Institute of Economic Development and Research
SCHOOL OF ECONOMICS
University of the Philippines

Discussion Paper No. 73-7 March 21, 1973

A NOTE ON DECOMPOSITION OF THE GINI RATIO ACROSS REGIONS

by

Mahar Kelly Mangahas, 1944-

A NOTE ON DECOMPOSITION OF THE GINI RATIO ACROSS REGIONS

Mahar Mangahas

In recent approaches to the measurement of the contributions of subsectors or regions of the economy to national income inequality, the measure used --the index of decile inequality, the variance of the logarithm of income-- has been linear or quadratic, inasmuch as such measures are relatively simple to decompose [1, 2]. However, the most common measure of income inequality is the Gini or concentration ratio deriving from the Lorenz curve. The purpose of this note is to indicate that the national-level Gini ratio can be expressed as a weighted average of regional Gini ratios and of certain Gini-type ratios constructed from pairwise regional comparisons of the size distribution of income.

Let f_k^* be the cumulative proportion of families up to the k^{th} income class, and y_k^* the cumulative proportion of income received by those families, for $k = 1, \ldots, G$. The Gini ratio is computed by triangular approximation as

$$L = 1 - 2 \sum_{k=1}^{G} \left[1/2\ (f_k^* - f_{k-1}^*)(y_k^* - y_{k-1}^*) + (f_k^* - f_{k-1}^*) y_{k-1}^* \right]$$

where $f_0^* = y_0^* = 0$. The summation expression on the right-hand-side is the approximate area underneath the Lorenz curve. This reduces to

$$L = 1 - 2 \sum_{k=1}^{G} \left[1/2 \ (f_k^* - f_{k-1}^*) y_k^* + 1/2 \ (f_k^* - f_{k-1}^*) y_{k-1}^* \right]$$

$$L = 1 - \sum_{k=1}^{G} (f_k^* - f_{k-1}^*)(y_k^* + y_{k-1}^*)$$

$$(1) \qquad L = 1 - \sum_{k=1}^{G} f_k \ (y_k^* + y_{k-1}^*)$$

where $f_k = f_k^* - f_{k-1}^*$ is simply the proportion of families within the k^{th} income class. We also define $y_k = y_k^* - y_{k-1}^*$ as the proportion of total incomes enjoyed by families within the k^{th} income class.

Now define

$$f = \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_G \end{bmatrix} \qquad \text{and} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \vdots \\ \cdot \\ y_G \end{bmatrix} .$$

Then note that

$$y^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \cdot \\ \cdot \\ \cdot \\ y_G^* \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_G \end{bmatrix} = Cy \quad ,$$

where C is the matrix with ones on and below the diagonal, and zeros elsewhere. Furthermore,

$$y^*_{-1} = \begin{bmatrix} y^*_0 \\ y^*_1 \\ \cdot \\ \cdot \\ \cdot \\ y^*_{G-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_G \end{bmatrix} = (C - I)y$$

where I is the $G \times G$ identity matrix. In matrix notation, the Gini ratio is then

$$L = 1 - f'(y^* + y^*_{-1})$$

$$= 1 - f'(Cy + (C - I)y)$$

(2) $$L = 1 - f'H\,y$$

where

$$H = (2C - I) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 2 & 1 & 0 & \dots & 0 \\ 2 & 2 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ 2 & 2 & 2 & \dots & 1 \end{bmatrix},$$

a matrix with twos below the diagonal, ones on the diagonal, and zeros above the diagonal.

In general, if r and s are any two $G \times 1$ vectors containing percentage distributions, and if one wishes to compare equality between the two distributions by cumulating them, the Gini-type measure of inequality is given by $(1 - r'Hs)$. In particular, let the vectors f and y refer to national-level data and let f_j and y_j be $G \times 1$ vectors similarly defined for the j^{th} region, with $j = 1, \ldots, R$. Then the regional-level Gini ratios are

$$(3) \qquad L_j = 1 - f_j' H y_j \qquad , \qquad j = 1, \ldots, R \qquad .$$

If n is the total number of families in the nation, then nf is the $G \times 1$ vector whose k^{th} element is the total number of families in the k^{th} income class. Let X be a $G \times G$ diagonal matrix whose k^{th} diagonal element is mean family income in the k^{th} income class. Then nXf is the $G \times 1$ vector whose k^{th} element is the total family income earned by families belonging to the k^{th} income class. Total family income in the nation is then

$$(4) \qquad v = \iota' Xf \cdot n$$

where ι is a $G \times 1$ vector of ones. Then y is given by

$$(5) \qquad y = (n/v)Xf = (\iota' Xf)^{-1} Xf = (1/m)Xf \qquad ,$$

where m is the mean family income in the nation. Since f determines y, f is the basic data vector, and may be considered synonymous with "the size distribution of income". The mean income levels per class, or the diagonals of X, depend on the distribution of families within each class's upper and lower bounds. As a simplification, X may be considered identical for each region and for the nation as a whole; in principle at least one can always arrive at approximately equal X's by simply constructing a large enough number of income classes, with very narrow intervals.

