ECONOMIC ANALYSIS AND WORKERS' MANAGEMENT, 2, XXIV (1990), 157—169

GENERALIZED GINI COEFFICIENT: AN ALTERNATIVE APPROACH

*Branko MILANOVIĆ**

The purpose of this note is to propose an alternative and intuitively simpler derivation of the generalized Gini coefficient, and to show how a number of different applications follow directly from this alternative derivation. In Section 1 we present the derivation; Sections 2 and 3 give some further applications and derivations.

1. DERIVATION

As is well known, 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 (sum of areas E and F).

The height of each strip such as aa' (Figure 1, see page 168) is equal to $\sum_{i=1}^{j} p_i - \sum_{i=1} y_i$ where p_i = proportion of recipients in the i-th group, and y_i = proportion of total income received by the i-th group. The expression $\sum_{i=1}^{j} p_i - \sum_{i=1}^{j} y_i$ gives the height of the line aa' which corresponds to the population group j. Consequently, the area of that strip will be equal to $(\sum_{i=1}^{j} p_i - \sum_{i=1}^{j} y_i)\, p_j$. The whole area E is then given by

$$\textit{area } E = \sum_{j=1}^{n} (\sum_{i=1}^{j} p_i - \sum_{i=1}^{j} y_i)\, p_j \qquad (1)$$

* The World Bank, Washington.
Institut Ekonomskih nauka, Beograd.
I gratefully acknowledge comments made by the referee. I am alone responsible for the remaining errors.

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

$$area\ (E + F) = \sum_{j=1}^{n} (\sum_{i=1}^{j} p_i - 0)\, p_j = \sum_{j=1}^{n} \sum_{i=1}^{j} p_i\, p_j. \qquad (2)^1$$

In matrix notation area E can be written as p' (Ap — Ay) where A is a square matrix (nxn) that has l's along and below the main diagonal, p = column vector of p_i's and y = column vector of y_i's. 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 p'Ap.
The Gini coefficient (G) becomes

$$G \overset{n}{\underset{1}{=}} p' (Ap - Ay) (p' Ap)^{-1} = p' A (p - y) (p' Ap)^{-1} \qquad (3)$$

Assuming that all groups are composed of the same number of individuals we can write p = p_ou where p_o = relative (percentage) size of the group and u = unit column vector.
Then (3) becomes

$$G = p_o u'A (p_o u - y) (p_o u' A p_o u)^{-'} =$$

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

$$= \frac{1}{p_o} u'A (p_o u - y) (u'Au)^{-1} =$$

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

[1] The area should be equal to 1/2. Now suppose that all p_i are equal, so that $p_i = \frac{1}{n}$. Then, $\sum_{j=1}^{n} \sum_{i=1}^{j} p_i p_j = \frac{1}{n^2} \sum_{j}^{n} \sum_{i}^{j} 1 = \frac{1}{n^2} (1 + 2 + 3 + \ldots + n) = = \frac{1}{n^2} \frac{n+1}{2} n = \frac{n+1}{2n}$. When n tends to infinity, the last expression is equal to (1/2).

$$= \frac{2}{n(n+1)} \frac{1}{p_o} u'A(p_o u - y) \quad (4)$$

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

If each group is composed of an individual income recipient $p_o = \frac{1}{n}$ and (4) can be further simplified

$$G = \frac{2}{n+1} u'A \left(\frac{1}{n} u - y \right) \quad (5)$$

Proportion of total income received by each individual is $y_i = m_i/n\bar{m}$ where m_i = income received by i-th individual, $\bar{m}$ = average income of the population. Then if m is the ordered column vector of m_i's, relation (5) becomes

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

$$= K_o w \left(u - \frac{1}{\bar{m}} m \right) \quad (6)$$

where $K_o = \frac{2}{n(n+1)}$ and $w = u'A$. Writing it all out,

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

Relation (6) gives our final expression for the Gini coefficient.[2]

[2] Relation (6) is also equal to

$$G = = 1 - \frac{1}{\bar{m}} K_o wm$$

where the part $\frac{1}{m} K_o wm$ must be equal to twice the area below the Lorenz curve.

