispersion of a distribution. It was viewed, for a quite long time, not too different from other dispersion measures such as variance and standard deviation. But when coming to a decision as to which inequality measure should be adopted in a study, economists found that it was rather difficult to select one statistics over others without any justification in terms of social welfare implication. Thus, they started to search the link between the existing inequality measures and their underlying social welfare functions. It is now known that many well-known inequality measures indeed have direct, although implicit, relations with social welfare functions and that the measured inequality can be interpreted as social welfare loss due to inequality. With this intellectual premise, the social welfare implication of a Gini index value can now be interpreted with a greater clarity.

Economists also examined how the Gini index as an aggregate inequality measure could be decomposed according to either income sources or subpopulation groups. A great effort has been made to specify the conditions under which such decompositions are feasible. Even when decompositions are feasible, it is not always clear what meaningful interpretation each and every decomposed parts of the Gini index has. More specifically, when subgroup decomposition is made of the Gini index, one term called the crossover term appears difficult to interpret. Over time, this view has changed. Now many economists found that this term can be viewed as a measure of income stratification or the degree to which the incomes of different social groups

cluster.

The remaining paper is organized as follows. Section 2 introduces necessary mathematical symbols and basic definitions. Section 3 reviews the evolution of computation methods of the Gini index. Section 4 revisits the literature on Pigou-Dalton's principle of transfers and social welfare implication of the Gini index. Income source and subgroup decomposition are discussed in Section 5. Finally, Section 6 concludes.

2 Mathematical Symbols and Definitions

There are generally two different approaches for analyzing theoretical results of the Gini index: one is based on discrete distributions; the other on continuous distributions. The latter demands certain conditions on continuity while the former does not require such conditions. The discrete income distribution is easy to understand in some cases while the continuous income distribution can simplify some derivations in some situations. The two can be unified as shown in Dorfman (1979).

When the distribution function is discrete, y takes n values that can be denoted by an $n \times 1$ column vector $\mathbf{y} = [y_1, y_2, \ldots, y_n]^\top$ such that the elements in the vector are arranged in non-decreasing order: $y_1 \leq y_2 \leq \cdots \leq y_n$. The values of y are bounded below by $a = 0$ and above by $b < +\infty$, or $y_i \in [a, b]$ for $i = 1, 2, \ldots, n$. The notation "$\sim$" is used to sort $\mathbf{y}$ in the opposite order; i.e., the elements of $\widetilde{\mathbf{y}} = [\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_n]^\top$ are arranged in non-increasing order: $\widetilde{y}_1 \geq \widetilde{y}_2 \geq \cdots \geq \widetilde{y}_n$. The discrete cumulative distribution function is $F_i = \frac{i}{n}$, while the probability of y taking on the value y_i is $f_i = \frac{1}{n}$ for $i = 1, 2, \ldots, n$.

$F(y_k) = \sum_{i=1}^{k} \frac{i}{n}$ is the cumulative probability up to y_k and can be interpreted as the population share of those whose incomes are less than or equal to y_k.[3] The mean income, μ_y, is given by $\mu_y = \frac{1}{n} \sum_{i=1}^{n} y_i$. The cumulative income shares of the population up to the individual whose income is ranked ith from the lowest to the highest are given by

$$L_i = \frac{1}{n\mu_y} \sum_{j=1}^{i} y_j, \tag{1}$$

for $i = 1, 2, \ldots, n$. L_0 is defined as zero while $L_n = 1$. Note that L_i's are arranged in non-decreasing order. Sometimes, economists wish to use $\widetilde{L}_i$ to indicate L_i's that are arranged in non-increasing order [see equation (29)].

When the income distribution is continuous, y can be viewed as a value of the cumulative distribution function of income $F(y)$ or a distribution function of income $f(y)$. In general, y is bounded below by $a = 0$ and above by $b < +\infty$. $F(a) = 0$ while $F(b) = 1$. $F(y^*) = \int_a^{y^*} f(y)dy$ is the cumulative probability up to y^*. The mean income, μ_y, is given by $\mu_y = \int_a^b y dF(y) = \int_a^b y f(y)dy$. The cumulative income share of the population up to the individual whose income is y^* is given by

$$L(p^*) = L(F(y^*)) = \frac{1}{\mu_y} \int_a^{y^*} y f(y)dy. \tag{2}$$

The Lorenz curve was hinted by Sir Leo Chiozza Money (1905) and originally proposed by Mr. M. O. Lorenz in 1907.[4] It is denoted as $L(p) =$

---

[3]When the complex sampling survey designs are used to collect income data, the researcher must deal with the issues of statistical inference and sampling weights.

[4]See *Publications of the American Statistical Association*, Vol. ix, pp. 209ff.

$L(F(y))$, the proportion of the total income of the economy that is received by the lowest $100p$ of the population for all possible values of p. In other words, the graph of F and L is the Lorenz curve with $0 \leq F \leq 1$ and $0 \leq L \leq 1$. For a discrete distribution, $\frac{k}{n} = F(y_k)$, $F^{-1}(\frac{k}{n}) = y_k$, and the Lorenz curve is $L_k = L(\frac{k}{n}) = L(F(y_k))$ for $0 \leq \frac{k}{n} \leq 1$. For a continuous distribution, $p = F(y)$, $F^{-1}(p) = y$, and the Lorenz curve is $L(p) = L(F(y))$ for $0 \leq p \leq 1$. Figure 1 shows a Lorenz curve is below the 45 degree line. This reflects that the income share grows at a much slower rate as the population share increases and that there exists a higher degree of income concentration within the population.

Figure 1. Lorenz Curve
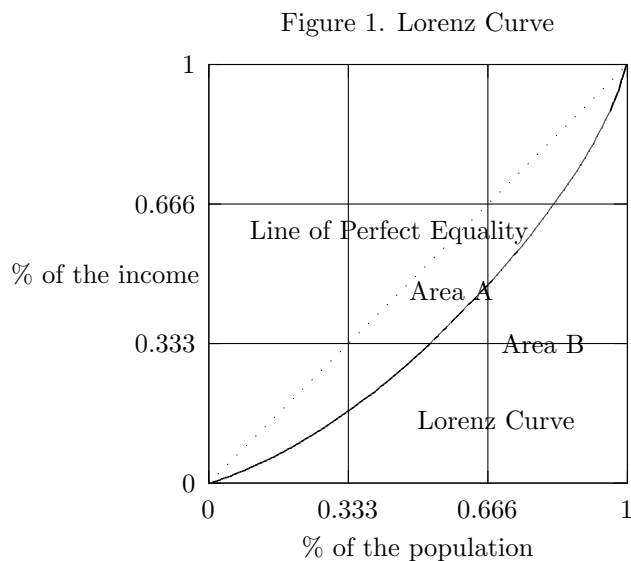
3   Evolution of Computational Methods

The Gini index as is called today was, according to Dalton (1920, p. 354), named after the fact that "a remarkable relation has been established be-

tween this measure of inequality and the relative mean difference, the former measure being always equal to half the latter." This remarkable relation was first given by Gini in 1912. Dalton (1920, p. 353) therefore called this mean difference as "Professor Gini's mean difference."

