BULLETIN OF ECONOMIC RESEARCH

Bulletin of Economic Research 46:1, 1994, 0307-3378

THE GINI-TYPE FUNCTIONS: AN ALTERNATIVE DERIVATION

*Branko Milanovic**

ABSTRACT

The purpose of this note is to propose an alternative and intuitively simpler derivation of the Gini coefficient (in Section 1), to show how it can be generalized and how then a number of coefficients (concentration coefficient, Kakwani's progressivity index) are obtained directly from this generalization (Section 2), and finally to use this approach to obtain some Gini relationships (Section 3).

1. THE DERIVATION OF THE GINI COEFFICIENT[2]

The Gini coefficient is equal to the area above the Lorenz curve (area E in Figure 1) divided by the area below the 45 degrees line (the sum of areas E and F).

The height of each strip such as aa' (Figure 1) is equal to

$$\sum_{i=1}^{j} p_i - \sum_{i=1}^{j} y_i$$

where p_i = proportion of recipients in the ith group, and y_i = proportion of total income received by the ith group. The expression

$$\sum_{i=1}^{j} p_i - \sum_{i=1}^{j} y_i$$

*I gratefully acknowledge comments by Professor Kakwani and two anonymous referees. I am responsible for the remaining errors.

[2] A slightly revised version of this section was published in Milanovic (1989).

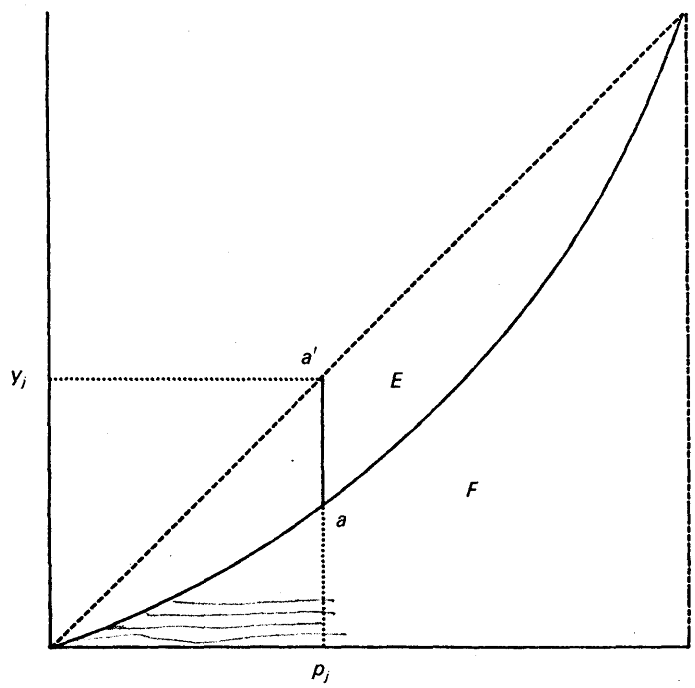

Fig. 1.

gives the height of the line aa' which corresponds to the population group j. Consequently, the *area* of that strip will be equal to

$$\left(\sum_{i=1}^{j} p_i - \sum_{i=1}^{j} y_i\right) p_j.$$

The whole area E is then

$$\text{area } E = \sum_{j=1}^{n} \left(\sum_{i=1}^{j} p_i - \sum_{i=1}^{j} y_i\right) p_j$$

where n = total number of population groups. By a similar reasoning the area $E+F$ will be equal to[3]

$$\text{area}(E+F) = \sum_{j=1}^{n} \left(\sum_{i=1}^{j} p_i - 0\right) p_j = \sum_{j=1}^{n} \sum_{i=1}^{j} p_i p_j.$$

[3] Note that, if all p_i are equal so that $p_i = (1/n)$, the area $(E+F)$ tends, for a sufficiently large n, to 0.5.