2. SOME APPLICATIONS

First note that $w = u'A$ is a row-vector of the form $[n \; n-1 \; n-2 \; .. 1]$, and that $K_o = 2/[n(n+1)]$ is the inverse of the sum of elements of w.

This then clearly implies that the Gini coefficient in expression (6) can be interpreted as the *weighted average of differences* between one's importance as the member of a community (vector u composed of 1's) and one's importance as an income-receiving unit (vector $(1/\overline{m})m$). Clearly, if these two things coincide, an individual's income is the same as the average income, and for all i's: $u_i - \frac{1}{\overline{m}} m_i = 0$. Income distribution is perfectly equal if $(u - \frac{1}{\overline{m}} m) = 0$, $G = 0$.

Weights range from n to 1 (divided by K_o) where the greatest weight is attached to the lowest income recipient. The weights decline uniformly as income level increases and the highest income unit receives a weight which is n times less than the weight given to the individual with the lowest income. This shows that the Gini coefficient weights proportionately more discrepancies between one's importance as a member of a community and his importance as an income-recipient at *low levels of income.*

In general, we can call vector w divided by the sum of its elements K_o, the *weight vector.* A variety of weighting schemes could be imagined. The best examples are Suits' and Kakwani's measures of poverty. Suits' measure of poverty is simply expression (6) with the weights vector w such that element w_i is equal to the aggregate income of all individuals with incomes greater than the income of i-th individual. Kakwani's measure, on the other hand, takes for weights the *number* of individuals with higher incomes. 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.

It is important to note that the generalized Gini coefficient includes two types of "value judgements": first, weights $K_o w$, which, as we saw, can be modifed depending on what aspect of inequality we want to emphasize, and second, "the yardstick", vector u, which can also be varied depending on what we believe to be the appropriate equality criterion. In general we can call vector u the *yardstick vector.* Analogously, vector $(1/m)m$ may be called the *outcome vector,* since it shows what the actual situation or outcome is.

We can illustrate the value judgment character implied in vector u with a few examples. A unit column vector u, as in equation (6), implicitly assumes that if each member of a community had the same income equality would be prefect. However, if we knew (hypothetically) what each individual's level of utility is, we could modify the yardstick by having the elements of the vector $u_i = k_i$ where k_i is the i-th individual marginal utility of income. If, as Edgeworth thought marginal utility of income was directly proportional to the level of income, the inequality thus measured would, for the *same* distribution of money incomes, be much less than the conventional inequality calculated using a unitary yardstick of u. On the other hand, if marginal utilities were inversely related to income, vector u would display high values for low i's, and the degree of inequality would be much greater. By varying the yardstick we also obtain the measure of tax progressivity (or vertical equity). Our yardstick now becomes an individual's relative income ($m_i/\overline{m}$), and the outcome vector the relative tax paid ($t_i/\overline{t}$), where $\overline{t}$ = the average tax paid. The generalized Gini thus naturally transforms into

$$G = K_o\, w\left(\frac{1}{\overline{m}}\, m - \frac{1}{\overline{t}}\, t\right) \tag{6a}$$

This expression is, when w is of the form [n n — 1 n — 2 . . . 1], identical to Kakwani's measure of tax progressivity. It may be noted that the yardstick vector in (6a) is what was the outcome vector in the original formulation of the generalized Gini in equation (6) while the outcome vector is now the relative taxation level (tax paid by the i-th individual divided by the avergae tax paid). Geometrically, the yardstick vector will be always on the abscissa (on Figure 1) and the outcome vector on the ordinate. If we want to find, for example, the own Gini coefficient of taxation we would have the cumulative percentage of recipients ranked by amount of taxes paid on the horizontal axis, and cumulative percentage of taxes paid on the vertical axis. If, on the other band, we want to asses progressivity of taxation, we would have cumulative percentage of income on thehorizontal, and cumulative percentage of taxes paid on the vertical axis (assuming horizontal equity to be observed). In either case the same formula will be used to calculate the area 2E — and the yardstick or the outcome vector will depend on the nature of the problem at hand.

