THE GINI COEFFICIENT AND NEGATIVE INCOME*

By CHAU-NAN CHEN, TIEN-WANG TSAUR *and* TONG-SHIENG RHAI

IT has widely been agreed that by far the best single measure of income inequality is the Gini coefficient of concentration (Morgan 1962). However, as Kendall and Stuart (1963, p. 47) put it, the coefficient suffers from "the disadvantage of being affected very much by the value of the mean measured from some arbitrary origin, and are not usually employed unless there is a natural origin of measurement or comparisons are being made between distributions with similar origins". The disadvantage becomes obvious when negative incomes are included in the distribution, for, in this case it is both theoretically (Hagerbaumer 1977) and empirically (Pyatt–Chen–Fei 1980)[1] possible for the coefficient to take on a value greater than one. In this paper we attempt to reformulate and normalize the Gini coefficient so that comparability can be attained between the distributions without negative incomes and the distributions with some negative incomes.

We consider n families and write Y_j for the income of the jth family, which is to be ordered from lowest to highest:

$$Y_1 \leqslant Y_2 \leqslant \ldots \leqslant Y_n. \tag{1}$$

Some of Y_j can be negative, but the total income is positive:

$$\sum_1^n Y_j > 0, \tag{2}$$

and so is the mean:

$$\mu = \sum_1^n Y_j/n > 0. \tag{3}$$

Then the Gini coefficient may be defined as

$$G = M/2\mu \tag{4}$$

where $M = (2/n^2) \sum_{j=1}^{n} \sum_{i<j} (Y_j - Y_i)$ is the mean difference. Write $y_j = Y_j/n\mu$ as the income share of the jth family.[2] (4) can then be expressed as

$$G = \left(\frac{1}{n}\right) \sum_{j=1}^{n} \sum_{i<j} (y_j - y_i). \tag{5}$$

* We are indebted to John Fei for his helpful comments.

[1] The estimate of the Gini coefficient of all other income (i.e., total family income subtract wage income, profit income and agricultural income) in the agricultural family by Pyatt, Chen and Fei (1980) for 1964 is 1.028.

[2] It goes without saying that $y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n$ and that $\sum_1^n y_j = 1$.

Expanding (5), we have

$$G = \left(\frac{1}{n}\right)((y_2 - y_1) + (y_3 - y_1) + (y_4 - y_1) + \ldots + (y_n - y_1)$$
$$+ (y_3 - y_2) + (y_4 - y_2) + \ldots + (y_n - y_2)$$
$$+ (y_4 - y_3) + \ldots\ldots\ldots\ldots$$
$$+ (y_n - y_{n-1}))$$
$$= \left(\frac{1}{n}\right) \sum_1^n y_j(2j - (n+1))$$
$$= \left(\frac{1}{n}\right) \sum_1^n y_j(n - 1 - 2(n - j))$$
$$= 1 - \left(\frac{1}{n}\right) \sum_1^n y_j(1 + 2(n - j))$$
$$= 1 + \left(\frac{2}{n}\right) \sum_1^k jy_j - \left(\frac{1}{n}\right) \sum_{k+1}^n y_j(1 + 2(n - j)). \qquad (6)$$

where k is defined in such a way so that $\sum_1^k y_j = 0$.

We are now ready to examine the range of the coefficient. The minimum value is obviously zero, which is attained when every family receives its proportional share of the total income so that $y_j = 1/n$ for all j. If one family earns all the income and the other families together earn nothing at all, so that $k = n - 1$, $y_n = 1$, $\sum_1^{n-1} y_j = 0$, (6) reduces to

$$G = 1 - \frac{1}{n} + \left(\frac{2}{n}\right) \sum_1^{n-1} jy_j. \qquad (7)$$

If incomes are nonnegative, so that $y_j = 0$ for $j < n$, and it the number of families, n, becomes greater and greater, G approaches one. However, if negative incomes are present, so that $y_j \gtreqless 0$, then G may exceed unity, for, with $y_j \gtreqless 0$, it is apparent that

$$\left(\frac{2}{n}\right) \sum_1^{n-1} jy_j > 0.$$

In general, $G \gtreqless 1$ depends on

$$\left(\frac{2}{n}\right) \sum_1^k jy_j \gtrless \left(\frac{1}{n}\right) \sum_{k+1}^n y_j(1 + 2(n - j)). \qquad (8)$$

It may be worth giving a numerical example here. Suppose that the income pattern of ten families is (−500, −300, −300, −100, 200, 300, 300, 400, 500, 500) as given by Schutz (1951, p. 119). G in this case will take a value of 1.94. This hardly appeals to our intuition once we realize that G

will take a value of only .99 when one percent of the families earn all the income while each of the rest of the families earns nothing at all. This suggests that the Gini coefficient may overestimate the inequality of income distribution when negative incomes are included. We propose to adjust or divide G by $1+(2/n)\sum_1^k jy_j$[3] to obtain a normalized Gini coefficient, G^*:[4]

$$G^* = \frac{1+\frac{2}{n}\sum_1^k jy_j - \frac{1}{n}\sum_{k+1}^{n} y_j(1+2(n-j))}{1+\frac{2}{n}\sum_1^k jy_j} \tag{9}$$