The computational methods for the Gini index include the geometric approach, Gini's mean difference approach (or the relative mean difference approach), covariance approach, and matrix form approach. Each approach has its own appeal and is desirable in some way but all can be unified and are consistent with one another. These methods and their technical justifications are examined in the following.

3.1 Geometric Approach

The attractiveness of the Gini index to many economists is that it has an intuitive geometric interpretation. That is, the Gini index can be, as in Figure 1, defined geometrically as the ratio of two geometrical areas in the unit box: (a) the area between the line of perfect equality (45 degree line in the unit box) and the Lorenz curve, which is called Area A and (b) the area under the 45 degree line, or Areas $A + B$. Because Areas $A + B$ represents the half of the unit box, that is, $A + B = \frac{1}{2}$, the Gini index, G, can be written as

$$
\begin{aligned}
G &= \frac{A}{A+B} \\
&= 2A \\
&= 1 - 2B.
\end{aligned}
\tag{3}
$$

If one works with a discrete income distribution, he or she can compute

F_i's and L_i's and then the area below the Lorenz curve

$$B = \frac{1}{2} \sum_{i=0}^{n-1} (F_{i+1} - F_i)(L_{i+1} + L_i). \qquad (4)$$

Substituting equation (4) into equation (3) yields the Gini index G:[5]

$$G = 1 - \sum_{i=0}^{n-1} (F_{i+1} - F_i)(L_{i+1} + L_i). \qquad (5)$$

To illustrate how to use the above definition, let the hypothetical income distribution be $y_1 = 0$, $y_2 = 1$, $y_3 = 2$. For this distribution, the Lorenz curve can be described by the points $(L_1 = 0, F_1 = \frac{1}{3})$, $(L_2 = \frac{1}{3}, F_2 = \frac{2}{3})$, and $(L_3 = 1, F_3 = 1)$ in the unit box . As indicated in Figure 2, Area B is the sum of the area of a small triangle $(\frac{1}{18})$, the area of a square $(\frac{1}{9})$, and the area of a large triangle $(\frac{1}{9})$:

$$B = \frac{1}{18} + \frac{1}{9} + \frac{1}{9} = \frac{5}{18}.$$

The Gini index, as indicated in equation (5),

$$G = 1 - \left[\left(\frac{1}{3} \right)(0) + \left(\frac{1}{3} \right)\left(\frac{1}{3} \right) + \left(\frac{1}{3} \right)\left(\frac{4}{3} \right) \right] = \frac{4}{9}.$$

---

[5]There are various expressions of this definition. For example, Yao [1999, p. 1251, equation (1)] adopted a spread sheet approach using this method. Osberg and Xu [2000, p. 59, equation (14)] modified the definition to accommodate the complex sampling survey data.

Figure 2. Lorenz Curve and Gini Index
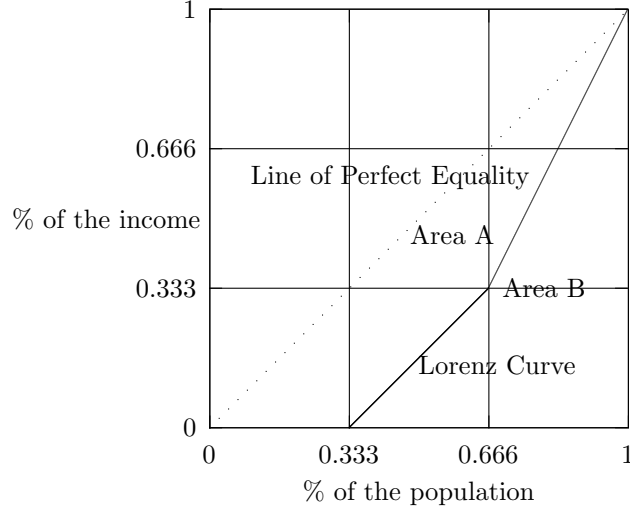
Several alternative formulations in fact follow the same tradition, Rao (1969) showed that the Gini index can be defined as

$$G = \sum_{i=1}^{n-1} \left(F_i L_{i+1} - F_{i+1} L_i \right). \tag{6}$$

This formulation can be shown to be consistent to equation (5): given $F_n = L_n = 1$ and $F_0 = L_0 = 0$, the Gini index defined in equation (5) can be rewritten as

$$G = 1 + \sum_{i=0}^{n-1} \left(F_i L_{i+1} - F_{i+1} L_i \right) - \sum_{i=0}^{n-1} \left(F_{i+1} L_{i+1} - F_i L_i \right).$$

Since the last term on the right-hand side is one, we have equation (6).

11

Sen (1973) defined the Gini index as

$$G = \frac{n+1}{n} - \frac{2}{n^2\mu_y} \sum_{i=1}^{n} (n+1-i)\, y_i. \tag{7}$$

This formulation illustrates that the income-rank-based weights are inversely associated with the sizes of incomes. That is, in the index the richer's incomes get lower weights while the poor's income get higher weights. Sen's definition[6] can be derived from equation (6) by noting

$$
\begin{aligned}
G &= \sum_{i=1}^{n-1} (F_i L_{i+1} - F_{i+1} L_i) \\
&= \sum_{i=1}^{n} (F_{i-1} L_i - F_i L_{i-1}) \\
&= \sum_{i=1}^{n} (F_i (L_i - L_{i-1}) - (F_i - F_{i-1}) L_i) \\
&= \frac{1}{n^2\mu_y} \sum_{i=1}^{n} \left(i y_i - \sum_{j=1}^{i} y_j \right)
\end{aligned} \tag{8}
$$

given $F_i = \frac{i}{n}$, $L_i = \frac{1}{n\mu_y} \sum_{j=1}^{i} y_j$, $F_i - F_{i-1} = \frac{1}{n}$, and $L_i - L_{i-1} = \frac{y_i}{n\mu_y}$. The expression for G can be further manipulated as

$$
\begin{aligned}
G &= \frac{1}{n^2\mu_y} \left[\sum_{i=1}^{n} i y_i - \sum_{i=1}^{n} \sum_{j=1}^{i} y_j \right] \\
&= \frac{1}{n^2\mu_y} \left[\sum_{i=1}^{n} i y_i - \sum_{i=1}^{n} (n+1-i) y_i \right] \\
&= \frac{1}{n^2\mu_y} \left[\sum_{i=1}^{n} (n+1) y_i - 2 \sum_{i=1}^{n} (n+1-i) y_i \right] \\
&= \frac{n+1}{n} - \frac{2}{n^2\mu_y} \sum_{i=1}^{n} (n+1-i) y_i.
\end{aligned} \tag{9}
$$

The last equality is consistent with equation (7).