We shall illustrate this with a simple example. Consider the distribution of managerial positions in an international organization by nationality of managers. If each country is given the same weight (weight of 1), the yardstick vector is a unit vector, and we simply calculate the usual Gini coefficent. In the event it is equal to 0.79. Let the yardstick vector now reflect the country's capital contribution to the organization, so that the Lorenz curve is obtained by charting the cumulative percentages of managerial positions against cumulative percentages of capital (with countries ranked according to their capital contribution). Obviously, we now implicitly assume that countries that have a higher contribution are also "entitled" to a greater number of leading positions in the organization. The Gini thus calculated is much less, 0.62.[3]

3. FURTHER DERIVATIONS AND DECOMPOSITION

We can now proceed with the derivation of several further relations using the expression (6). It allows us to determine easily the change in G due to an infinitesimal increase in m_i (such that the mean income does not change). We directly obtain

$$\frac{dG}{dm_i} = -(n-i+1)\,K_o\,\frac{1}{\overline{m}} \qquad (7)$$

The Gini coefficient goes down with an infinitesimal increase in any income (including the highest).

If we deal with a change in m_i such that the mean income changes as well (dm_i/n), we get

$$dG = \frac{\partial G}{\partial m_i}\,dm_i + \frac{\partial G}{\partial \overline{m}}\,d\overline{m} = -(n-i+1)\frac{K_o}{\overline{m}}\,dm_i +$$

$$+\frac{1}{\overline{m}^2}K_o\,wm\,d\overline{m} = \frac{K_o}{\overline{m}}\left[-(n-i+1) + \frac{1}{\overline{m}n}\,wm\right]dm_i.$$

Using the fact that $G = 1 - K_o\,(wm/m)$ and $K_o = 2/\,[n\,(n+1)]$ the last relation can be transformed into

$$dG = \frac{1}{Y}\left[1 - G - \frac{2\,(n-i+1)}{(n+1)}\right]dm_i \qquad (8)$$

where Y = total income of the community ($\overline{m}n$). Equation (8) shows that, depending on whose income increases, G may go up or down. For low incomes (low value of i), the part in brackets will be negative. For example, for $i = 1$, $dG = (1/Y)\,[1 - G - (2n/n+1)]\,dm_i = -(G+1)/Y\,dm_i < 0$. If the income of high income recipients goes up, dG would be positive. For example, if the "wealthiest" recipient's income goes up, $dG = (1/Y)\,[1 - G - (2/n+1)]\,dm_i = (1/Y)\,(1-G)\,dm_i$, which must be greater than zero. Equation (8) also shows that the decrease in G is

[3] Note that this underestimates the extent of "progressivity" linked with increase in capital contribution since horizontal equity is not observed: countries with higher capital contribution do not uniformly have more managerial positions.

greater (a) the greater the original Gini coefficient, and (b) the lower total income. Finally, from equation (8) we can determine i for which an increase in income (m_i) would not produce a change in the Gini coefficient ($dG = 0$). This i will, for a given total income, vary in function of G.

If we have 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,

$$dG = \frac{\partial G}{\partial m_i} dm_i + \frac{\partial G}{\partial m_j} dm_j = -K_o (n - i + 1) \frac{1}{\bar{m}} dm_i +$$

$$+ K_o (n - j + 1) \frac{1}{\bar{m}} dm_j = \frac{-2}{n} \frac{1}{\bar{m}} \left[\frac{j - i}{n + 1} \right] dm_i.$$

If a ("progressive") transfer of income takes place between the individual with the highest and the individual with the lowest income, $dG = [-2(n-1)] / [n\bar{m}(n+1)]$. If the distance between two individuals is given, i.e. $i - j =$ constant, then the effect on G will be the same, regardless where in the distribution these two individuals are located.