G^* will be equal to G if incomes are nonnegative, since in this case $k=0$ and the term $(2/n)\sum_1^k jy_j$ vanishes. Thus the minimum value of G^* is also zero. The maximum value of G^*, on the other hand, is unity. When one family earns all the income and the remaining $(n-1)$ families together earn nothing at all: $k=n-1$, $y_n=1$, $\sum_1^{n-1} y_j=0$, and the number of families becomes greater and greater: $n\to\infty$, G^* will approach unity:

$$G^* = \frac{1+\left(\frac{2}{n}\right)\sum_1^{n-1} jy_j - \frac{1}{n}}{1+\left(\frac{2}{n}\right)\sum_1^{n-1} jy_j} \to 1 \tag{10}$$

Like G in the case of nonnegative incomes, G^* ranges between zero and unity.[5]

A geometrical interpretation can be made by linking the Gini coefficient with the Lorenz curve. A hypothetical Lorenz curve is drawn in Figure 1 for the case $n=8$, $k=5$. Since the n families are ordered according to increasing income shares, the Lorenz curve does not cross the horizontal axis until positive incomes have balanced negative incomes. We shall now show that the Gini coefficient G is equal to one plus twice the area under the horizontal axis, 2A, minus twice the area under the Lorenz curve, 2C.

[3] If it so happens that $\sum_1^k y_j<0$, $\sum_1^{k+1} y_j>0$, then $\frac{2}{n}\sum_j^k jy_j$ should be written as

$$\frac{2}{n}\sum_1^k jy_j+\frac{1}{2n}\sum_1^k y_j\left\{\frac{\sum_1^k y_j}{y_{k+1}}-(1+2k)\right\}.$$

[4] (9) can be alternatively expressed as

$$G^*=\frac{2\,\mathrm{Cov}\,(y_j,\ j)}{1+(2k/n)\,\mathrm{Cov}\,(y_i,\ i)} \qquad \begin{matrix} i=1,2,\ldots k \\ j=1,2,\ldots n \end{matrix}$$

where Cov stands for the covariance.

[5] The value of G^* for the Schutz numerical example cited is 0.9065.

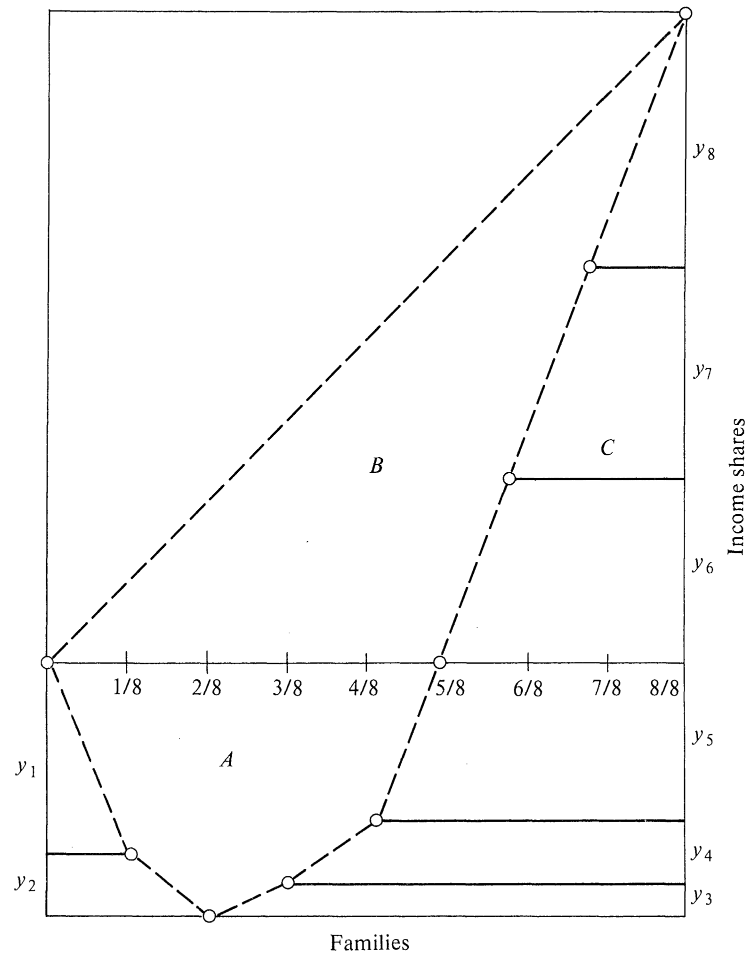

Figure 1. A Lorenz Curve with Negative Incomes

From inspection of Figure 1 we know that the area under the horizontal line, A, is