In matrix notation area E can be written as $\boldsymbol{p}'(\boldsymbol{Ap}-\boldsymbol{Ay})$ where $\boldsymbol{A}$ is a square matrix $(n \times n)$ that has 1's along the below the main diagonal, $\boldsymbol{p}$ = column vector of p_i and $\boldsymbol{y}$ = column vector of y_i. If $n=2$ we would have

$$[p_1\ p_2]\left(\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}\begin{bmatrix}p_1\\ p_2\end{bmatrix}-\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix}\right)=p_1(p_1-y_1)+p_2(p_1+p_2-y_1-y_2).$$

Similarly, the area $E+F$ can be written $\boldsymbol{p}'\boldsymbol{Ap}$.

The Gini coefficient (G) then becomes

$$G=\boldsymbol{p}'(\boldsymbol{Ap}-\boldsymbol{Ay})(\boldsymbol{p}'\boldsymbol{Ap})^{-1}=\boldsymbol{p}'\boldsymbol{A}(\boldsymbol{p}-\boldsymbol{y})(\boldsymbol{p}'\boldsymbol{Ap})^{-1} \tag{1}$$

Assuming that all groups are composed of the same number of individuals we can write $\boldsymbol{p}=p_0\boldsymbol{u}$ where p_0 = relative (percentage) size of the group and $\boldsymbol{u}$ = unit column vector.

Then (1) becomes

$$G=p_0\boldsymbol{u}'\boldsymbol{A}(p_0\boldsymbol{u}-\boldsymbol{y})(p_0\boldsymbol{u}'\boldsymbol{A}p_0\boldsymbol{u})^{-1}=p_0\boldsymbol{u}'\boldsymbol{A}(p_0\boldsymbol{u}-\boldsymbol{y})\frac{1}{(p_0)^2}(\boldsymbol{u}'\boldsymbol{Au})^{-1}$$

$$=\frac{1}{p_0}\boldsymbol{u}'\boldsymbol{A}(p_0\boldsymbol{u}-\boldsymbol{y})(\boldsymbol{u}'\boldsymbol{Au})^{-1}=\frac{1}{p_0}\boldsymbol{u}'\boldsymbol{A}(p_0\boldsymbol{u}-\boldsymbol{y})\left[\frac{n}{2}(n+1)\right]^{-1}$$

$$=\frac{2}{n(n+1)}\frac{1}{p_0}\boldsymbol{u}'\boldsymbol{A}(p_0\boldsymbol{u}-\boldsymbol{y}) \tag{2}$$

where we make use of $\boldsymbol{u}'\boldsymbol{Au}=n+(n-1)+(n+2)+\ldots+1=(n+1/2)n$.

If each group is composed of an individual income recipient $p_0=1/n$ and (2) can be further simplified

$$G=\frac{2}{n+1}\boldsymbol{u}'\boldsymbol{A}\left(\frac{1}{n}\boldsymbol{u}-\boldsymbol{y}\right). \tag{3}$$

Proportion of total income received by each individual is $y_i=m_i/n\bar{m}$ where m_i = income received by ith individual, $\bar{m}$ = average income of the population and n = total population. Then if m is the ordered column vector of $m_i(m_i<m_j)$ relation (3) becomes

$$G=\frac{2}{n+1}\boldsymbol{u}'\boldsymbol{A}\left(\frac{1}{n}\boldsymbol{u}-\frac{1}{n\bar{m}}\boldsymbol{m}\right)=\frac{2}{n(n+1)}\boldsymbol{u}'\boldsymbol{A}\left(\boldsymbol{u}-\frac{1}{\bar{m}}\boldsymbol{m}\right)=K_0\boldsymbol{w}\left(\boldsymbol{u}-\frac{1}{\bar{m}}\boldsymbol{m}\right) \tag{4}$$

where $K_0=\dfrac{2}{n(n+1)}$ and $\boldsymbol{w}=\boldsymbol{u}'\boldsymbol{A}$.

Also,

$$G = 1 - \frac{1}{\bar{m}} K_0 \boldsymbol{wm} \qquad (5)$$

since $K_0 \boldsymbol{wu} = 1$. Equation (5) gives a very straightforward and easy way to calculate the Gini coefficient.

Writing it all out,

$$G = K_0 [n \; n-1 \; n-2 \ldots 1] \left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} - \frac{1}{\bar{m}} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \cdot \\ \cdot \\ \cdot \\ m_n \end{bmatrix} \right)$$

Equations (4) and (5) give our final expression for the Gini coefficient.