Finally, if all individuals' incomes are infinitesimally increased

$$\frac{\partial G}{\partial m} = \frac{-1}{\bar{m}} K_o w$$ where $\partial G / \partial m$ is a row vector of change in G due to increases in individual m_i's. Since all incomes increase by the same amount (dm) we shall have

$$dG = -\frac{1}{\bar{m}} K_o wu\, dm. \tag{9}$$

Average income will also increase by the same amount, and

$$dG = \frac{1}{\bar{m}^2} K_o wm\, dm. \tag{10}$$

Combining (9) and (10) we obtain

$$dG = \frac{-1}{\bar{m}} K_o wu + \frac{1}{\bar{m}^2} K_o wm = \frac{-1}{\bar{m}} K_o w \left(u - \frac{1}{\bar{m}} m \right) = \frac{-1}{\bar{m}} G\, dm$$

The percentage change in the Gini coefficient is inversely proportional to the mean income and directly proportional to dm. Thus, for example, an across-the-board increase in income equal to 1/5 of the average income, will result in reduction of the initial Gini coefficient by 20 percent.

A proportional increase in all incomes will, of course, leave the Gini unchanged. Equation (9) then becomes $dG = \frac{-1}{\bar{m}} a K_o wm$ where a is the percentage increase in income levels. Equation (10) is $dG = \frac{1}{\bar{m}^2} am K_o wm = \frac{1}{\bar{m}} a K_o wm$. The two relations cancel out.

Decomposition by factor components. 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 now rewrite (6)

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

$$\left(u - \frac{1}{\bar{m}} x - \frac{1}{\bar{m}} z \right) = K_o w \left(u - s_x \frac{x}{\bar{x}} - s_z \frac{z}{\bar{z}} \right) \tag{11}$$

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

Developing further,

$$G = K_o w \left[(s_x + s_z) u - s_x (1/\bar{x}) x - s_z (1/\bar{z}) z \right] =$$

$$= K_o w \left[s_x (u - (1/\bar{x}) x \right] + K_o w \left[s_z (u - (1/\bar{z}) z \right] =$$

$$= s_x K_o w \left[u - (1/\bar{x}) x \right] + s_z K_o w \left[u - (1/\bar{z}) z \right] =$$

$$= s_x C_x + s_z C_z$$

where we made use of the fact that factor shares $s_x + s_z = 1$ and $C_x = K_o w \left[u - (1/\bar{x}) x \right]$ is the concentration coefficient of the labor income. The Gini coefficient is the weighted sum of concentration coefficients of different income sources — the result obtained by Fei, Ranis and Kuo (1978), and Pyatt, Chen and Fei (1980).

Note that the Gini coefficient of labor income alone (when recipients are ranked by the level of labor income) can be written $G_x = K_o w \left[u - (1/\bar{x}) \hat{x} \right]$, where $\hat{x}$ is the ordered vector of labor incomes. Since the weight vector is such that values for lower i's are higher, it must then be true that the Gini coefficient is greater (or equal, if the rankings according to labor income and total income coincide) than the concentration coefficient (C_x).

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

$$K_o [n \; n-1 \; n-2 \ldots 1] \left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} - \frac{1}{\bar{m}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} - \frac{1}{\bar{m}} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_n \end{bmatrix} \right) \quad (12)$$

we see that there are two effects: change in individuals x's and change in the overall mean income ($\bar{m}$).

The increase in i-th labor income (x_i) has the following effect on G: $dG = K_o(n - i + 1) \frac{-1}{\bar{m}} dx_i$. Since the increase is proportional across all recipients we get

$$dG = K_o \frac{1}{\bar{m}} \sum_{i=1}^{n} (n - i + 1)\, dx_i = K_o \frac{1}{\bar{m}} a\, wx = -K_o \frac{s_x}{\bar{x}} a\, wx \quad (13)$$

where $dx_i = a x_i$.

Change in $\bar{m}$, on the other hand, has the following effect on the Gini coefficient:

$$\frac{dG}{dm} = \frac{K_o}{\bar{m}^2} \sum_{i=1}^{n} (n - i + 1)(x_i + z_i).$$

Now, since $d\bar{m} = d\bar{x} = a\bar{x}$, we can write the last expression

$$dG = \frac{K_o}{\bar{m}^2} a\bar{x}\, wm = \frac{K_o}{\bar{m}} s_x\, a\, wm \quad (14)$$

