ABSTRACT

The Use of the Gini Ratio in Measuring Distributional Impacts

Charles Riemenschneider

Economists often use the Gini Ratio for measuring inequality in income or program benefit distributions without fully understanding the implications and limitations of this statistic. This paper delineates some of the limitations in the use of the Gini Ratio as it is usually calculated. It also examines several new techniques for calculating the Gini Ratio which mitigate some of these shortcomings.

THE USE OF THE GINI RATIO IN MEASURING DISTRIBUTIONAL IMPACTS

by

Charles H. Riemenschneider

A Research Report

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Agricultural Economics

1976

TABLE OF CONTENTS

THE USE OF THE GINI RATIO IN MEASURING DISTRIBUTIONAL IMPACTS*

INTRODUCTION

Over the past fifteen years equity in the distribution of income has become an important issue in the political arena. Along with the rise of this issue has come the desire to measure inequity. Measurement of inequity is closely tied to the desire for social indicators. Henriot (p. 3) defines social indicators as "quantitative data that serve as measures of socially important conditions of society." There would be little argument that inequities in the distribution of income or program benefits are socially important conditions of society. As Bauer (p. 20) points out there are many areas where policies have long been made without any yardstick for measuring whether, as a result of policy actions, conditions are getting better or worse.

Many economists have used the Lorenz Curve and the related Gini Ratio to provide this yardstick for measuring inequity in income and the Gini Ratio appears to have many of the characteristics which are desirable in a social indicator: it is a single statistic which describes the distribution, conceptually it is rather easy to understand, computationally it is simple, and it appears to be relatively value free. But on more careful examination some of these characteristics do not seem to hold well and other problems become apparent in the use of the Gini Ratio. The problem that arises is not so much in the use of the Gini Ratio per se, but rather its use without understanding its limitations as a measure of inequity in income distributions.

The distinction between equity and equality needs clarification before proceeding any further with the discussion of the Gini Ratio. While most economists recognize the distinction between equity and equality, discussions of income distribution can still be found in current literature where this distinction is not clearly made, e.g. Sen. As Bronfenbrenner (p. 10) points out the equity of a given distribution of income is fundamentally a matter of ethical judgment and thus is a subjective concept. On the other hand, the equality of an income distribution is for the most part a mathematical or statistical matter and thus is basically objective. Thurow (p. 22) makes a similar distinction. Equity with respect to income in his terms deals with the determination of a just distribution of income. Justice does not necessarily require equality. Due to the subjective nature of equity a universally acceptable method of measuring inequity is not available to the economist. So in comparing income distributions, measures of inequality, such as the Gini Ratio, are often used for lack of a better measure because it is generally possible to measure inequality objectively while inequity is not as readily measurable in an objective sense.

This creates a problem in that usually it is the equity of a given distribution of income that is important from society's viewpoint and not the equality of that distribution. Both Thurow and Bronfenbrenner arrive at the same conclusions: 1) that the connection between these two concepts, equity and equality, is not readily apparent and 2) that a general specification of an equitable distribution is not possible. This places some limitations on the use of the Gini Ratio, which will be discussed in the later sections of this paper.

The general methodological question being raised here is that of how to describe a distribution by a single statistic without losing needed information. A statistical description of a distribution usually contains three major pieces of information: central tendency, dispersion, and symmetry. Each of these dimensions has potential economic implications in the context of an income distribution and can provide relevant information for policy decisions. However, the relative importance of any one of these dimensions depends on the questions that need to be answered before a policy can be formulated or a decision made. Thus the appropriateness of a statistic which purports to describe a distribution will depend upon its capacity to capture and transmit information on the dimensions needed for decisions. As of yet no single statistic has been devised which conveys sufficient information about all three major dimensions, the central tendency, dispersion, and symmetry, which the statistician would use to describe a distribution.

From a methodological point of view there is one other aspect of the measurement of inequality that also requires comment before discussing the Gini Ratio. This is the difference between so called normative and positive measures of inequality. Positive or objective measures of income inequality tend to be those that attempt to measure inequality in some objective sense by using statistical measures of income variation; for instance the Gini Ratio and the coefficient of variation are usually considered as members of this class. Normative measures usually try to measure inequality on the basis of some value notion of social welfare in which social welfare decreases as inequality increases. Atkinson's Index and Dalton's Index are often listed as members of this class.(Sen.) This dichotomy of normative and objective measures is not very useful for, in a strictly logical sense, any concept of inequality has both normative and objectives aspects and thus so do measures of inequality. In the past the Gini Ratio has been characterized as being descriptive with no normative implications (Bonnen, p. 465). This is true in the sense that the Gini Ratio is a measure of inequality and makes no reference to any value judgment as to what is equitable. However, even as a measure of inequality the Gini Ratio does have some normative implications as any specific statistic must. It implies a specific type of inequality since it has as its basis a comparison of the observed distribution of income with Lorenz's line of equality. This Lorenzian equality implies equal incomes for everyone in a given time period. There are other ways of measuring inequality.

Another type of equality that one might wish to consider in determining an equitable income distribution is that of equal incomes for persons at the same stage in the life cycle. As Paglin (p. 598) shows, the Lorenzian concept of equality when used with annual income data neglects any life cycle differences in income and treats these as a type of inequality. Thus, in using the Gini Ratio to compare income distributions, one is not totally avoiding all subjectivity in that there is an implicit value judgment made simply by using the Gini Ratio. The so called normative measures of inequality, on the other hand, tend to be based on only one normative aspect which is used for the comparison of distributions is usually dependent on some objective aspect of inequality (Sen, p. 3). Thus the so called normative measures of inequality are not wholly normative, just as the so called objective measures are not wholly objective. So to claim that one measure of inequality is objective while another is normative really makes little sense.

