

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

Discussion Paper No. 74-2

February 19, 1974

A NOTE ON DECOMPOSITION OF THE  
GINI RATIO ACROSS REGIONS

(Revised)

by

Mahar Mangahas

Note: IEDR Discussion Papers are preliminary versions circulated privately to elicit critical comment. References in publications to Discussion Papers should be cleared with the author.

IEDRDP '9  
12.4.97

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  $k^{\text{th}}$  income class, and  $y_k^*$  the cumulative proportion of income received by those families, for

---

\*Part of a research project on Sectoral Analysis and the Size Distribution of Income, supported by the International Labor Organization. The author has benefited from criticism by J. Encarnacion.

$k = 1, \dots, G$ . The Gini ratio is defined as

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

where  $f_0^* = y_0^* = 0$ . The summation expression on the right-hand-side is the area underneath the Lorenz "curve", where plotted points are joined by straight lines. This reduces to

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

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

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

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

Now define

$$f = \begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ f_G^* \end{bmatrix}$$

and

$$y = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_G^* \end{bmatrix}$$

Then

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

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

$$y_{-1}^* = \begin{bmatrix} y_0^* \\ y_1^* \\ \vdots \\ y_{G-1}^* \end{bmatrix} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{bmatrix} = (C - I)y$$

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

$$\begin{aligned} L &= 1 - f'(y^* + y_{-1}^*) \\ &= 1 - f'(Cy + (C - I)y) \end{aligned}$$

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

where

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

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  $j^{\text{th}}$  region, with  $j = 1, \dots, R$ . Then the regional-level Gini ratios are

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

If  $n$  is the total number of families in the nation, then  $nf$  is the  $G \times 1$  vector whose  $k^{\text{th}}$  element

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

$$(4) \quad v = \mathbf{1}'Xf \cdot n$$

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

$$(5) \quad y = (n/v)f = (\mathbf{1}'Xf)^{-1} Xf = (1/m)Xf$$

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

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

where  $P = HX$  may be viewed as a matrix of constants, on account of the argument in the preceding paragraph. With  $H$  triangular,  $X$  diagonal, and all elements in  $H$  and  $X$  positive, it follows that the matrix  $P$  is positive definite.<sup>1/</sup> For the regions, we similarly obtain

$$(7) \quad 1 - L_j = (1/m_j) f_j' P f_j, \quad j = 1, \dots, R,$$

where  $m_j = (f_j' X f_j)$  is the mean family income in the  $j^{\text{th}}$  region.

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

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

where  $\phi_j$  is the proportion of all families in the nation who live in region  $j$ ; thus  $\sum \phi_j = 1$ . Consolidating the

---

<sup>1/</sup> Thus, strictly speaking,  $L$  may get very close to one, but never quite reaches it.

regional size distributions of income into a  $G \times R$  matrix  $F$ , where

$$F = [f_1 \ f_2 \ \dots \ f_R] \quad ,$$

then we have

$$f = F\phi$$

Therefore (6) becomes

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

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

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

and whose off-diagonals are

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

Note that

$$(11) \quad (f_i - f_j)' P (f_i - f_j) = f_i' P f_i + f_j' P f_j - f_i' P f_j - f_j' P f_i \quad ,$$

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

Then the sum of the elements in (10) is

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

We now focus on the expression  $(f_i - f_j)' P (f_i - f_j)$ . Consider two regions which are internally completely equal. In the first region, let all families be found in income class  $k_1$ ; and in the second, let all families be in class  $k_2$ , which we set arbitrarily as a higher class than  $k_1$ . Then

$$f_1 - f_2 = \begin{pmatrix} : \\ 1 \\ : \\ -1 \\ : \end{pmatrix}$$

where the term 1 is in the  $k_1^{\text{th}}$  place, the term -1 in the  $k_2^{\text{th}}$  place, and there are zeros elsewhere. In this case,

$$\begin{aligned} (f_1 - f_2)' P (f_1 - f_2) &= (\dots 1 \dots -1 \dots) H X \begin{pmatrix} : \\ 1 \\ : \\ -1 \\ : \end{pmatrix} \\ &= (\dots 1 \dots -1 \dots) H \begin{pmatrix} : \\ x_{k_1} \\ : \\ -x_{k_2} \\ : \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} \dots & 1 & \dots & -1 & \dots \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_{k_1} \\ \vdots \\ 2x_{k_1} - x_{k_2} \\ \vdots \end{pmatrix},$$

where the column vector has  $x_{k_1}$  in the  $k_1^{\text{th}}$  place and  $2x_{k_1} - x_{k_2}$  in the  $k_2^{\text{th}}$  place (actually, all terms beginning with the  $k_1^{\text{th}}$  are non-zero, but only the two indicated are essential). Then

$$(f_1 - f_2)'P(f_1 - f_2) = x_{k_1} - 2x_{k_1} + x_{k_2} = x_{k_2} - x_{k_1}$$

The result is the same if region one happens to be the richer region. Thus, if two regions are internally equal, then  $(f_i - f_j)'P(f_i - f_j)$  is the range between their respective means. This is a maximum when all families in one region are in the poorest class, while all families in the other region are in the richest class, in which case the maximum value is  $x_G - x_1$ , the range of mean incomes across all classes. We therefore define the Gini difference-ratio between regions  $i$  and  $j$  as

$$(12) \quad L_{ij} = \frac{1}{x_G - x_1} (f_i - f_j)' P (f_i - f_j)$$

This symmetric expression is zero if and only if the percentage distributions of families by income class are identical for the two regions. The expression is one when all families in one region are "equally very poor" and all families in the other region are "equally very rich". Negative values for  $L_{ij}$  are excluded by the positive-definiteness of  $P$ .

The difference-ratio compares two regions' size distributions of income, not merely their means. Two unequal distributions may have equal means; nevertheless  $L_{ij}$  will be positive. For instance, consider five income classes: lower, lower-middle, middle, upper-middle, and upper. Suppose three regions had the same mean income, but (a) region one had all its families in the middle class, (b) region two had half of its families in the upper-middle class and half in the lower middle class, and (c) region three had half of its families in the lower class and half in the upper class. Then it can easily be shown that  $L_{12}$ ,  $L_{13}$  and  $L_{23}$  are all positive, and furthermore that  $L_{13} > L_{12}$ , as we would intuitively desire.

Lastly, to take an extreme case, suppose all regions in the country had the same mean family income, but different size distributions. Then the variance-decomposition<sup>2/</sup> of income inequality would indicate no between-region inequality at all, whereas the various  $L_{ij}$  would be positive.

The sum of the elements in (10) may now be written

$$\sum_{i>j} \left[ (\phi_i \phi_j m_i / m)(1 - L_i) + (\phi_i \phi_j m_j / m)(1 - L_j) - \frac{\phi_i \phi_j (x_G - x_1)}{m} L_{ij} \right]$$

Combining this sum with the sum of the terms in (9) gives

$$\begin{aligned} 1 - L &= \sum_i \sum_j \frac{\phi_i \phi_j m_j}{m} (1 - L_j) - \sum_{i>j} \frac{\phi_i \phi_j (x_G - x_1)}{m} L_{ij} \\ &= \sum_j \frac{\phi_j m_j}{m} (1 - L_j) - \sum_{i>j} \frac{\phi_i \phi_j (x_G - x_1)}{m} L_{ij} \end{aligned}$$

---

<sup>2/</sup> If the decomposition is taken on the variance of the logs of income, then the supposition is that geometric means of family income are the same across regions.