Combining (13) and (14) we get final effect on the Gini coefficient

$$dG = K_o\, s_x\, a \left[-w \frac{x}{\bar{x}} + w \frac{m}{\bar{m}} \right]$$

which can be further transformed

$$dG = K_o s_x aw \left[\left(u - \frac{x}{\bar{x}} \right) - \left(u - \frac{m}{\bar{m}} \right) \right]$$

$$= s_x a K_o w [u - (x/\bar{x})] - s_x a K_o w [u - (m/\bar{m}] =$$

$$= s_x a C_x - s_x a G = s_x a (C_x - G) \quad (15)$$

When a tends to infinity, relation (15) becomes

$$\frac{dG}{da} = S_x (C_x - G)$$

which is exactly the relation derived by Stark, Taylor and Yitzhaki (1986). A small 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 in (12) would also change and vectors x and y would both change so that the effect on the Gini would be indeterminate. Equation (15) accordingly gives the effect of a uniform and *infinitesimal* increase in one source of income.

Received: 29. 08. 1989
Revised: 14. 02. 1990

ANNEX

Relationship between definition of generalized Gini presented here and in Kakwani (1986).

Kakwani (1986, p. 65) defines a general class of inequality measures:

$$G(k) = \frac{n-1}{n [\phi(k) - n)} \frac{1}{m} \sum_{i=1}^{n} (\bar{m} - m_i) (n - i + 1)^k$$

where all the symbols are as explained, and $\phi(k) = \sum_{i=1}^{n} i^k$; if $k = 1$, $\phi(k) = n(n+1)/2$ and we have the Gini coefficient:

$$G(1) = \frac{n-1}{(n/2)(n^2 - n)} \times$$

$$\times \left[\frac{1}{\overline{m}} \sum_{i=1}^{n} \overline{m} (n-i+1) - \frac{1}{\overline{m}} \sum_{i=1}^{n} m_i (n-i+1) \right] = \frac{2}{n^2}.$$

$$\times \left[\sum_{i=1}^{n} (n-i+1) - \frac{1}{\overline{m}} \sum_{i=1}^{n} m_i (n-i+1) \right] = \frac{2}{n^2} \frac{n(n+1)}{2} -$$

$$- \frac{1}{\overline{m}} \sum_{i=1}^{n} m_i (n-i+1) = 1 + \frac{1}{n} - \frac{2}{n^2 \overline{m}} \sum_{i=1}^{n} m_i (n-i+1). \quad (*)$$

The generalized Gini from equation (6) can be written as

$$G = \frac{2}{n(n+1)} \sum_{i=1}^{n} (n-i+1) \frac{1}{\overline{m}} m - m_i) = \frac{2}{\overline{m}n(n+1)} \cdot$$

$$\times \left[\sum_{i=1}^{n} (n-i+1) \overline{m} - \sum_{i=1}^{n} m_i (n-i+1) \right] =$$

$$= \frac{2n(n+1)\overline{m}}{2\overline{m}n(n+1)} - \frac{2}{\overline{m}n(n+1)} \sum_{i=1}^{n} m_i (n-i+1) =$$

$$= 1 - \frac{2}{\overline{m}n(n+1)} \sum_{i=1}^{n} m_i (n-i+1) =$$

$$= 1 - \frac{2}{\overline{m}n(n+1)} \sum_{i=1}^{n} m_i (n-i+1) \qquad (**)$$

If we now write $A^* = \frac{2}{n\overline{m}} \sum_{i=1}^{n} (n-i+1) m_i$, Kakwani's formula in (*) becomes $G = 1 + [(1 - A^*)/n]$. Our formula in (**) is $G = 1 - [A^*/(n+1)]$. The difference between the two is equal to $\frac{1}{n} - \frac{A^*}{n(n+1)}$.

When n tends to infinity the difference becomes nil.