While it is beyond the scope of this paper to discuss other measures of inequality, one may find it necessary to consider different measures of inequality before deciding on the one best suited for his purpose. Sen and

Champernowne each discuss other measures of inequality in addition to the Gini Ratio such as the coefficient of variation, the standard deviation of logarithms, Theil's entropy coefficient, Dalton's Index and Atkinson's Index. It is important to note that these measures can yield different rankings from the same group of distributions so that it is necessary to be familiar with the sensitivities and shortcomings of the measure chosen.

The use of the Gini Ratio in much of the agricultural economics literature has dealt with comparing the distribution of benefits from commodity programs or the distribution of farm income, e.g. Bonnen, Boyne, Hill, Reinsel. The examples that are used in the remainder of this paper are based on these types of distributions.

The purpose of the following sections of this paper is to delineate some of the limitations of the Gini Ratio as it is usually calculated and used while at the same time examining several new techniques for calculating the Gini Ratio which attempt to deal with some of these limitations.

II. Calculation of the Gini Ratio

Conceptually the Gini Ratio is easier to comprehend if one first looks at the Lorenz Curve. A Lorenz Curve is derived by plotting the cumulative fraction of total income against the cumulative fraction of units receiving this income where the income receiving units are arranged from poorest to richest units or income classes. A typical Lorenz Curve is shown in Figure 1. The Lorenz Curve will coincide with the Line of Equality if every unit has the same income. In the absence of complete equality, the Lorenz Curve will lie below the diagonal, as in Figure 1. The Lorenz Curve lies on or below the diagonal with a slope that does not decrease as one moves from lower to higher income classes of the population.[5] Since the Lorenz Curve is plotted in cumulative fractions this curve must run from one corner of the unit square to the other. The Gini Index of Concentration, or more commonly the Gini Ratio, can be derived from the Lorenz Curve and is the proportion of the total area under the diagonal that is between the Lorenz Curve and the diagonal. From Figure 1 it can be expressed as follows:

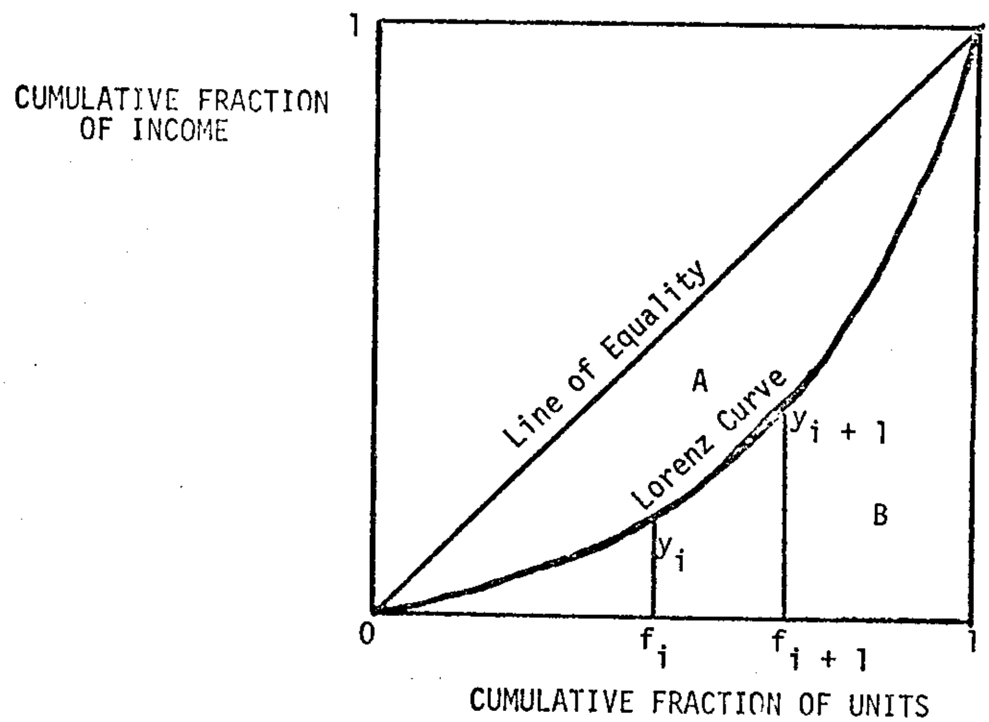
$$1)\quad \text{Gini Ratio} = \frac{A}{A+B} = \frac{\text{Area between the curve and diagonal}}{\text{Area under the diagonal}}$$

By convention the area of the complete square is defined to be 1 so the area under the diagonal is 1/2. Thus the expression can be rewritten as follows:

$$\text{Gini Ratio} = \frac{1/2 - \text{Area under curve}}{1/2}$$

$$2)\quad = 1 - 2\,(\text{Area under Curve})$$

Figure 1


1
CUMULATIVE FRACTION
OF INCOME
Line of Equality
A
Lorenz Curve
y_{i+1}
B
y_i
0
f_i
f_{i+1}
1
CUMULATIVE FRACTION OF UNITS