2. *THE GENERALIZATION: THE GINI-TYPE FUNCTIONS*

Note first that $\boldsymbol{w} = \boldsymbol{u}'\boldsymbol{A}$ is a row-vector of the form $[n \; n-1 \; n-2 \ldots 1]$, and that $K_0 = 2/[n(n+1)]$ is the inverse of the sum of elements of $\boldsymbol{w}$.

This implies that the Gini coefficient can be interpreted as the *weighted average of differences* between one's importance as the member of a community (vector $\boldsymbol{u}$ composed of 1's) and one's importance as an income-receiving unit (vector $(1/\bar{m})\boldsymbol{m}$). Clearly, if these two things coincide, individual's income is the same as the average income and for all i's: $1-(1/\bar{m})\,\bar{m} = 0$. Income distribution is perfectly equal if $(\boldsymbol{u}-(1/\bar{m})\,\boldsymbol{m}) = 0$, and $G = 0$.

Weights range from n to 1 (divided by K_0) where the highest weight is attached to the lowest income recipient. The weights decline uniformly as income increases. The Gini coefficient weighs proportionately more discrepancies between one's importance as a member of a community and ones's importance as an income-recipient at low levels of income.

In general, we can call vector $\boldsymbol{w}$ divided by the sum of its elements K_0, the *weight vector*. A variety of weighting schemes can be imagined. Good examples are Suits' and Kakwani's measures of tax progressivity. In Suits' measure the weight vector $\boldsymbol{w}$ is such that element w_i is equal to the aggregate income of all individuals with incomes greater than the income of ith individual. Kakwani's measure, on the other hand, takes for weights the number of individuals with higher incomes. Kakwani's weights vector is thus identical to the one in equation (4). As Kakwani (1987, pp. 432–3) writes, the two measures reflect different value judgements about relative deprivation; in one case 'deprivation is captured by knowing how many people are richer'; in the other, by what the aggregate income of the richer is.

Vector $\boldsymbol{u}$ can be called the *criterion vector* since it embodies what we believe to be an appropriate 'equality' criterion. In the case of the Gini coefficient, it is assumed that every person is equally worthy and $\boldsymbol{u}$ is a unitary vector. But we shall see that other cases are possible. The vector $(1/\bar{m})\,\boldsymbol{m}$ can be called the *outcome vector*, since it shows what the actual situation (the outcome) is.

If we let the criterion vector be the relative (with respect to the average) pre-tax income of an individual, $(1/\bar{m})\,\boldsymbol{m}$, and the outcome vector, the relative tax paid, $(1/\bar{t})\,\boldsymbol{t}$ where $\bar{t}$ = average tax and $\boldsymbol{t}$ vector of taxes paid (ordered according to the level of pre-tax income), we obtain Kakwani's tax progressivity index P_t (see Kakwani (1986, p. 80)):

$$P_t = K_0\,\boldsymbol{w}\left(\frac{1}{\bar{m}}\,\boldsymbol{m} - \frac{1}{\bar{t}}\,\boldsymbol{t}\right). \tag{6}$$

If we develop (6), we can readily see that tax progressivity index is equal to the difference between the concentration coefficient of taxes C_t and the Gini coefficient of pre-tax income.

$$P_t = K_0\,\boldsymbol{w}\left(\boldsymbol{u} - \frac{1}{\bar{t}}\,\boldsymbol{t} - \boldsymbol{u} + \frac{1}{\bar{m}}\,\boldsymbol{m}\right) = C_t - G. \tag{7}$$