Fei and Ranis (1974) and Fei, Ranis, Kuo (1978) defined the Gini index

---

[6]Sen (1997) gives a slightly different definition using the incomes sorted in non-increasing order.

as a linear function of a u_y-index:

$$G = \frac{2}{n}u_y - \frac{n+1}{n}. \tag{10}$$

where the u_y-index is given by

$$u_y = \frac{\sum_{i=i}^{n} iy_i}{\sum_{i=1}^{n} y_i}.$$

Substituting $u_y = \frac{\sum_{i=1}^{n} iy_i}{\sum_{i=1}^{n} y_i}$ into equation (10) yields

$$
\begin{aligned}
G &= \frac{2}{n^2\mu_y}\sum_{i=1}^{n} iy_i - \frac{n+1}{n} \\
&= \frac{n+1}{n} - \frac{2(n+1)}{n} + \frac{2}{n^2\mu_y}\sum_{i=1}^{n} iy_i \\
&= \frac{n+1}{n} - \frac{2}{n^2\mu_y}\sum_{i=1}^{n}(n+1-i)y_i,
\end{aligned} \tag{11}
$$

which is consistent with equation (7).

If one deals with a continuous income distribution, he or she can express the area under the Lorenz curve as

$$B = \int_0^1 L(p)dp. \tag{12}$$

Substituting equation (12) into equation (3) yields the Gini index for the continuous income distribution as

$$G = 1 - 2\int_0^1 L(p)dp. \tag{13}$$

Clearly, it is much simpler to understand the Gini index geometrically. However, its computation may be tedious.

13

3.2 Gini's Mean Difference Approach

Differing from the geometric approach, Gini (1912) showed that the geometric approach is in fact related to the statistical approach via a concept called the (absolute and relative) mean difference. That is, the Gini index as a ratio of two areas defined above is always equal to the half of the relative mean difference that will be explained later.

According to David (1968), the relative mean difference discussed by and named after Gini (1912) was in fact discussed much earlier by F. R. Helmert and other German writers in the 1870's. In 1912, Gini's book was published in Italian and hence was not accessible to English-speaking economists at the time. In 1921 when commenting on Dalton's (1920) work, Gini (1921) explained his work (1912) and related literature in English in a short Economic Journal article. Since then, the Gini index and Gini's relative mean difference were made known in the literature of income inequality measurement in the English-speaking world.

Following Gini (1912), Kendall and Stuart (1958) in their well-known book *Advanced Theory of Statistics* stated the Gini index as the half of Gini's relative mean difference because it was indeed an important statistical result at that time. There is no doubt that many generations of statisticians learned this result through the classical work of Kendall and Stuart.

Gini's absolute mean difference for a discrete distribution is defined as

$$\Delta = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|. \tag{14}$$

where y_i and y_j are the variates from the same distribution. The absolute

mean difference for a continuous distribution is defined similarly as the mean difference between any two variates of the same distributions:

$$\Delta = E|y_i - y_j| \tag{15}$$

where E is the mathematical expectation operator. The relative mean difference is defined as

$$\frac{\Delta}{\mu_y} = \frac{E|y_i - y_j|}{\mu_y}. \tag{16}$$

That is, the relative mean difference equals the absolute mean difference divided by the mean of the income distribution. In addition to the above result, Shalit and Yitzhaki (1984) also provided several alternative ways to express Gini' relative mean difference.

The Gini index is the one-half of Gini's relative mean difference

$$G = \frac{\Delta}{2\mu_y}. \tag{17}$$

The above expression is also written as

$$G = \frac{\frac{1}{n^2} \sum_{i=1}^n \sum_{i=1}^n \max(0, y_i - y_j)}{\mu_y} \tag{18}$$

because $\sum_{i=1}^n \sum_{j=1}^n |y_i - y_j| = 2 \sum_{i=1}^n \sum_{i=1}^n \max(0, y_i - y_j)$ [see Pyatt (1976)]. Anand (1983) showed that equation (17) is consistent with the geometric definition given in equation (5). For a discrete distribution, the absolute mean difference Δ can be rewritten as $\frac{2}{n^2} \sum_{i=1}^n \sum_{j \leq i} (y_i - y_i)$ and the Gini

index can be expressed as

$$
\begin{aligned}
G &= \frac{1}{n^2 \mu_y} \sum_{i=1}^{n} \sum_{j \leq i} (y_i - y_i) \\
&= \frac{1}{n^2 \mu_y} \sum_{i=1}^{n} \left(i y_i - \sum_{j=1}^{i} y_j \right).
\end{aligned}
\tag{19}
$$

This result is consistent with equation (8). In other words, Gini's mean difference approach is consistent with the geometric approach.

Following the tradition of Kendall and Stuart, Dorfman (1979) proposed a simple formula for the Gini index for the continuous income distribution, that is

$$
G = \frac{\Delta}{2\mu_y} = 1 - \frac{1}{\mu_y} \int_a^b (1 - F(y))^2 dy
\tag{20}
$$

where Gini's absolute mean difference for a continuous income distribution is given by equation (15). They also noted that Gastwirth (1972) proposed a similar formula without a proof which was attributed to Kendall and Stuart (1977) who also omitted the proof. This formula can be derived as follows. Because

$$
|y_i - y_j| = y_i + y_j - 2\min(y_i, y_j),
$$

Gini's absolute mean difference is written as

$$
\Delta = E|y_i - y_j| = 2\mu_y - 2E\min(y_i, y_j).
$$

To learn more about the term $E\min(y_i, y_j)$, it is necessary to know the probability for $\min(y_i, y_j)$, that is

$$
\Pr\left[\min(y_i, y_j) \leq y\right] = 1 - \Pr\left(y_i > y\right)\Pr\left(y_j > y\right) = 1 - (1 - F(y))^2.
$$

Incorporating this probability into Gini's absolute mean difference yields

$$\Delta = E|y_i - y_j| = 2\mu_y - 2\int_a^b yd\left(1 - (1 - F(y))^2\right) = 2\mu_y + 2\int_a^b yd\left(1 - F(y)\right)^2.$$

Substituting Δ into G yields

$$G = \frac{\Delta}{2\mu_y} = \frac{2\mu_y + 2\int_a^b yd\left(1 - F(y)\right)^2}{2\mu_y}.$$

Since $a = 0$ and b is finite, $a(1 - F(a))^2 = b(1 - F(b))^2 = 0$,

$$G = 1 + \frac{1}{\mu_y}\left(y(1 - F(y))^2|_a^b - \int_a^b (1 - F(y))^2\,dy\right) = 1 - \frac{1}{\mu_y}\int_a^b (1 - F(y))^2 dy.$$

The definition of the Gini index based on the relative mean difference has its root in the statistics. However, its computation could be complex. It is the covariance approach, which will be discussed below, that can facilitate the computation of the Gini index using the commonly used covariance procedure in most statistical software packages.

3.3   Covariance Approach

It was known that the Gini's absolute mean difference can be expressed as a function of the covariance between variate-values and ranks as noted in Stuart (1954, 1955):

$$\Delta = 4\int_a^b y\left(F(y) - \frac{1}{2}\right)f(y)dy. \tag{21}$$

17

But, finding of the link between this fact and the computation of the Gini index occurred much later.