The calculation of the Gini Ratio can be done in at least two ways, one for grouped data and another for ungrouped data. The traditional method for grouped data uses a linear approximation of the curve, in that it estimates the area under the curve by drawing straight lines between the data points and then taking the area of each polygon determined from this and summing these areas to get the area under the curve (Morgan). The area under any of these line segments can be expressed as follows:

$$(f_{1+1} - f_i) \frac{(y_i + y_{i+1})}{2}$$

Where:

f_i = cumulative fraction of units

y_i = cumulative fraction of income

Summed over all intervals the area under the curve is approximated by:

$$\sum_{i=1}^{K} (f_{i+1} - f_i) \frac{(y_i + y_{i+1})}{2}$$

Where: K = number of intervals (groups)

Substituting this into the equation for the Gini Ratio above, a formula for estimating the Gini Ratio can be obtained.

$$\text{Gini Ratio} = 1 - 2 \sum_{i=1}^{K} (f_{i+1} - f_i) \frac{(y_i + y_{i+1})}{2}$$

$$= 1 - \sum_{i=1}^{K} (f_{i+1} - f_i) (y_i + y_{i+1})$$

(See Morgan, Miller, Bonnen, and Hill for similar presentations.)

For estimating the Gini Ratio from ungrouped data, more than one method is available. The Gini Ratio can be shown to equal one-half the relative mean difference which is the arithmetic mean of the absolute values of the differences between all pairs of individual incomes, i.e.

4) Gini Ratio $= \frac{1}{2N^2\mu} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| y_i - y_j \right|$

Where N = total number of individuals

y_i & y_j = the income of individual i and j respectively

μ = the arithmetic mean income

Through manipulation of the above expression the Gini Ratio can also be expressed as follows:

5) Gini Ratio $= 1 - \frac{1}{N^2\mu} \sum_{i=1}^{N} \sum_{j=1}^{N} \min (y_i, y_j)$

or

6) Gini Ratio $= 1 + \frac{1}{N} - \frac{2}{N^2\mu} [y_1 + 2y_2 + \ldots N y_N]$

for $y_1 \geq y_2 \geq \ldots \geq y_N$

(This presentation is based on Sen, p. 29-31).

If income is distributed with perfect equality, i.e. everyone with the same income, then the Gini Ratio is zero and if there is perfect inequality, i.e. one has all the income, the Gini Ratio is one. Thus the Gini Ratio is bounded by zero and one with a lower Gini Ratio implying a more equal distribution of income.

III. Limitations In the Use of the Gini Ratio

The Gini Ratio is only a relative measure of income distribution, it says nothing about absolute income levels. So the income of each member of the population can double without affecting the Gini Ratio. This leads to difficulties in comparing different state or county income distributions or different distributions of benefits from public programs. For instance in the 1973 U.S. cotton program the average set-aside payment per farm in the Southeastern region of the U.S. was $2,231 and the Gini Ratio was .678 while in the Western region the average payment per farm was $12,295 while the Gini Ratio was .679. Thus in this case the mean absolute payment was over 450% higher in the West while the Gini Ratios were nearly identical. The Gini Ratio is mean independent, as the example above shows, and therefore to make value judgments, concerning the equity of different distributions, based on the Gini Ratio alone, may not account for some of the most relevant information.

A problem of shifting universes can arise in comparing some public program benefit distributions over time. Between 1970 and 1971 the cotton program was changed to eliminate small farm payments, because of this many small farms in the Southeast dropped out of the program in 1971. Total Southeast cotton program participation declined from 190,961 farms in 1970 to 77,703 farms in 1971. It is safe to assume that this reduction in total participants arose mostly from the reduction of small farms since in 1970 there were 146,059 small farm participants in this region with 100,913 receiving payments without planting any cotton. Given this information, it would be easy to conclude that the amount of inequality had probably increased between 1970 and 1971, since there were fewer small farms in the program and there was essentially no change in total payments. A conclusion from this for anyone not familiar with the Gini

Ratio might be that the Gini Ratio should increase. But this was not the case. The Gini Ratio for this distribution of payments from the cotton program in the Southeast, actually declined from .717 in 1970 to .680 in 1971. This occurs because the populations on which the Gini Ratio is based had changed. Since the Gini Ratio is only a relative measure and the range of income over which beneficiaries were distributed had actually decreased, the Gini Ratio also declined. Thus using the Gini Ratio to compare distributions over time where the universe has shifted significantly often does not account for phenomena which the researcher may consider significant. This analysis also raised questions about making comparisons between farm income distributions over time, particularly during the time period when farm population was declining rapidly.

Richard Benson (p. 446) discussed what he called a "bias" in the Gini Ratio caused by aggregation. This "bias" is not really a statistical bias at all but only a failure to understand one of the limitations in the use of the Gini Ratio, that being the problem of a shifting universe. His conclusions though are essentially correct in that because of this problem a Gini Ratio from one state or region alone does not imply anything about the national Gini Ratio or even the effect of combining different state distributions.

