On the Interpretation and Disaggregation of Gini Coefficients

Author(s): Graham Pyatt

Source: *The Economic Journal*, Jun., 1976, Vol. 86, No. 342 (Jun., 1976), pp. 243-255

Published by: Oxford University Press on behalf of the Royal Economic Society

Stable URL: https://www.jstor.org/stable/2230745

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Royal Economic Society and *Oxford University Press* are collaborating with JSTOR to digitize, preserve and extend access to *The Economic Journal*

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

The Economic Journal, **86** (*June* 1976), 243–255
Printed in Great Britain

ON THE INTERPRETATION AND DISAGGREGATION OF GINI COEFFICIENTS[1]

I. INTRODUCTION

The contemporary interest in questions of inequality has resulted in a great many papers concerned with problems of measurement and of interpretation. Meanwhile, and not least as a result of major contributions by Atkinson (1970) and Sen (1973), the traditional framework of Lorenz curves and Gini coefficients has retained its adherents, although Theil (1967) has shown that a measure of inequality based on entropy has advantages. Specifically, Theil's measure can be neatly decomposed so that, in a grouped population, total inequality depends on inequality within groups and inequality between them. No such simple decomposition is available for the Gini coefficient, yet its direct relationship to the Lorenz curve has resulted in persistent attempts to derive a disaggregation which can be used in empirical work. A major concern in the present paper is to further such endeavour.

In outline, the present paper is in three main sections following this introduction. The first section gives an interpretation of the Gini coefficient in terms of the expected value (in the statistical sense) of a game in which each individual is able to compare himself with some other drawn at random from the total population. This provides a basis for the Gini coefficient, which has some novelty and is of potential use in analysis. The second section illustrates this by decomposition of the Gini coefficient in terms of conditional expectations of the value of the game. A development of this result shows that the decomposition can be expressed as involving three terms. The first depends on the Gini measure of inequality within subgroups of the total population. The last depends on differences in the average value of income[2] between groups. In between, there is a term which depends on the extent to which the income distributions for different groups overlap each other. This same disaggregation has been derived previously by Bhattacharya and Mahalanobis (1967) using a more conventional approach in terms of absolute differences. The justification for the present paper must rest, therefore, on the simplicity of the proofs and the novelty of the approach. One implication of the decomposition is to obtain a measure of the extent to which the total can be accounted for simply by consideration of differences in means between groups.

Finally, in section four, some potential applications of the decomposition algorithm to the analysis of migration and discrimination are suggested. Thus,

[1] I am grateful to a number of people who have sent comments and related papers in response to an earlier draft, and would particularly like to thank Clive Bell and John Fei in this context, as well as Nanak Kakwani, Mark Leiserson, Farhad Mehran, Gus Ranis, Amartya Sen and Wouter van Ginneken. Of course I am fully responsible for this final version.

[2] It is convenient to discuss the subject-matter in terms of income distributions if only because income inequality has been, and remains, the main motivation for work in this area. However, the results can of course be applied to a wide range of issues and are not specific to income.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

while reservations about the use of Gini coefficients to measure inequality may remain, it is hoped that the present paper may encourage empirical analysis directed towards understanding causes of inequality, albeit within the limitations of the Gini framework.

II. AN INTERPRETATION OF THE GINI COEFFICIENT

The Gini coefficient of inequality among a set of numbers[1] $y_1, y_2, \ldots, y_n$ can be expressed in various ways. One such way which brings out the relationship of the coefficient to interpersonal comparisons is:[2]

$$G = \frac{(1/2n^2) \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|}{(1/n) \sum_{i=1}^{n} y_i}, \tag{1}$$

i.e. the ratio of the mean absolute difference between all pairs (y_i, y_j) to twice the mean level of the variable y. To develop this, note that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j| = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \max(0, y_i - y_j), \tag{2}$$

where max $(0, y_i - y_j)$ denotes the higher of the two things within the bracket. It follows that

$$G = \frac{(1/n^2) \sum_{i=1}^{n} \sum_{j=1}^{n} \max(0, y_i - y_j)}{(1/n) \sum_{i=1}^{n} y_i}. \tag{3}$$

This last expression is the basis for the interpretation of the Gini coefficient on which our subsequent analysis rests.

Consider the following statistical game. For each individual we conduct an experiment. First, some income, y, is selected at random from the population of incomes $y_1, \ldots, y_n$. If the income selected is greater than the actual income of the individual then he can retain the value selected: otherwise he retains his actual income. Clearly, no individual could lose from participating in this experiment; and all individuals apart from the richest would have a mathematical expectation of gaining from it. For individual i the expected gain is given by

$$\frac{1}{n} \sum_{j=1}^{n} \max(0, y_j - y_i) \geqslant 0 \quad \text{for all } i, \tag{4}$$

and if we now average these expected gains over all individuals, i, we obtain an expression which is the numerator of equation (3).