Now, substitution of (6) into (5) yields

$$G = 1 - P_t - K_0\,\frac{1}{\bar{t}}\,\boldsymbol{wt}. \tag{8}$$

By analogy with (5), the concentration coefficient of after-tax income (m^*) can be written

$$C_{m^*} = K_0\,\boldsymbol{w}\left[\boldsymbol{u} - \frac{1}{\bar{m}^*}\,\boldsymbol{m}^*\right] = 1 - \frac{1}{\bar{m}^*}\,K_0\,\boldsymbol{wm}^*$$

where $\bar{m}^* = \bar{m} - \bar{t}$ = average post-tax income, and $\boldsymbol{m}^*$ = vector of post-tax income ordered according to pre-tax income. Writing out C we obtain:

$$C_{m^*} = 1 - \frac{1}{\bar{m}^*}\,K_0\,\boldsymbol{w}(\boldsymbol{m} - \boldsymbol{t}) = 1 - \frac{1}{(1-g)\,\bar{m}}\,K_0\,\boldsymbol{wm} + \frac{1}{(1-g)\,\bar{m}}\,K_0\,\boldsymbol{wt} \tag{9}$$

where we use the fact that the average post-tax income is equal to (1 − average tax rate) *times* average pre-tax income, or $\bar{m}^* = (1-g)\bar{m}$, where g = average tax rate. (Note that $\bar{t}$ is the average tax paid in absolute amount while g is the average tax rate).

Substituting $(1/\bar{m}) K_0 \mathbf{wm} = 1 - G$ from equation (5) and $K_0 \mathbf{wt} = (1 - P_t - G)\bar{t}$ from equation (8) into (9), we obtain

$$C_{m^*} = 1 - \frac{1}{(1-g)}(1-G) + \frac{\bar{t}}{(1-g)\bar{m}}(1-P_t-G)$$

$$= 1 - \frac{1}{(1-g)}(1-G) + \frac{g}{(1-g)}(1-P_t-G) = G - \frac{g}{(1-g)}P_t \tag{10}$$

where we use $g = \bar{t}/\bar{m}$. The relation (10) is exactly the same as derived by Kakwani (1987, p. 433). It clearly shows the relationship between the concentration coefficient of post-tax income, the Gini coefficient of pre-tax income and the tax progressivity index.

It should be now clear that G, G_{m^*}, C_t, or P_t are only different types of a generalized Gini coefficient with criterion and outcome vectors taking different values (the weight vector is here always the same). We can thus write the Gini coefficient as

$$G(\mathbf{m}) = Z(1; \mathbf{m} \mid \mathbf{m})$$

where we use Z-operator to represent general class of Gini-type functions of the type (4): weights are of the form $K_0 w$, while the criterion and outcome vector can take any value. In the operator Z the first value shows the criterion vector (in the case of the Gini coefficient all 1's) and the second value, the outcome vector. The variable according to which the ranking is done is displayed after the sign |, so that $m \mid m$ denotes a vector of pre-tax income with recipients also ranked according to pre-tax income.

The concentration coefficient of post-tax income and progressivity index of taxes can then be written respectively as

$$C(\mathbf{m}^* \mid \mathbf{m}) = Z(1; \mathbf{m}^* \mid \mathbf{m})$$

$$P(\mathbf{t} \mid \mathbf{m}) = Z(\mathbf{m} \mid \mathbf{m}; \mathbf{t} \mid \mathbf{m}).$$

This formulation clearly shows that all Gini-type functions are calculated with respect to some criterion whether it is implicit (unit vector in the case of a concentration coefficient) or explicit (pre-tax income in the case of a progressivity index).

3. FURTHER GINI DERIVATIONS AND DECOMPOSITION

Using expressions (4) and (5) we can show how several additional relations can be easily derived.

Changes in the Gini

Let there be a change in income of ith individual ($\mathrm{d}m_i$) and corresponding change in mean income ($\mathrm{d}m_i/n$). By total differentiation of (5) we obtain

$$\mathrm{d}G=\frac{\delta G}{\delta m_i}\,\mathrm{d}m_i+\frac{\delta G}{\delta \bar{m}}\,\mathrm{d}\bar{m}=-(n-i+1)\frac{K_0}{\bar{m}}\,\mathrm{d}m_i+\frac{1}{\bar{m}^2}K_0\boldsymbol{wm}\,\mathrm{d}\bar{m}$$

$$=\frac{K_0}{\bar{m}}\left(-(n-i+1)+\frac{1}{\bar{m}n}\boldsymbol{wm}\right)\mathrm{d}m_i. \qquad (11)$$