A third practical problem that is often encountered when using the linear approximation method comes about due to the lack of availability of ungrouped data or the availability of data grouped in too few cells. Since this method for calculating the Gini Ratio from grouped data relies on a linear approximation of the curve, as the number of cells decreases, the line segments that approximate the curve get longer and this seriously biases the approximation of the curve. By reducing the number of cells into which data is grouped, ceteris paribus, the area under the curve increases, hence the Gini Ratio decreases. By aggregating the same data into six cells instead of eleven for the distribution of set-aside payments for the United States under the 1971 cotton program, the Gini Ratio decreased from .680 to .656. The reason for this is that the Gini Ratio, when calculated in the normal method for grouped data, fails to consider the dispersion of income or benefits within each cell. Rather it is assumed that each unit within the cell has the mean income of that cell. The Gini Ratio is the sum of two components, the dispersion of the cells about the mean income of the sample and the dispersion of income within each cell. So that the failure to consider this second component in the linear approximation method causes the Gini Ratio to be underestimated. Since most farm income data is grouped according to the USDA's economic classes of farms, one is often faced with data grouped into less than 8 cells.

Another shortcoming encountered in the use of the Gini Ratio is that it attempts to describe an entire distribution with only one statistic. The distribution of income has many facets and any attempt to describe this distribution with only one number necessarily must leave out much information. The same Gini Ratio can be obtained from very different distributions, so the Gini Ratio gives little information about the nature of the skewness of the distribution or Lorenz Curve. For example, the same Gini Ratios can be obtained from Lorenz Curves that cross and therefore these Gini Ratios may be based on quite different income distributions. Thus there is no way to determine which income classes are getting relatively higher or lower income shares by looking at the Gini Ratio alone.

The sensitivity of the Gini Ratio to income transfers can cause problems in its use. From a practical standpoint, differences in sensitivity to income transfers among different inequality measures can cause different conclusions to be reached concerning the effect of certain types of income transfers. In any case one would normally wish to have an index which shows Pigou-Dalton efficiency, i.e. a transfer of income from one person to any poorer person would always decrease the index of inequality and vice versa (Champernowne, p. 789). The Gini Ratio has Pigou-Dalton efficiency but it is not equally sensitive to all types of income transfers. This can be seen by looking at equation 6 for calculating the Gini Ratio from ungrouped data. The Gini Ratio is based on a weighted sum of different individual's income levels with the weights being determined by the rank ordering of the individuals according to income levels. Sensitivity to income transfers is not only determined by the income levels of the individuals involved in the transfer but also it depends on the number of individuals between the income levels of the individuals involved in giving and receiving the transfer (Sen, p. 32-33).

Several problems encountered when comparing different Gini Ratios have already been identified. In addition there is the problem of determining the statistical difference between Gini Ratios. In the case of Gini Ratios based on the distribution of an entire population, the Gini Ratio is a population parameter and any difference between parameters is statistically significant. In the usual application one is forced to calculate Gini Ratios from a sample of the population. To test for significant differences between Gini Ratios in this case requires a measure of variance of the Gini Ratio. There are no known formulas for describing the variance of the sample distribution, as a total distribution, or for a Gini Ratio based on a sample. Conceptually one might use multiple samples or random subsample replication[1] techniques on each sample if the samples are of sufficient size. When limited to secondary data multiple samples are usually impossible. Even when comparing Gini Ratios from entire populations, such as all participants in a farm program, problems arise when grouped data is used because the Gini Ratio is understated when using the linear approximation method of estimation. This method of estimation does not allow one to measure the extent to which the Gini Ratio is understated so that a direct comparison of Gini Ratios is inappropriate and the means for a test of significance is still lacking.

Income can be measured in terms of different income receiving units, for instance dollars per farmily or dollars per person. The income receiving unit chosen is very important in determining the Gini Ratio. David and Morgan and more recently Paglin (p. 603) have shown that as the level of aggregation of the income receiving unit increases (e.g. from persons in families to households) the Gini Ratio will increase for the same total income. Many of the efforts to measure the distribution of benefits from farm programs have used the "ASCS farm" as the income receiving unit. The ASCS definition of farm does not coincide with and tends to produce more income units than the USDA's definition of "farm" or "farmer." So if one is interested in the distribution of program benefits among farmers, the Gini Ratio based on the ASCS farm unit probably underestimates the inequality of the benefit distribution among farmers because the level of aggregation of income receiving units is greater.

This measurement unit problem also causes further problems in making comparisons across populations or through time. If the other obstacles can be overcome, one must still be sure that the measurement unit for those receiving income or program benefits is the same in order to make comparisons meaningful.

One possible limitation, already mentioned in the introduction, deals with the definition of equality that is implied by the Lorenz Curve and Gini Ratio. Lorenz's Line of Equality implies equal incomes for all ages or any other cross classification; in measuring the distribution of benefits from commodity programs most authors have used the Line of Equality to imply equal benefits for each farm regardless of size. Whether one's own definition of equity is compatible with this definition of equality should be a significant factor in determining the usefulness of the Gini Ratio for many applications.

IV. Techniques to Deal With These Limitations

. It is impossible to overcome many of the limitations in the use of the Gini Ratio. Many are inherent in the definition of this statistic. If one were to attempt to make the Gini Ratio less of a relative measure or to change its sensitivity to income transfers, it would be necessary to redefine the Gini Ratio. Nonetheless, the need to understand these limitations remains particularly with respect to making welfare or policy judgments based on this statistic.