In the context of discrete income distributions, Anand (1983) showed the Gini index can be computed by [7]

$$G = \frac{2cov(y_i, i)}{n\mu_y}; \tag{22}$$

that is, the Gini index can be expressed as a function of the covariance between incomes and their ranks. In the context of the discrete income distribution, Anand demonstrated how the definition of equation (22) is justified. He noted that the mean of the ranks is given by

$$\bar{i} = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2}$$

and the covariance is expressed as

$$
\begin{aligned}
cov(y_i, i) &= \tfrac{1}{n}\sum_{i=1}^{n}\left(y_i - \mu_y\right)\left(i - \bar{i}\right) \\
&= \tfrac{1}{n}\sum_{i=1}^{n} iy_i - \tfrac{n+1}{2}\mu_y.
\end{aligned}
$$

Thus, the Gini index can be written as

$$G = \frac{2cov(y_i, i)}{n\mu_y} = \frac{2}{n^2\mu_y}\sum_{i=1}^{n} iy_i - \frac{n+1}{n}$$

which is consistent with equation (11).

Independently, Lerman and Yitzhaki (1984) also reported the same result

---

[7]Based on the author's email communication with Sudhir Anand, the author learned that Sudhir Anand's thesis, which is the basis of Anand (1983), was completed in 1978.

for continuous income distributions. Yitzhaki (1982) noted that according to Lomnicki (1952) the Gini's absolute mean difference of the distribution F can be written as

$$\Delta = \int_a^b \int_a^b |x - y| f(x) f(y) dx dy = 2 \int_a^b F(x) \left[1 - F(x)\right] dx.$$

Lerman and Yitzhaki (1984)[8] showed that, by integration by parts, with $u = F(y)\left[1 - F(y)\right]$ and $v = y$, Δ can be written as[9]

$$\Delta = 4 \int_a^b y \left(F(y) - \frac{1}{2} \right) f(y) dy.$$

By transformation of variables, write $f(y)dy = dF$ and change from the bounds $[a, b]$ for y to the bounds $[0, 1]$ for F,

$$\Delta = 4 \int_0^1 y(F) \left(F - \frac{1}{2} \right) dF.$$

Note that F is uniformly distributed between 0 and 1 so that its mean is $\frac{1}{2}$. Thus, Gini's absolute mean difference is given by

$$\Delta = 4cov\left[y, F(y)\right]$$

and the Gini index is defined as $G = \frac{\Delta}{2\mu_y}$

$$G = \frac{2cov\left[y, F(y)\right]}{\mu_y}. \tag{23}$$

---

[8]This result in Charavarty (1990, p. 88) is called the Stuart(1954)-Lerman-Yitzhaki (1984) proposition.

[9]This is consistent to Stuart[1954, pp. 39-40, equations (13)–(15)].

In the context of continuous income distributions, Lambert (1989) used a slightly different approach to the same problem. He first noted that the Lorenz curve $L(p)$ has the following property:

$$L'(p) = \frac{dL(p)}{dp} = \frac{dL(p)/dy}{dp/dy} = \frac{yf(y)/\mu_y}{f(y)} = \frac{y}{\mu_y}. \tag{24}$$

Then using integration by parts he rewrote the Gini index from equation (13) as

$$
\begin{aligned}
G &= 1 - 2\int_0^1 L(p)dp \\
&= 2\int_0^1 pL'(p)dp - 1 \\
&= 2\int_a^b \frac{yF(y)}{\mu_y}f(y)dy - 1.
\end{aligned}
\tag{25}
$$

Since $cov[y, F(y)] = E[yF(y)] - E(y)E[F(y)]$ and $E[F(y)] = \frac{1}{2}$, the Gini index can be derived from equation (25) as

$$G = \frac{2\left[\int_a^b yF(y)f(y)dy - \frac{\mu_y}{2}\right]}{\mu_y} = \frac{2Cov[y, F(y)]}{\mu_y}. \tag{26}$$

Although the derivations of Anand (1983), Lerman and Yizhaki (1984) and Lambert (1989) differ, each result is a variant of the other. To compute the Gini index using Anand's approach, first obtain rank for each observation y_i, i ; then, compute the covariance between y_i and i, $cov(y_i, i)$. The resulting covariance must be divided by the number of the observations n, $cov(y_i, \frac{i}{n}) = \frac{1}{n}cov(y_i, i)$, since i/n terms are empirical cumulative distribution of $F(y)$. Finally, the Gini index is computed by $G = \frac{2}{n\mu_y}cov(y_i, i)$. This is consistent with the result of Lerman and Yitzhaki (1984) and Lambert (1989). This approach is also extended by Shalit (1985) so that a regression can be used

20

to compute the Gini index.

The advantage of the covariance approach is that the computation of the Gini index can be facilitated by using the covariance procedure in existing statistical software packages.

3.4   Matrix Form Approach

In the literature, the matrix form approach was proposed by Pyatt (1976) and Silber (1989) for decomposition purposes.[10]

Pyatt (1976) focused on equation (18) in which the Gini index is interpreted as a ratio of two terms: (a) the numerator $\frac{1}{n^2} \sum_{i=1}^{n} \sum_{i=1}^{n} \max(0, y_i - y_j)$ is the "average expected gain" to be expected by the population if each and every individual in the population is allowed to compare his or her income y_i with any other chosen individual's income y_j and to take y_j when y_j is greater than y_i and (b) the denominator is the mean income μ_y. Consider that the population can be divided into k subpopulation groups with group i has p_i proportion of the population. It is possible to express the average expected gain as

$$\sum_{i=1}^{k} \sum_{i=1}^{k} E(gain|i \rightarrow j) \Pr(i \rightarrow j) \tag{27}$$

where $\Pr(i \rightarrow j) = p_i p_j$ for all $i, j = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} p_i = 1$. Let $\mathbf{E}$ be a $k \times k$ matrix with elements being $E_{ij} = E(gain|i \rightarrow j)$. Stack p_i's into a $k \times 1$ column vector $\mathbf{p}$. Let the average income for group i be m_i. Stack m_i's into a $k \times 1$ column vector $\mathbf{m}$. Hence $\mathbf{m'p} = \sum_{i=1}^{k} m_i p_i = \mu_y$. Thus, the Gini

---

[10]Also see Yao (1999) for a spread sheet application.

index, corresponding to equation (18), can be defined as

$$G = (\mathbf{m'p})^{-1}\mathbf{p'Ep}. \tag{28}$$

Silber (1989) proposed another elegant approach for computing the Gini index. The derivation starts from the Gini index's definition:

$$G = \sum_{j=1}^{n} \widetilde{L}_j \left[\frac{(n-j)}{n} - \frac{(j-1)}{n} \right], \tag{29}$$

where $\widetilde{L}_i$ is the proportion of total income earned by the individual whose income has the ith rank in the income distribution with

$$\widetilde{L}_1 \geq \widetilde{L}_2 \geq \cdots \geq \widetilde{L}_j \cdots \geq \widetilde{L}_n.$$

(This means that the richest individual is ranked 1st while the poorest individual is ranked nth.) First of all, examine how this definition is linked to the previous ones. Note that $\widetilde{L}_j = \frac{y_j}{\sum_{j=1}^{n} y_j}$ implies that y_j's are arranged in non-increasing order while y_i's in our previous discussion are arranged in non-decreasing order. Therefore, it is useful to find out the system link between two index systems. It turns out that $i = n - j + 1$ and that equation (29) can be rewritten as

$$G = \sum_{i=1}^{n} \frac{y_i}{\sum_{i=1}^{n} y_i} \left[\frac{i-1}{n} - \frac{n-i}{n} \right],$$

which in turn equals

$$
\begin{aligned}
G & = \frac{1}{\sum_{i=1}^{n} y_i} \left[\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \left(\frac{2n - 2i + 1}{n} \right) y_i \right] \\
& = 1 - \frac{1}{n^2 \mu_y} \sum_{i=1}^{n} (2n - 2i + 1) y_i \\
& = \frac{n+1}{n} - \frac{2}{n^2 \mu_y} \sum_{i=1}^{n} (n + 1 - i) y_i.
\end{aligned}
$$