Using the fact that $G=1-K_0(\boldsymbol{wm}/\bar{m})$ and $K_0=[2/n(n+1)]$ we can write (11) as

$$\mathrm{d}G=\frac{\mathrm{d}m_i}{Y}\left(1-G-\frac{2(n-i+1)}{(n+1)}\right) \qquad (12)$$

where Y= total income of the community ($\bar{m}n$). Equation (12) shows that, depending on whose income increases, G may go up or down. For low incomes (low value of i), the part between brackets will be negative. For example, for $i=1$, $\mathrm{d}G\doteq-(G+1)\,\mathrm{d}m_i/Y<0$. If income of the wealthiest recipient goes up, $\mathrm{d}G=(1/Y)[1-G-(2/n+1)]\,\mathrm{d}m_i\doteq(1-G)/Y\mathrm{d}m_i>0$.

Let now all incomes increase by the same amount ($\mathrm{d}m$); the mean income goes up by $\mathrm{d}m$ as well and we obtain:

$$\mathrm{d}G=\frac{-1}{\bar{m}}K_0\boldsymbol{wu}\,\mathrm{d}m+\frac{1}{\bar{m}^2}K_0\boldsymbol{wm}\,\mathrm{d}\bar{m}=\frac{-1}{\bar{m}}K_0\boldsymbol{w}\left(\boldsymbol{u}-\frac{1}{\bar{m}}\boldsymbol{m}\right)=\frac{-1}{\bar{m}}G\,\mathrm{d}m. \qquad (13)$$

The percentage change in the Gini coefficient is inversely proportional to the mean income and directly proportional to $\mathrm{d}m$. Thus, for example, an across-the-board increase in incomes equal to 1/5 of the average income, will result in reduction of the initial Gini coefficient by 20 per cent.

Transfers

If there is an infinitesimal transfer of income from a person with income m_j to a person with income $m_i(m_j>m_i)$, it can be shown that the change in the Gini coefficient will depend on the distance $j-i$. In effect,

$$\mathrm{d}G=\frac{\delta G}{\delta m_i}\,\mathrm{d}m_i+\frac{\delta G}{\delta m_j}\,\mathrm{d}m_j=-K_0(n-i+1)\frac{1}{\bar{m}}\,\mathrm{d}m_i$$

$$+K_0(n-j+1)\frac{1}{\bar{m}}\,\mathrm{d}m_i=\frac{-2}{n\bar{m}}\left(\frac{j-i}{n+1}\right)\mathrm{d}m_i.$$

Gini's sensitivity to a *given* transfer will be greater where the frequency distribution of recipients is 'denser' because for a given amount of transfer more people are likely to lie between the two persons concerned, j and i. The sensitivity will accordingly be the greatest at the mode of the distribution. The change in the Gini is less the smaller the relative size of the transfer $(\mathrm{d}m_i/Y)$.

Decomposition by Income Sources

Suppose now that total income is composed of two sources x and z (say, labor and capital) so that $m_i = x_i + z_i$. We can rewrite (4)

$$G = K_0 \boldsymbol{w}\left(\boldsymbol{u} - \frac{1}{\bar{m}}\boldsymbol{m}\right) = K_0 \boldsymbol{w}\left[\boldsymbol{u} - \frac{1}{\bar{m}}(\boldsymbol{x}+\boldsymbol{z})\right]$$

$$= K_0 \boldsymbol{w}\left(\boldsymbol{u} - \frac{1}{\bar{m}}\boldsymbol{x} - \frac{1}{\bar{m}}\boldsymbol{z}\right) = K_0 \boldsymbol{w}\left(\boldsymbol{u} - s_x\frac{1}{\bar{x}}\boldsymbol{x} - s_z\frac{1}{\bar{z}}\boldsymbol{z}\right)$$

where $\boldsymbol{x}$ and $\boldsymbol{z}$ are column vectors of labor and capital income ordered according to total income, and $s_x = \bar{x}/\bar{m}$ and $s_z = \bar{z}/\bar{m}$ shares of the two sources in total income.