Recently techniques have been suggested in the literature which reduce the effect of some of the shortcomings described in the previous section. As Gastwirth points out the linear approximation method for estimating the Gini Ratio from grouped data is calculated from the mean of each group (cell) and ignores the distribution of observations around the mean within the cells. To deal with this he recommends that instead of using one value for the Gini Ratio, it is better to calculate an upper bound and a lower bound thus in a fashion accounting for the dispersion within cells. The lower bound is calculated using the linear approximation method defined earlier. This yields the lower bound because it assumes that all incomes within a cell are equal to the mean income of the cell, so in this case there is no dispersion within the cell. The upper bound is calculated bv maximizing the spread within the cell. The upper bound is defined as follows:

$GU = GL + D$

where GL = the lower bound of the Gini Ratio and

$$D = \frac{1}{\mu} \sum_{i=1}^{n+1} (f_i - f_{i-1})^2 \, (a_i - \mu_i) \, \frac{(\mu_i - a_{i-1})}{(a_i - a_{i-1})}$$

where $n + 1$ = the number of cells

a_i = the upper endpoint of the cell i in dollars

$(f_i - f_{i-1})$ = the fraction of units in cell i

μ_i = the mean income in cell i in dollars

μ = the overall mean income of the population in dollars

(Based on Mehran and Gastwirth)

This method provides one intuitive way in which Gini Ratios estimated by linear approximation from grouped data can be compared. It would seem unreasonable to conclude that two Gini Ratios were significantly different if the lower or upper bound of one fell between the lower and upper bounds of the other. Gastwirth (p. 310) suggests a use of these bounds in another way, they can be used to design the grouping intervals or cell sizes necessary to obtain the desired degree of accuracy. Usually the number of groups required to get close bounds is rather large (about 20) because most group boundaries are not chosen to minimize grouping correction (Gastwirth, p. 310).

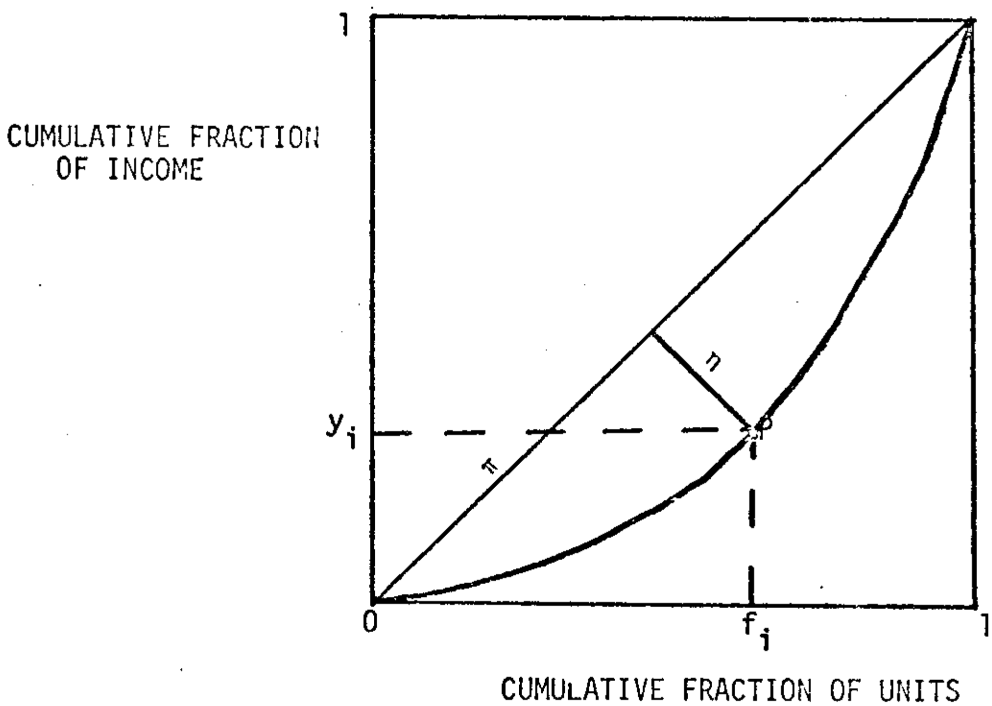
This method also gives one some idea of the possibilities of errors in the use of the Gini Ratio, as calculated by the linear approximation method, caused by reducing the number of cells from which this statistic is calculated. The lower bound is reduced by reducing the number of cells and the upper bound is increased. To obtain the upper bound for the Gini Ratio calculated from a distribution of payments from a farm commodity program, it is necessary to place dollar values on the endpoints of each cell in order to use the formula stated above because with this type of data the endpoints of each cell are normally given in acres not dollars. This does not give an exact upper bound, as Gastwirth's method does when used in cases where the classes are divided by income, because the dollar values of the acreage endpoints are only approximate. The upper bound for the 1971 cotton program set-aside payment distribution for the United States is .6949 and the lower bound is .6808. When the same data is grouped in six cells the upper and lower bounds are .7075 and .6563 respectively.[2]

Kakwani and Podder (1976) have devised another method for estimating the Gini Ratio from grouped observations by estimating the equation of the Lorenz Curve and then using the integral of this curve to estimate the area under the curve and thus the Gini Ratio. In the past many researchers have attempted to fit some well known density function such as the Pareto or lognormal to existing income distributions and then use this to obtain an estimate of the Gini Ratio. These efforts have not been an overwhelming success in most cases because such functions have rarely been a good fit for the actual data. Where Kakwani and Podder's approach differs is that instead of trying to fit some predetermined function to the data they estimated the Lorenz Curve from the data using regression techniques. To get a good fit for actual data the functional form of the Lorenz Curve must be specified correctly. In their original article Kakwani and Podder (1973) used a functional form which when applied to other distributions was not really adequate.[3]