$$A = -\frac{1}{n}(\tfrac{1}{2}y_1 + (\tfrac{1}{2}y_2 + y_1) + (\tfrac{1}{2}y_3 + y_2 + y_1) + \dots$$
$$+ (\tfrac{1}{2}y_k + y_{k-1} + \dots + y_1)) \tag{11}$$
$$= \frac{-1}{2n}\sum_{1}^{k} y_j(1 + 2(k-j)) = \frac{1}{n}\sum_{1}^{k} j y_j,$$

and the area under the Lorenz curve, C, by the same token is

$$C = \frac{1}{2n}\sum_{k+1}^{n} y_j(1 + 2(n-j)). \tag{12}$$

Substituting (11) and (12) into (6), we obtain

$$G = 1 + 2(A - C). \tag{13}$$

Thus,

$$G \gtreqless 1, \text{ depending on } A \gtreqless C. \tag{14}$$

Substituting (11) and (12) into (9), we now have

$$G^* = \frac{1 + 2(A - C)}{1 + 2A}. \tag{15}$$

To rephrase the argument in terms of geometry:

(a) If incomes are nonnegative so that $A = 0$, then $G^* = 1 - 2C = G$. And if each family has equal share of total income so that $C = \frac{1}{2}$, then $G^* = 0$.

(b) If some of incomes are negative so that $A > 0$, then the smaller is the size of C, the greater will be G^*. In extreme, if only one out of a great number of families earns the total income, that is, if $C \to 0$, then G^* approaches unity.

A more intuitive explanation for our results can be made by rewriting G^* in (15) as a fraction:

$$G^* = \frac{A + B}{(A + B) + C}. \tag{16}$$

The area $(A + B)$ is usually called the area of concentration whereas the area C can be called the area of equalization. When negative incomes are absent ($A = 0$), it so happens that the sum of the two areas $(B + C)$ has a definite size of $\frac{1}{2}$, hence $G^* = G = B/(\frac{1}{2})$. But when negative incomes are present, these two areas no longer sum to a definite size and hence G^* should be defined as (16) rather than $(A + B)/(\frac{1}{2})$. The conventional Gini goes wrong because it treats the indefinite size of $(A + B + C)$ as a definite size of $\frac{1}{2}$.

We have shown that the Gini coefficient can be properly adjusted to suit both the nonnegative and negative income cases. However, it must be admitted that the ambiguity problem raised by Hagerbaumer (1977) still remains unsolved when the distributions with nonnegative incomes and the distributions with some negative incomes have identical Gini coefficients (the Lorenz curves intersect). It must also be admitted that the normalized coefficient proposed is hardly general since it cannot handle the case where negative and positive incomes sum to nil. In spite of these drawbacks, the normalized coefficient seems to have the following advantages:

(a) it retains all the basic properties of the conventional Gini coefficient with nonnegative incomes;

(b) similarly to the conventional Gini coefficient, it, too, has a simple geometric interpretation;

(c) unlike the method suggested by Schutz (1951) for handling negative incomes, it is directly comparable with the Gini coefficient with nonnegative incomes and therefore enables us to make full use of existing empirical results based on the Gini coefficient method.

As negative income is an unfamiliar concept, it may be worth explaining the ways in which it can arise. In our society there are two major types of economic decision-making units, namely productive enterprises (firms, farms, corporations, unincorporated enterprises) and families. It is easy to see that some of the productive enterprises may not be profitable or even make a loss. For these enterprises the net income is strictly negative. If there is a coincidence between an enterprise and a family, such as a family farm, then the family farm income may be negative. The same is true for a single proprietorship. In the case of corporations, the loss of the company will be reflected as a negative income for the property income receiving families. To sum up, to the extent that the productive enterprises are not profitable due to business depression or bad harvest, some of the families will tend to have negative income.[6]

Institute of Three Principles of the People, Academia Sinica, Taipei and *National Taiwan University, Taipei*

REFERENCES

MORGAN, J., "The Anatomy of Income Distribution," *Review of Economics and Statistics,* XLIV (Aug. 1962), 270–83.

HAGERBAUMER, J. B., "The Gini Concentration Ratio and the Minor Concentration Ratio: A Two-Parameter Index of Inequality," *Review of Economics and Statistics,* LIX (Aug. 1977), 377–79.

KENDALL, M., and A. STUART, *The Advanced Theory of Statistics I,* 2nd ed. (London: Charles Griffen and Company, 1963).

PYATT, G., C. N. CHEN, and J. C. H. FEI, "The Distribution of Income by Factor Components," *Quarterly Journal of Economics,* XCV (Nov. 1980), 451–73.

SCHUTZ R. R., "On the Measurement of Income Inequality," *American Economic Review,* XLI (Mar. 1951), 107–22.

[6] The agricultural income takes on negative values for some agricultural families in Taiwan from 1964 to 1976. Thus the Gini coefficients for this income in Table 2 and Table 3 of Pyatt, Chen and Fei (1980) may have overestimated the inequality of distribution.