Developing further,

$$G = K_0 \boldsymbol{w}[(s_x + s_z)\,\boldsymbol{u} - s_x(1/\bar{x})\,\boldsymbol{x} - s_z(1/\bar{z})\,\boldsymbol{z}]$$

$$= K_0 \boldsymbol{w}[s_x(\boldsymbol{u} - (1/\bar{x})\,\boldsymbol{x}] + K_0 \boldsymbol{w}[s_z(\boldsymbol{u} - (1/\bar{z})\,\boldsymbol{z}]$$

$$= s_x K_0 \boldsymbol{w}[\boldsymbol{u} - (1/\bar{x})\,\boldsymbol{x}] + s_z K_0 \boldsymbol{w}[\boldsymbol{u} - (1/\bar{z})\,\boldsymbol{z}] = s_x C_x + s_z C_z$$

where we made use of the fact that factor shares $s_x + s_z = 1$. It can be thus shown that the Gini coefficient is the weighted sum of concentration coefficients of income sources — the result obtained by Fei, Ranis and Kuo (1978) and Pyatt, Chen and Fei (1980).

Consider now the following problem. Let one source of income (labor) increase proportionately across all income recipients so that $\Delta x_i = \alpha x_i$ for all i. How would the overall Gini be affected?

There are two effects: change in individuals x's and change in the overall mean income.

Using (13) we know that the first term, the increase in labor income across all individuals, has the following effect on G:

$$\mathrm{d}G = \frac{-1}{\bar{m}} K_0 \boldsymbol{wu}\,\mathrm{d}m = \frac{-\alpha}{\bar{m}} K_0 \boldsymbol{wx} = \frac{-\alpha s_x}{\bar{x}} K_0 \boldsymbol{wx} \qquad (14)$$

where we use $s_x = \bar{x}/\bar{m}$.

The change in the overall mean $\mathrm{d}\bar{m} = \mathrm{d}\bar{x} = \alpha\bar{x}$ results in

$$\mathrm{d}G = \frac{1}{\bar{m}^2} K_0 \boldsymbol{wm}\,\mathrm{d}\bar{m} = \frac{\alpha\bar{x}}{\bar{m}^2} K_0 \boldsymbol{wm} = \frac{\alpha s_x \bar{m}}{\bar{m}^2} K_0 \boldsymbol{w}\bar{m}. \qquad (15)$$

Combining (14) and (15), we obtain

$$dG = \alpha s_x K_0 \mathbf{w}\left(\frac{1}{\bar{m}}\mathbf{m} - \frac{1}{\bar{x}}\mathbf{x}\right) = \alpha s_x K_0 \mathbf{w}\left(\mathbf{u} - \frac{1}{\bar{x}}\mathbf{x} - \mathbf{u} + \frac{1}{\bar{m}}\mathbf{m}\right)$$

$$= \alpha s_x C_x - \alpha s_x G = \alpha s_x (C_x - G). \tag{16}$$

When α tends to zero, relation (16) becomes

$$\frac{dG}{d\alpha} = s_x (C_x - G) = s_x P_x \tag{17}$$

Equation (17) is also derived by Stark, Taylor and Yitzhaki (1986). An infinitesimal and proportional increase in one source of income will therefore raise or lower the overall Gini coefficient depending on whether that source's concentration coefficient is greater or smaller than the Gini coefficient. It is important to note that throughout we assume that increase in x does not disturb the ranking of individuals by their overall income. If it does, then the rankings of recipients would change and the effect on the Gini would be indeterminate.

4. CONCLUSIONS

The note shows that the Gini coefficient is a special case of the general class of Gini-type fuctions which are the weighted average of the difference between the criterion and outcome vectors. The criterion vector embodies our value judgment about a given phenomenon (e.g. all individuals are equal); it represents the base-line scenario against which we compare the outcome vector (different individuals get different incomes). Finally, the weight vector embodies our *a priori* judgment as to what kind of discrepancies between the criterion and outcome vector should be deemed more or less important. A number of results between various Gini-type functions can be relatively easily derived using the alternative formulation proposed here.

The World Bank
Washington, DC

Received May 1992
final revision accepted
February 1993

REFERENCES

Fei, J., Ranis, G. and Kuo, S. (1978). 'Growth and the Family Distribution of Income by Factor Components', *Quarterly Journal of Economics*, February, pp. 17–53.