The last equality in the above is consistent with equation (7). Second, Equation (29) can be readily shown to be

$$
G = \sum_{i=1}^{n} \widetilde{L}_i \left[\sum_{j \geq i} \frac{1}{n} - \sum_{j \leq i} \frac{1}{n} \right]. \tag{30}
$$

Equation (30) is equivalent to equation (29) because the sum in the former is identical to the sum in the latter as shown below:

j	$\widetilde{L}_j \left[\frac{(n-j)}{n} - \frac{(j-1)}{n} \right]$	i	$\widetilde{L}_i \left[\sum_{j \geq i} \frac{1}{n} - \sum_{j \leq i} \frac{1}{n} \right]$
$j = 1$	$\widetilde{L}_1 \left[\frac{n-1}{n} - \frac{1-1}{n} \right] = \widetilde{L}_1 \left(\frac{n-1}{n} \right)$	$i = 1$	$\widetilde{L}_1 \left[\frac{n}{n} - \frac{1}{n} \right] = \widetilde{L}_1 \left(\frac{n-1}{n} \right)$
$j = 2$	$\widetilde{L}_2 \left[\frac{n-2}{n} - \frac{2-1}{n} \right] = \widetilde{L}_2 \left(\frac{n-3}{n} \right)$	$i = 2$	$\widetilde{L}_2 \left[\frac{n-1}{n} - \frac{2}{n} \right] = \widetilde{L}_2 \left(\frac{n-3}{n} \right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$j = n$	$\widetilde{L}_n \left[\frac{n-n}{n} - \frac{n-1}{n} \right] = \widetilde{L}_n \left(\frac{1-n}{n} \right)$	$i = n$	$\widetilde{L}_n \left[\frac{1}{n} - \frac{n}{n} \right] = \widetilde{L}_n \left(\frac{1-n}{n} \right)$

Third, equation (30) can be readily written compactly in matrix form as

$$
G = \mathbf{e}' \mathbf{G} \widetilde{\mathbf{L}} \tag{31}
$$

23

or

$$G = \begin{bmatrix} \dfrac{1}{n}, \dfrac{1}{n}, \dots, \dfrac{1}{n} \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 & \cdots & -1 \\ 1 & 0 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \widetilde{L}_1 \\ \widetilde{L}_2 \\ \vdots \\ \widetilde{L}_n \end{bmatrix}.$$

where $\mathbf{e}$ is a column vector of n elements of $1/n$, $\widetilde{\mathbf{L}}$ is a column vector of n elements being respectively equal to $\widetilde{L}_1, \widetilde{L}_2, \dots, \widetilde{L}_n$, and $\mathbf{G}$ is the $n \times n$ G-matrix whose elements G_{ij} are equal to -1 when $i < j$, to 1 when $i > j$, to 0 when $i = j$.

4   Social Welfare Implication

From the statistics point of view, the Gini index is a function of Gini's mean difference and hence it was initially, and still is, interpreted a measure of dispersion. Pyatt (1976), however, went a bit further and gave the Gini index an interpretation as the average gain to be expected, if each and every individual is allowed to compare his or her income with the income of another individual and to keep the income that is higher. But this interpretation is statistical in nature and more convenient for subpopulation group decomposition rather than for measuring social welfare (loss) due to inequality.

In fact, Dalton in his 1920 paper, following Pigou (1912, p. 24), had attempted to raise a minimum criterion for an inequality measure. It is now called Pigou-Dalton's principle of transfers. To establish this principle, he said:

24

"We have, first, what may be called the principle of transfers. ... we may safely say that, if there are only two income-receivers, and a transfer of income takes place from the richer to the poorer, inequality is diminished. There is, indeed, an obvious limiting condition. For the transfer must not be so large, as more than to reverse the relative positions of the two income-receivers, and it will produce its maximum result, that is to say, create equality. And we may safely go further and say that, however great the number of income-receivers and whatever the amount of their incomes, any transfer between any two of them, or, in general, any series of such transfers, subject to the above condition, will diminish inequality." (Dalton, 1920, p. 351)

Dalton (1920) also noted that the Gini index can be viewed as half of the Gini's relative mean difference. According to Dalton, as the relative mean difference satisfies the principle of transfers, the Gini index must satisfies the same principle and be judged as a desirable inequality measure.

Jenkins (1991), among others, used the total differential approach to evaluate whether the Gini index indeed satisfies the principle of transfers when the transfers are very small. To do so he assumed that the transfer is mean-preserving (i.e., μ_y is fixed) and that there is a transfer from to the richer individual i to the poorer individual j but this transfer will not change the fact the relative positions of the rich and the poor in the income distribution. Taking the total differential of equation (7) with respective to y_i and y_j yields

$$\partial G = (\partial G/\partial y_j)dy_j - (\partial G/\partial y_i)dy_i = \frac{2(j-i)}{n^2\mu_y}dy < 0 \qquad (32)$$

given that $dy_i = -dy_j$, $j < i$, and $|dy_i| = |dy_j| = dy$. Thus, the Gini index indeed satisfies the principle of transfer. That is, when the transfer occurs, the value of the Gini index will decrease.

Although the Gini index indeed satisfies the principle of transfers, there was little discussion about the social welfare implication of inequality measures including the Gini index after Dalton (1920). For example, Gini (1921) himself, in response to Dalton's work (1920), suggested that the measure of inequality (such as the one he proposed) was of income and wealth not of economic welfare.

The normative approach, which relates an inequality measure directly to an underlying social welfare function, appeared much later. Kolm (1969) advocated the use of social welfare function in measuring income inequality. Atkinson (1970) noted that the social welfare implication was particularly important when one came to select a summary statistics of income inequality. He wrote:

> "Firstly, the use of these measures often serves to obscure that fact that a complete ranking of distributions cannot be reached without fully specifying the form of the social welfare function. Secondly, examination of the social welfare functions implicit in these measures shows that in a number of cases they have properties which are unlikely to be acceptable, and in general these are no grounds for believing that they would accord with social values. For these reasons, I hope that these conventional measures will be rejected in favour of direct consideration of the properties that we should like the social welfare function to display."

(Atkinson, 1970, p. 262)

Sen (1973) also discussed this approach as a generalization of Atkinson's measure. It is Blackorby and Donaldson (1978) who examined the issue further in a systematical fashion, established the general results and applied them to inequality measures including the Gini index.