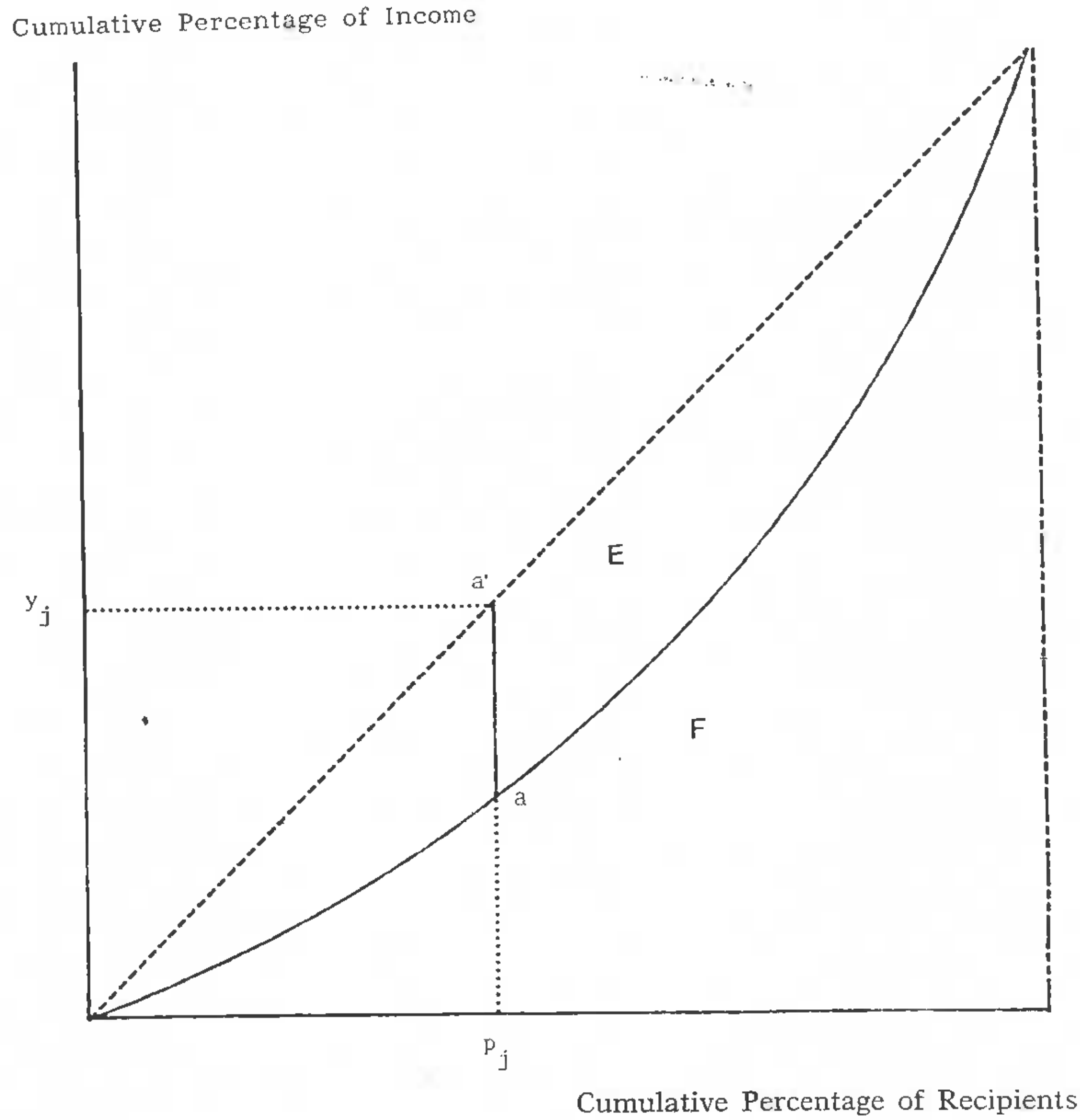

Figure 1

REFERENCES

Fei, John, Gustav Ranis and Shirley Kuo (1978), "Growth and the Family Distribution of Income by Factor Components", *Quarterly Journal of Economics,* February, pp. 17—53.

Kakwani, Nanak (1986), *Analyzing Redistribution Policy,* Cambridge: Cambridge University Press.

Kakwani, Nanak (1987), "Measures of Tax Progressivity and Redistribution Effect: A Comment", *Public Finance,* No. 3, 1987, p. 431.

Pyatt, Graham, Chau-nan Chen and John Fei (1980), "The Distribution of Income by Factor Components", *Quarterly Journal of Economics*, November, pp. 451—473.

Stark, Oded, Edward Taylor and Shlomo Yitzhaki (1986), "Remittances and Inequality", *The Economic Journal*, September, pp. 722—740.