However, in a more recent article Kakwani and Podder (1976) develop a new coordinate system for the Lorenz Curve, then estimate the curve using the new coordinates and a new functional form. To derive the new coordinates let P be any point on the Lorenz Curve with coordinates (f_i, y_i) then the new coordinates π and η (see Figure 2) are given by:

$$\pi_i = \frac{1}{\sqrt{2}} (f_i + y_i) \qquad \eta_i = \frac{1}{\sqrt{2}} (f_i - y_i)$$

Figure 2


1
CUMULATIVE FRACTION
OF INCOME
η
y_i
π
P
0
f_i
1
CUMULATIVE FRACTION OF UNITS

Then a Lorenz Curve of the form $\eta = a\,\pi^{\alpha}(\sqrt{2} - \pi)^{\beta}$ can be estimated, where a, α and β all are greater than zero. In actual estimation of the Lorenz Curves for the distribution of payments from the cotton program ordinary least squares was used to estimate the equation:

$$\log \eta_i = \log a + \alpha \log \pi_i + \beta \log (\sqrt{2} - \pi_i) + e$$

Where the values for η_i and π_i are obtained from the values of f_i and y_i that are used in the usual linear approximation method. Once the regression coefficients (a, α, β) are known then the Gini Ratio for this Lorenz Curve can be obtained from:

$$\text{Gini Ratio} = 2\int_0^{\sqrt{2}} a\,\pi^{\alpha}(\sqrt{2} - \pi)^{\beta}\, d\pi$$

$$= 2\, a\, (\sqrt{2})^{1 + \alpha + \beta}\; B(1 + \alpha, 1 + \beta)$$

Where $B(1 + \alpha, 1 + \beta)$ is the BETA Function

This method of estimation gave very accurate results for all the distributions used to test it in this research.[4] From a mathematical standpoint this method is superior to the usual method in that it gives a curvilinear approximation of the Lorenz Curve instead of a linear approximation. The method also gives more information about the distribution. The regression coefficients give three numbers with which to describe the distribution and these can be used to obtain a measure of skewness of the distribution and the Lorenz Curve. If $\alpha > \beta$ it is skewed toward (0,0), if $\alpha = \beta$ the Lorenz Curve symmetrical and if $\alpha < \beta$ it is skewed toward (1,1). So if $\alpha > \beta$ the lower and higher income groups get the highest income shares. If $\alpha < \beta$ then the middle income groups get the highest income shares.

One important shortcoming of the linear approximation method that Kakwani and Podder's method seems to mitigate is the bias caused by reducing the number of cells into which the data is grouped. As was stated earlier when the number of cells used in calculating the Gini Ratio of the 1971 cotton program set-aside payments distribution for the United States was reduced from eleven cells to six cells the Gini Ratio declined from .6808 to .6563 using the linear approximation method. But using Kakwani and Podder's method the Gini Ratio only declines from .6868 to .6865 when the number of cells is reduced from eleven to six. Similar results were obtained for other years. These latter estimates of the Gini Ratio using Kakwani and Podder's method for 1971 cotton program data both fall within the upper and lower bounds calculated by Gastwirth's technique and intuitively do not seem significantly different. It appears that Kakwani and Podder's method can be used with more success than the usual linear approximation method for grouped data when problems arise due to data grouped in too few cells. Even without empirical justification one would expect Kakwani and Podder's method to reduce the bias in the Gini Ratio due to aggregation into a limited number of cells since a curvilinear approximation of the Lorenz Curve to some extent takes into account the dispersion within cells.

Kakwani and Podder's method has one drawback, that being that it is more costly than the linear approximation method. The relatively simple arithmetic calculation of the Gini Ratio using the linear approximation method costs very little in terms of researcher or computer time when compared to the multiple regression techniques needed before one can calculate the Gini Ratio using Kakwani and Podder's method. The relative cost is increased substantially if an iterative process is used to estimate the BETA function needed to calculate the Gini Ratio. Thus, the improved accuracy and information brought about through this method does not come at zero price. The absolute cost of calculation from Kakwani and Podder's method should not be prohibitive in most cases, so the improved accuracy and added information about the skewness of the distribution should make this the preferred method for most research.