The way to link the Gini index to its underlying social welfare function is to define the Gini index in terms of the equally-distributed-equivalent-income (EDEI), or the representative income proposed by Atkinson (1970), Kolm (1969), and Sen (1973). Using this approach, an inequality measure, I, can be written as a function of the EDEI income, ξ, and the mean income, μ_y.

$$I = 1 - \frac{\xi}{\mu_y}. \tag{33}$$

If I is defined based on the Gini social welfare function, then I is denoted by I^G or, simply, G. Given this setup, if ξ is identical to μ_y, then I is zero. That is, there is no inequality in the income distribution from which ξ and μ_y are computed. If, on the other hand, ξ is less than μ_y (say, the former is only 70% of the latter), then I will be greater than zero but bounded by 1 (I will take a value of 0.3). That is, there is some degree of inequality. Of course, it is crucial to know how ξ is derived. Generally speaking, for a particular social welfare function or social evaluation function, an EDEI given to every individual could be viewed as identical in terms of social welfare to an actual income distribution.

To explain the idea further, let $W(\mathbf{y}) \equiv \phi(\overline{W}(\mathbf{y}))$ be a homothetic (ordinal) social welfare function of income with ϕ being an increasing function and

27

$\overline{W}$ being a linearly homogeneous function. Let $\mathbf{1}$ be a column vector of ones with an appropriate dimension. Then, $W(\xi \cdot \mathbf{1}) = W(\mathbf{y})$ or $\overline{W}(\xi \cdot \mathbf{1}) = \overline{W}(\mathbf{y})$. Given that $\overline{W}$ is positively linearly homogeneous, an EDEI is computed by $\xi = \frac{\overline{W}(\mathbf{y})}{\overline{W}(\mathbf{1})} = \Xi(\mathbf{y})$. The social welfare function, W, and the EDEI, Ξ, have an one-to-one corresponding relationship. The homotheticity of the social welfare function makes the indifference curves blowing-out or -in proportionally. Under this condition, the EDEI function, $\Xi(\mathbf{y})$, is linearly homogeneous in $\mathbf{y}$; that is, doubling $\mathbf{y}$ will also double ξ.

The above idea can be further explained by the case of an income distribution of two individuals. Figure 3 shows an actual income distribution denoted by the point $\mathbf{y}$ (that is, the income of the first individual, y_1, is 2 while that of the second individual, y_2, is 5). In Figure 3, income distributions on the 45 degree line from the origin represent perfect equality (that is, $y_1 = y_2$) of the two-person world. The social welfare function W has indifference curves like I_1 (the dotted convex curve) and I_2 (the solid convex curve). Since the actual income distribution $\mathbf{y}$ is on indifference curve I_2, EDEI, the point at which I_2 and the 45 degree line cross, will give the same level of social welfare. Thus, with a particular social welfare function W, an actual income distribution y and its corresponding EDEI are indifferent.

Figure 3. Social Welfare Function and EDEI
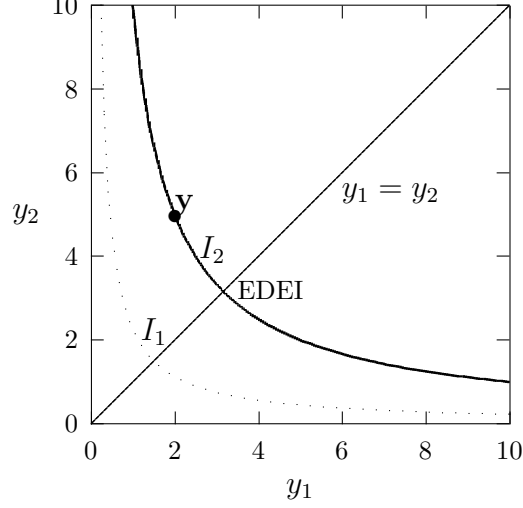
The Gini index can be defined with the Gini EDEI[11]

$$\Xi_G(\mathbf{y}) \equiv \frac{1}{n^2} \sum_{i=1}^{n}(2n - 2i + 1)y_i \tag{35}$$

that corresponds to the Gini social welfare function, $\overline{W}_G(\mathbf{y}) = \frac{1}{n^2} \sum_{i=1}^{n}(2n - 2i + 1)y_i$.[12] The Gini social welfare function attaches a higher weight to a lower level of income and vice versa. The weight is determined by the rank of an income rather than the size of the income. The Gini index can be defined

---

[11]If we sort the elements in $\mathbf{y}$ in non-increasing order and denote the new vector as $\widetilde{\mathbf{y}}$, then

$$\Xi_{\widetilde{G}}(\widetilde{\mathbf{y}}) = \frac{1}{n^2} \sum_{i=1}^{n}(2i - 1)\widetilde{y}_i \tag{34}$$

with $\Xi_G(\mathbf{y}) = \Xi_{\widetilde{G}}(\widetilde{\mathbf{y}})$. This is because $y_i = \widetilde{y}_{n-i+1}$ and $\widetilde{y}_i = y_{n-i+1}$.

[12]This is because

$$\overline{W}(\mathbf{1}) = \frac{1}{n^2} \sum_{i=1}^{n}(2n - 2i + 1) = 1.$$

in terms of the Gini EDEI and the mean income as[13]

$$I^G \equiv G \equiv 1 - \frac{\Xi_G(\mathbf{y})}{\mu_y} = 1 - \frac{1}{n^2 \mu_y} \sum_{i=1}^{n} (2n - 2i + 1) y_i. \qquad (37)$$

Note that while $G \equiv 1 - \frac{\Xi_G(\mathbf{y})}{\mu_y}$ taking values between 0 (perfect equality) and 1 (perfect inequality) measures inequality; $1 - G \equiv \frac{\Xi_G(\mathbf{y})}{\mu_y}$ also taking values between 0 (perfect inequality) and 1 (perfect equality) measures equality. The term $1 - G$ can be viewed as a measure of equality.

The social welfare implication of the Gini index can be better appreciated by the above formulation. For example, when the Gini index is .3, this means that according to the Gini social welfare function the inequality reduces the social welfare to the level that only 70% of the current total income if distributed equally among the population can achieve. The inequality can be reduced to zero if the current total income can be distributed equally among the population. When this understanding is established, a society which is inequality averse would prefer the income distribution with a lower Gini index value to the one with a higher Gini index value.

Of course, different inequality measures may have very different social welfare implications. Comparisons between the Gini index and other measures such as the coefficient of variation and Theil indices[14] were made based

---

[13]Alternatively,

$$\widetilde{G}(\widetilde{\mathbf{y}}) \equiv 1 - \frac{\Xi_{\widetilde{G}}(\widetilde{\mathbf{y}})}{\mu_y} = 1 - \frac{1}{n^2 \mu_y} \sum_{i=1}^{n} (2i - 1) \widetilde{y}_i, \qquad (36)$$

where $\widetilde{\mathbf{y}}$ has elements in non-increasing order. The two equations are identical because $\overline{\mathbf{y}} = \overline{\widetilde{\mathbf{y}}}$, $y_i = \widetilde{y}_{n-i+1}$, and $\widetilde{y}_i = y_{n-i+1}$. Note that in equations (37) and (36) $G(\mathbf{y}) = \widetilde{G}(\widetilde{\mathbf{y}})$ but $G(\cdot)$ and $\widetilde{G}(\cdot)$ have different functional forms and the elements in $\mathbf{y}$ and $\widetilde{\mathbf{y}}$ are sorted differently.