From (2) and (5) we obtain

(6) $$1 - L = (1/m)f'HXf = (1/m)f'Pf \quad ,$$

where $P = HX$ may be viewed as a matrix of constants, on account of the argument in the preceding paragraph. For the regions, we similarly obtain

(7) $$1 - L_j = (1/m_j)f_j'Pf_j \quad , \qquad j = 1, \ldots, R \quad ,$$

where $m_j = \iota'Xf_j$ is the mean family income in the j^{th} region.

The next problem is to determine how L and the L_j are related. Define

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \cdot \\ \phi_R \end{bmatrix}$$

where ϕ_j is the proportion of all families in the nation who live in region j; thus $\Sigma\phi_j = 1$. Consolidating the regional size distributions of income into a $G \times R$ matrix F, where

$$F = [f_1 \; f_2 \; \ldots \; f_R] \quad ,$$

then we have

$$f = F\phi \quad .$$

Therefore (6) becomes

$$(8) \qquad 1 - L = (1/m)\phi' F' P F\phi \quad .$$

We now recognize that $1 - L$ is the sum of all the terms of an $R \times R$ matrix whose diagonal elements are

$$(9) \qquad (1/m)\phi_j^2 f_j' P f_j = (m_j/m)\phi_j^2 (1 - L_j) \quad , \quad j = 1, \ldots, R$$

and whose off-diagonals are

$$(10) \qquad (1/m)\phi_i \phi_j f_i' P f_j \quad . \qquad i \neq j;\ i, j = 1, \ldots, R \quad .$$

Note that

$$f_i' P f_j = f_i' H X f_j = m_j f_i' H (1/m_j) X f_j = m_j (f_i' H\, y_j) \quad .$$

The expression in parentheses can be interpreted as (one minus) the result if one were to compute a Gini ratio by comparing the distribution across income classes <u>of families in the i^{th} region</u> with the distribution across income classes <u>of incomes received by families in the j^{th} region</u>.

Furthermore, note that

$$(f_i - f_j)' P (f_i - f_j) = f_i' P f_i + f_j' P f_j - f_i' P f_j - f_j' P f_i \quad ,$$

where the last two terms on the right-hand-side are elements of "Gini cross-ratios" such as those in (10). We can then write the sum of the elements in (10) as

$$(11)\qquad \sum_{i>j} [(\phi_i\phi_j/m)f_i'Pf_i + (\phi_i\phi_j/m)f_j'Pf_j - (\phi_i\phi_j/m)(f_i - f_j)'P(f_i - f_j)] =$$

$$= \sum_{i>j} [(\phi_i\phi_j m_i/m)(1 - L_i) + (\phi_i\phi_j m_j/m)(1 - L_j) - (\phi_i\phi_j(m_i - m_j)/m)(1 - L_{ij})] \quad ,$$

where we recognize that

$$L_{ij} = 1 - (m_i - m_j)^{-1} (f_i - f_j)' P(f_i - f_j)$$

is a Gini ratio computed on the difference between the size distributions of income in regions i and j. Let us term this a "Gini difference-ratio". Combining (11) with the sum of the terms (9), we obtain

$$1 - L = \sum_i \phi_i \sum_j (\phi_j m_j/m)(1 - L_j) - \sum_{i>j} (\phi_i\phi_j(m_i - m_j)/m)(1 - L_{ij}) \quad .$$

Since $\Sigma\phi_i = 1$, $m = \Sigma\phi_j m_j$, and $\sum_{i>j} \phi_i\phi_j (m_i - m_j)/m = 0$, we derive

$$(12)\qquad L = \sum_j (\phi_j m_j/m)L_j + \sum_{i>j} (\phi_i\phi_j(m_j - m_i)/m)L_{ij} \quad .$$

This is a decomposition of the national Gini ratio as the sum of a weighted average (weights adding to one) of the regional Gini ratios and a weighted sum (weights adding to zero) of all possible Gini difference ratios. Thus the first expression measures the contribution of "within-region inequality" whereas the second measures the contribution of "between-region inequality".

An informative way of presenting the decomposition might be in the form of a table, as below. The diagonal elements would be $\phi_j m_j L_j / mL$ and the lower triangular elements would be $\phi_i \phi_j (m_i - m_j) L_{ij} / mL$. The sum of all the elements would be one, and relatively large numbers would indicate the most pressing sources of income inequality.

Relative Sizes of Regional Sources of National Income Inequality

	Region 1	Region 2	. . .	Region R
Region 1	Inequality Within 1			
Region 2	Inequality Between 2 and 1	Inequality Within 2		
⋮	⋮	⋮		
Region R	Inequality Between R and 1	Inequality Between R and 2	. . .	Inequality Within R

REFERENCES

[1] H.T. Oshima, "Income Inequality and Economic Growth: the Postwar Experience of Asian Countries," Malayan Economic Review, vol. XV, no. 2 (October 1970), pp. 7-41.

[2] M. Mangahas, "A Note on 'Income Inequality and Economic Growth: the Postwar Experience of Asian Countries,'" forthcoming in the Malayan Economic Review.