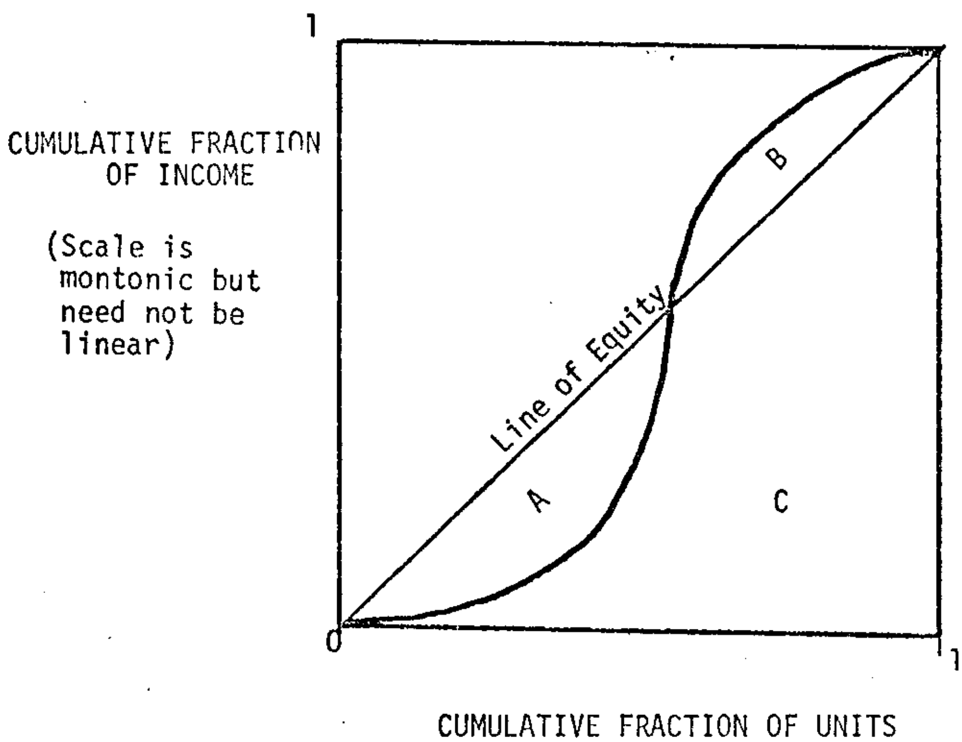
The normative implications of the Lorenzian definition of equality were outlined in an earlier section. This may cause one to look for a different measure of inequality. Paglin has suggested a method which modifies the Gini Ratio and allows for the use of a different definition of equality. His method breaks down the usual Gini Ratio, calculated from annual income data into two parts. One part is based on the dispersion of income due to the fact that individuals are at different stages of the life cycle. The second part is based on the remainder of the income dispersion from other reasons. Paglin defines equality as equal lifetime income but with equal income for each person or family at the same stage in the life cycle as opposed to the Lorenzian definition of equality which implies equal incomes for everyone regardless of stage in the life cycle. So the equality line in Paglin's case is really a Lorenz Curve where the cells are broken down according to age class instead of income class and then these age classes are ranked according to median income. Thus, the Paglin Gini is the difference between two Gini Ratios, one based on the distribution of income by age class, the other based on the distribution of income by income class. The usefulness of this method will depend on the type of data used, i.e. whether it is annual income data or not, and whether one's own definition of equity is consistent with Paglin's definition of equality.

Paglin's definition is only one of many that are possible. Thus, his conclusions, conceptual and empirical, carry only limited capacity for generalization.

The concept behind the Gini Ratio can be extended to measure inequity in certain cases. If one can specify what an equitable distribution is and this can be quantified with the cumulative fraction of income as a function of the cumulative fraction of units receiving income, then it is possible to measure deviations of any actual distribution from the equitable distribution. Using a monotonic non-linear scale between 0 and 1 for the cumulative fraction of income, the equitable distribution can be expressed as the diagonal of the unit square. Deviations from equity can then be plotted as deviations from the diagonal. The area between the plotted curve and the diagonal can be calculated again using the linear scale and the ratio of this area to the area under the diagonal would yield a measure of inequity with characteristics similar to the Gini Ratio. One noticeable difference between this method and the usual Lorenz based Gini Ratio is that in this new measure the concentration curve could lie above as well as below the diagonal depending on whether the actual distribution exhibited both positive and negative deviations from equity at different points in the distribution. (See Figure 3).

When the concentration curve lies above and below the diagonal, i.e. where there are both positive and negative deviations from equity, the formula for calculating the regular Gini Ratio will understate the real inequity. This is so since the usual Gini Ratio formula will only calculate the ratio of the area of net deviations from the diagonal to the area under the diagonal $\frac{A-B}{A+C}$, not the area of absolute deviations from the diagonal

Figure 3.


1
CUMULATIVE FRACTION OF INCOME
(Scale is montonic but need not be linear)
Line of Equity
A
B
C
0
1
CUMULATIVE FRACTION OF UNITS

$\frac{A+B}{A+C}$. (See Figure 3). Since we are interested in the total of all deviation from equity, the mathematical formula for calculating the standard Gini Ratio cannot be used to calculate this inequity ratio in all cases.

Conceptually it is possible to measure equity in the manner outlined above. However, the prior hurdle of defining and quantifying what an equitable distribution is, still must be overcome.

Paglin attempted to measure inequity by taking the difference between two Gini Ratios. His method has a similar shortcoming as the method above in that when there are both positive and negative deviations from equity, i.e. when the Lorenz Curve crosses the equity line (in his example when the Lorenz Curve crosses the P-reference line based on the age distribution of income), the Paglin Gini will understate the degree of inequity. While this may not be a problem for the example he uses, he states that this is a general method and can be used for any distribution that can be specified as equitable. There are many distributions of income or benefits which one or many believe is equitable, and his method could grossly misrepresent the amount of inequity in the actual distribution in these cases.

Conclusions and Policy Implications

Throughout this article numerous examples have been cited where difficulties arise in using or comparing Gini Ratios, particularly with respect to making welfare judgments based on the Gini Ratio. The problems in comparing Gini Ratios from different populations or public programs, the inherent biases in different methods of calculation, and the underlying assumptions of the Gini Ratio must all be considered by any economist or policy maker who uses this statistic.

There is a further question about the tradeoff between cost of calculation and the accuracy and information really needed. From a policy standpoint does increased accuracy in the Gini Ratio at the third decimal place really add to our understanding of the problem of income inequity? Furthermore does it aid us in our ability to solve this problem? What is the economic significance of differences between Gini Ratios? Too many economists leap from differences of unknown statistical significance to economic significance for policy without any demonstrable basis.