[14]See Theil (1967).

on the unit-simplex by Sen (1973, pp. 56–58, Figures 3.3 and 3.4), Blackorby and Donaldson (1978, p. 73, Figure 4), and Sen (1997, pp. 142–148, Figures A4.1 and A4.2).

5   Subgroup and Source Decomposition

The Gini index like any other indices is an aggregate summary statistics of income inequality and it can apply to a nation or regions/subgroups within the nation. However, researchers frequently wish to explore how inequality statistics in regions/subgroups contribute to the national inequality. Similarly, the Gini index is also applicable to each and every income component. How the inequalities in all income components contribute the overall income inequality is also of interest to social scientists. Chakravarty (1990) detailed subgroup and source decomposition in Section 2.6 of his book.

In general, the decomposition of an inequality measure can be conducted either on some kind partition of the population or on some division of the income. The former is referred to as subgroup decomposition and the latter is called as source decomposition. For subgroup decomposition, one wishes to see how subgroup inequality measures can be effectively related to the population inequality. By the same token, for source decomposition, one wishes to examine relationship between the aggregated inequality measure and the inequality measures from the components or sources.

When a researcher does source decomposition, he or she must apply the Gini index equation to each and every income components. Presenting these Gini indices themselves will not pose any problem. It is how to relate these

Gini indices for different income components to the Gini index of the aggregate incomes that becomes very challenging [see Silber (1993) and references therein]. Because the Gini index does not always permit a clear and explicit form of source decomposition, normally pseudo-Gini indices are used [see Chakravarty (1990)].

Subgroup decomposition is different. When the population is divided into K subgroups, the incomes for subgroup i also constitute a distribution and hence the Gini index for that subgroup, say G_i, can be computed. Similarly, the Gini index of the distribution of the mean incomes of these subgroups can also be computed and is called the between-group Gini index, or G_B. Now let the weight for each group i, b_i, be the product of the proportion of the population in subgroup i and the proportion of the aggregate income of the population in subgroup i. Then, the following relationship can be established:

$$G = G_B + \sum_{i=1}^{K} b_i G_i + R \tag{38}$$

where R is normally called the crossover term. Bhatacharya and Mahalanonis (1967) are perhaps the first economists working on the subgroup decomposition of the Gini index. Pyatt (1976) and Das and Parikh (1982) found the same result using the matrix form approach. Mookherjee and Shorrocks (1982) noted that the crossover term is an 'awkward interaction effect ... impossible to interpret with any precision." Shorrocks (1984) found that the Gini index is known to be decomposable (without the crossover term) when ranking incomes based on income sizes leads to the subgroup incomes cluster into subgroup income ranges without overlapping across subgroups.

Silber (1989), however, suggested that the crossover effect of subgroup

decomposition of the Gini index is not troublesome and, in fact, has a clear and intuitive interpretation. It measures the intensity of the permutation which occur when instead of ranking all the individual shares by decreasing (or increasing) income shares, one ranks them, firstly by decreasing (or increasing) value of the average income of the population subgroup to which they belong, and secondly, within each subgroup, by decreasing (or increasing) individual income share. Lambert and Aronson (1993) used a geometric approach to explain the crossover term and gave it a similar and good geometric interpretation.[15]

Yitzhaki and Lerman (1991) and Yitzhaki (1994) went a bit further suggesting that the crossover term would be a good indicator of income stratification. They noted that sociologists often find that the lack of the stratification dimension, when using the Theil index for subgroup decomposition, warrants some reconsideration when the issue of stratification is of interest to researchers.

To examine how 'stratification' among population subgroups to overall income inequality, Yitzhaki (1988, 1994) and Yizhaki and Lerman (1991) have also developed a 'pseudo-Gini coefficient', which mimics the Gini index in some respects. Their index decomposes inequality across population subgroups in a way which is different from the conventional way but is superficially similar: a crossover term, over and above between- and within-group terms, measures stratification in a precisely defined sense and has intuitively appealing properties [also see Lambert and Aronson (1993), 1224–1225].

Of course, some scholars have some reservations about using the Gini

---

[15]Sastry and Kelkar (1994) proposed a decomposition that is slight different from one proposed by Silber (1989).

index for subgroup decomposition. Cowell (1988) used specific examples to show that the Gini index could be a bad inequality measure because it allows the three things happen at the same time after some transfers have occurred: (a) mean income in every subgroup is constant; (b) inequality in every subgroup goes up; and (c) overall inequality for the population of all subgroups goes down. After a careful consideration, one can find that these can happen because within-subgroup, not across-subgroup, transfers can be made as such so that the conditions (a), (b), and (c) can be satisfied at the same time. Thus, this is unlikely to be a problem of the Gini index. A decreasing overall inequality reflects a social welfare gain in the population as a whole, while increasing inequality in every subgroup should be interpreted as a social welfare loss for each subgroup when evaluated independent of other subgroups of the population. Hence, when the social welfare implication is considered, the Gini index and its subgroup decomposition appear to have a clear interpretation.

6   Conclusion

Generally, it can be seen that over the past 80 years since Gini (1921) made his index known beyond Italy, our understanding of this index has improved and deepened substantially.

Now economists have learned that there are many ways to formulate and interpret the Gini index. That is, the Gini index can be computed based on the geometric approach, Gini's mean difference approach, covariance approach and matrix form approach. The Gini index is closely related to the

underlying Gini social welfare function. In addition, source and subgroup decomposition can assist the analysis of the inequality in income components and in subgroups. Now economists has a much better understanding of the crossover term when we deal with the subgroup decomposition of the Gini index.

The Gini index is also an important component of the Sen index of poverty intensity and the modified Sen index of poverty intensity [see, for example, Xu and Osberg (2001)]. The Gini index has been generalized to the S-Gini and E-Gini index to reflect various level of the inequality aversion [see, for example, Xu (2000) and the references therein]. Because the computation of the Gini index and other inequality measures is often made based on sample data. The statistical inferences based on the Gini index become more and more important. [see, for example, Osberg and Xu (2000), Xu (2000) and Biewen(2002) and the references therein]. Given the space of this survey, it is not possible to include these topics which will be left as another project in the future. Because the main focus of this survey is on the methodology, many good empirical studies using the Gini index are not cited here.

References

[1] Anand, Sudhir (1983). Inequality and Poverty in Malaysia: Measurement and Decomposition, New York: Oxford University Press.

[2] Atkinson, Anthony B. (1970). "On the Measurement of Inequality," Journal of Economic Theory, 2, 244–263.

[3] Atkinson, A. B. and F. Bourguignon (2000). "Introduction: Income Distribution and Economics," Handbook of Income Distribution, New York: Elvesier.

[4] Bhattacharya, N. and B. Mahalanobis (1967) "Regional Disparities in Household Consumption in India," Journal of the American Statistical Association, 62, 143–161.

[5] Biewen, Martin (2002). "Bootstrap Inference for Inequality, Mobility and Poverty measurement," Journal of Econometrics, 108, 317–342.

[6] Blackorby, Charles and David Donaldson (1978). "Measures of Relative Equality and Their Meaning in Terms of Social Welfare," Journal of Economic Theory, 18, 59–80.

[7] Chakravarty, S. R. (1990). Ethical Social Index Numbers, New York: Springer-Verlag.