The distinction between equity and equality is central to any discussion of income or program benefit distributions. The Gini Ratio is only a statistical measure of relative inequality. To use it as the measure of inequity tends to eliminate many other important aspects of inequity. Most statisticians would not rely on the variance alone to describe the total distribution, they would at least require additional information about the mean and skewness to adequately characterize the distribution. In the same vein, economists require more information than that provided by the Gini Ratio alone to characterize inequity in the distribution of income. As long as one is aware of the limitations of the Gini Ratio and uses it with these limitations in mind then the Gini Ratio can provide important information about inequality. However since the linkage between equity and equality is tenuous at best, a specific measure of inequality such as the Gini Ratio provides no automatic information about inequity. This information can only be obtained after equity is carefully defined and related to particular statistics on the income distribution.

Footnotes

*The author wishes to thank Professor James T. Bonnen for his assistance throughout this project and Professors Lester V. Manderscheid, Byron W. Brown, Glenn L. Nelson and two anonymous reviewers for their helpful comments on earlier drafts of this paper.

[1] See e.g. Finifter (pp. 112-175).

[2] In the case above the value for the endpoints are determined by taking the mean payment per acre from two cells which have a common endpoint and multiplying this mean times the acreage value of the endpoint in question to get a dollar value for that endpoint, this process is repeated for all endpoints.

[3] The curve did not fit the actual data well and the estimate of the Gini Ratio was well below the lower bound of the Gini Ratio as calculated by Gastwirth's method for the distributions studied. For instance, for the distribution of payments for the 1971 cotton set-aside program the estimate of the Gini Ratio using Kakwani and Podder's original functional form was .6273 while as stated earlier the lower bound from Gastwirth's method was .6808 in this case.

[4] This method was used to estimate Lorenz curves and Gini Ratios for various distributions of payments from the cotton program for the years 1969-1973.

[5] McKee and Day have shown that if the curve is plotted for a part of total income but the order of the recipients is based on total income, the curve for the distribution can lie above the diagonal. But this is not a true Lorenz Curve and the Gini Ratio calculated from it is not a true Gini Ratio.

References

Bauer, Raymond A., "Detection and Anticipation of Impact: The Nature of the Task" Social Indicators, ed. Raymond A. Bauer, Cambridge, M.I.T. Press (1966).

Benson, Richard A., "Gini Ratios: Some Considerations Affecting Their Interpretation" American Journal of Agricultural Economics. 52(1970):444-447.

Bonnen, James T., "The Distribution of Benefits From Selected U.S. Farm Programs" Rural Poverty in the United States. A report by the President's National Advisory Commission on Rural Poverty, Washington, D.C. May (1968).

Boyne, David H., "Changes in the Income Distribution in Agriculture" Journal of Farm Economics. 47(1965):1213-1224.

Bronfenbrenner, Martin, "Equality and Equity" The Annals. The American Academy of Political and Social Science. 409(1973):9-23.

Champernowne, D.G., "A Comparison of Measures of Inequality of Income Distributions" The Economic Journal. 84(1974):787-816.

David, Martin and James N. Morgan, "A New Interpretation of Statistics on Income Distribution," Proceedings of the Business and Economic Statistics Section, American Statistical Association, Washington, D.C., (1961).

Finifter, Bernard M., "The Generation of Confidence: Evaluating Research Findings by Random Subsample Replication" Social Methodology, ed. Hebert L. Costner, San Francisco, Jossey-Bass Inc., (1972).

Gastwirth, J.L., "The Estimation of the Lorenz Curve and Gini Index," The Review of Economics and Statistics. 54(1972):306-316.

References Continued

Henriot, Peter J., "Political Aspects of Social Indicators: Implications for Research" Social Science Frontiers, Occasional Publication Reviewing New Fields for Social Science Development, No. 4, Russell Sage Foundation, (1972).

Hill, Roger P., "The Distribution of Benefits From the 1968 Upland Cotton Program" Southern Journal of Agricultural Economics. 2(1970):83-90.

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

____________________, "On the Estimation of Lorenz Curves From Grouped Observations" International Economic Review. 14(1973):278-292.

McKee, Vernon C. and Lee M. Day, "Measuring the Effects of U.S. Department of Agriculture Programs on Income Distribution," Rural Poverty in the United States, A report by the President's National Advisory Commission on Rural Poverty, Washington D.C., May (1968).

Mehran, Farhad, "Bounds on the Gini Index Based on Observed Points on the Lorenz Curve" Journal of the American Statistical Association. 70(1975):64-66.

Miller, Herman P., "Trends in the Income of Families and Persons in the United States: 1947-1960" Bureau of Census Tech. Paper 8, (1963).

Morgan, James, "The Anatomy of Income Distribution" Review of Economics and Statistics. 44(1962):270-283.

Paglin, Morton, "The Measurement and Trend of Inequality: A Basic Revision" American Economic Review. 65(1975):598-609.

Reinsel, Edward L., "People with Farm Earnings . . . Sources and Distribution of Income" USDA ERS Report 498, Washington, D.C., March (1972).

Sen, Amartya, On Economic Inequality Oxford, Clarendon Press, (1973).

Thurow, Lester C., Generating Inequality New York, Basic Books, (1975).