[8] Cowell, Frank A. (1988). "Inequality Decomposition: Three Bad Measures," Bulletin of Economic Research, 40 (4) 309–311.

[9] Dalton, Hugh (1920). "Measurement of the Inequality of Income," The Economic Journal, 30, 348–361.

[10] Das, T., and A. Parikh (1982). "Decomposition of Inequality Measures and a Comparative Analysis," Empirical Economics, 7, 23–48.

[11] David, H. A. (1968). "Gini's Mean Difference Rediscovered," Biometrika, 55, 573–575.

[12] Dorfman, Robert (1979). "A Formula for the Gini Coefficient," Review of Economics and Statistics, 61, 146–149.

[13] Fei, John C. H. and Gustav Ranis (1974). "Income Inequality by Additive Factor Components," Economic Growth Center, Yale University.

[14] Fei, John C. H. , Gustav Ranis, and Shirley W. Y. Kuo (1979). "Growth and the Family Distribution of Income by Factor Components," Quarterly Journal of Economics, 92, 17–53.

[15] Gastwirth, Joseph L. (1972). "The Estimation of the Lorenz Curve and Gini Index," Review of Economics and Statistics, 54, 306–316.

[16] Gini, Corrado (1912). Variabilità e Mutabilità, Bologna: Tipografia di Paolo Cuppini.

[17] Gini, Corrado (1914). "Sulla Misura della Concertrazione e della Variabilità dei Caratteri," Atti del R. Istituto Veneto di SS.LL.AA., 73, 1203–1248.

[18] Gini, Corrado (1921). "Measurement of Inequality of Incomes," The Economic Journal, 31, 124–126.

[19] Jenkins, Stephen (1991). "The Measurement of Income Inequality," In Lars Osberg (ed.), Economic Inequality and Poverty: International Perspectives, Armonk, New York: M.E. Sharpe, Publishers.

[20] Kendall, Maurice G., and Alan Stuart (1958). The Advanced Theory of Statistics, Vol. 1, Distribution Theory, 1st ed. New York: Hafner Publishing Company.

[21] Kendall, Maurice G., and Alan Stuart (1977). The Advanced Theory of Statistics, Vol. 1, Distribution Theory, 4th ed. London: Charles Griffin, 48–51, 54.

[22] Kolm, S. Ch. (1969). "The Optimal Production of Social Justice," in J. Margolis and H. Guitton (eds.) Public Economics, New York/London: Macmillan.

[23] Lambert, Peter J. (1989). The Distribution and Redistribution of Income: A Mathematical Analysis, Cambridge, Massachusetts: Basil Blackwell Inc.

[24] Lambert, Peter J. and J. Richard Aronson (1993). "Inequality Decomposition Analysis and the Gini Coefficient Revisited," The Economic Journal, 103, 1221–1227.

[25] Lerman, Robert I. and Shlomo Yitzhaki (1984). "A Note on the Calculation and Interpretation of the Gini Index," Economics Letters, 15, 363–368.

[26] Lomnicki, Z. A. (1952). "The Standard Error of Gini's Mean Difference," Annals of Mathematical Statistics, 23, 635–637.

[27] Money, Leo George Chiozza (1905). Riches and Poverty, London: Methuen & Co. Ltd.

[28] Mookherjee, D. and A. F. Shorrocks (1982). "A Decomposition Analysis of the Trend in UK Income Inequality," The Economic Journal, 92, 886–902.

[29] Osberg, Lars and Kuan Xu (2000). "International Comparison of Poverty Intensity: Index Decomposition and Bootstrap Inference," Journal of Human Resources , 35, 51–81.

[30] Pigou, A. C. (1912). Wealth and Welfare, London: Mcmillan.

[31] Pyatt, Graham (1976). "On the Interpretation and Disaggregation of Gini Coefficient," The Economic Journal, 86, 243–255.

[32] Rao, V. M. (1969). "Two Decompositions of Concentration Ratio," Journal of the Royal Statistical Society, Series A, 132, 418–425.

[33] Sastry, D. V. S. and Ujwala. R. Kelkar (1994). "Note on the Decomposition of the Gini Inequality," Review of Economics and Statistics, 76 (3), 584–586.

[34] Sen, Amartya K. (1973). On Economic Inequality, Oxford: Clarendon Press.

[35] Sen, Amartya K. (1997). On Economic Inequality, Expanded edition with a substantial annexe by James E. Foster and Amartya Sen, Oxford: Clarendon Press.

[36] Shalit, Haim (1985). "Calculating the Gini Index of Inequality for Individual Data," Oxford Bulletin of Economics and Statistics, 47 (2), 185–189.

[37] Shalit, Haim and Shlomo Yitzhaki (1984). "Mean Gini, Portfolio Theory, and the Pricing of Risky Assets," Journal of Finance, 39, 1449–1468.

[38] Shorrocks, Anthony F. (1984). "Inequality Decomposition by Population Subgroup," Econometrica, 52 (6), 1369–1385.

[39] Silber, Jacques (1989). "Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality," Review of Economics and Statistics, 71, 107–115.

[40] Silber, Jacques (1993). "Inequality Decomposition by Income Source: A Note," Review of Economics and Statistics, 75, 545–547.

[41] Silber, Jacques (1999). Handbook of Income Inequality Measurement, Boston : Kluwer Academic.

[42] Stuart, A. (1954). "The Correlation between Variate-values and Ranks in Samples from a Continuous Distribution," British Journal of Statistical Psychology, Vol. 7, Part 1, 37–44.

[43] Stuart, A. (1955). "The Correlation between Variate-values and Ranks in Samples from Distributions Having No Variance," British Journal of Statistical Psychology, Vol. 8, Part 1, 25–27.

[44] Theil, Henri (1967). Economics and Information Theory, Chicago: Rand McNally, 1967.

[45] Xu, Kuan (2000). "Inference for Generalized Gini Indices Using the Iterated Bootstrap Method," Journal of Business and Economics Statistics, 18, 223–227.

[46] Xu, Kuan and Lars Osberg (2002). "The Social Welfare Implications, Decomposability, and Geometry of the Sen Family of Poverty Indices," Canadian Journal of Economics, 35, 138–152.

[47] Yao, Shujie (1999). "On the Decomposition of Gini Coefficients by Population CLass and Income Source: A Spreadsheet Approach and Application," Applied Economics, 31, 1249–1264.

[48] Yitzhaki, Shlomo (1982). "Stochastic Dominance, Mean Variance and Gini's Mean Difference," American Economic Reivew, 72, 178–185.

[49] Yitzhaki, Shlomo (1988). "On Strtification and Inequality in Isreal," Bank of Israel Review, 63, 36–51.

[50] Yitzhaki, Shlomo (1994). "Economic Distance and Overlapping of Distribution," Journal of Economics, 61, 147–159.

[51] Yitzhaki, Shlomo (1998). "More Than A Dozen Alternative Ways of Spelling Gini," Research in Economic Inequality, Vol. 8, 13–30.

[52] Yitzhaki, Shlomo and Robert I. Lerman (1991). "Income Stratification and Income Inequality," Review of Income and Wealth, 37 (3), 313–329.