If follows from these arguments that equation (3) can be interpreted as stating that the Gini coefficient is the average gain to be expected, if each

[1] The choice of formulation in terms of a discrete set of numbers is a matter of style. All the results in this paper can be derived via the theoretical distribution of a continuously distributed variate y.

[2] This result is proved by Kendal and Stuart (1963), pp. 48–9.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

individual has the choice of being himself or some other member of the population drawn at random, expressed as a proportion of the average level of income. And in these terms the link between the Gini coefficient and interpersonal comparisons is immediate and obvious.[1]

III. DISAGGREGATION BY GROUPS

The average expected gain defined in the previous section can be disaggregated in a variety of ways especially if the population can be divided into a number of interesting groups (e.g. by geographical area). In particular, we can write

$$\text{average expected gain} = \sum_{i=1}^{k} \sum_{j=1}^{k} E(\text{gain}/i \to j) \Pr(i \to j), \tag{5}$$

where the event "$i \to j$" refers to an individual being in population group i and drawing a member of group j to compare himself with in the hypothetical game previously specified. In this notation $E(\text{gain}/i \to j)$ is the average, taken over all individuals in group i, of their expected gain, given that they draw a member of group j to compare with in the game. Since the game specifies that sampling is random, we have

$$\Pr(i \to j) = p_i p_j \quad \text{for all} \quad i, j = 1 \ldots k, \tag{6}$$

where

$$\sum_{i=1}^{k} p_i = 1,$$

so that p_i is the proportion of the population that is in group i, and the total population is divided into k mutually exclusive and exhaustive groups.

These results can be combined in the matrix equation

$$G = (\mathbf{m}'\mathbf{p})^{-1}\,\mathbf{p}'\mathbf{E}\mathbf{p}, \tag{7}$$

where $\mathbf{p}$ and $\mathbf{m}$ are both k-element column vectors. The ith element of $\mathbf{p}$ is the population proportion p_i: the ith element of $\mathbf{m}$ is the average income of individuals in population group i. Hence

$$\mathbf{m}'\mathbf{p} = \frac{1}{n} \sum_{i=1}^{n} y_i, \tag{8}$$

i.e. $\mathbf{m}'\mathbf{p}$ is the average income of the total population. Finally, $\mathbf{E}$ is a $k \times k$ matrix with (i, j)th element given by $E(\text{gain}/i \to j)$ as defined in equation (5).

An empirical illustration of the disaggregation (7) is given in Table 1. This is taken from a 1973 survey of income distribution in Sri Lanka[2] in which the total population of income receivers is classified geographically according to their location in urban, rural or estate areas. The table shows the population proportions, p_i, in each area; the average income, m_i, of each subgroup; and

[1] A symmetric treatment in terms of the minimum of a pair of incomes is, of course, possible. Such an approach would be a development of Sen's (1973) observation, p. 33: "In any pair-wise comparison the man with the lower income can be thought to be suffering from some depression on finding his income to be lower. Let this depression be proportional to the difference in income. The sum total of all such depressions in all possible pair-wise comparisons takes us to the Gini coefficient."

[2] The data are from Central Bank of Ceylon (1974) as reported by Karunatilake (1974).

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

Table 1

Distribution of Income (in Rupees per Month) of Income Receivers in Sri Lanka, 1973, by Geographical Areas

Location of income receiver	Mean income, **m**	Population proportion, **p**	Conditional expectations $E(\text{gain}/i \to j) = E_{ij}$ Estate	Rural	Urban	Conditional expectations, **Ep**	$(\mathbf{m}'\mathbf{p})^{-1}\,\mathbf{Ep}$
Estates	227·4	0·177	84·1	277·4	410·2	266·5	0·60
Rural	457·5	0·648	47·2	169·3	279·3	167·0	0·38
Urban	600·6	0·175	36·9	136·1	240·3	136·9	0·31
	$\mathbf{m}'\mathbf{p} = 441{\cdot}9$	Total 1·000				$\mathbf{p}'\mathbf{Ep} = 179{\cdot}3$	$G = 0{\cdot}41$

the matrix, **E**, of conditional expected gains. From this the vector **Ep** is easily calculated: its elements are the expected gains for groups of individuals in our hypothetical game conditional only on the group in which they are located. As might be expected, these gains are inversely correlated with the group mean incomes, m_i: a member of the poorest group stands to gain most from the chance of being someone else.

Table 2 illustrates the fact that the inverse relationship between the elements of **m** and those of **Ep** does not have to be monotonic. The table is derived from the work of Soltow (1960) on the relationship between income inequality and education using U.S. data for 1956,[1] and shows a break in monotonicity for those who have 13–15 years of schooling. The conditions under which such a break in monotonicity may occur are discussed later in the argument.

Table 2

Distribution of Family Income ($ per annum) in the United States, 1956, by years of Schooling Completed by Head of Household

Years at school	Mean Income, **m**	Population proportion, **p**	Conditional expectations, **Ep**	$(\mathbf{m}'\mathbf{p})^{-1}\,\mathbf{Ep}$	Gini coefficient
< 8	3,170	0·232	4,388	0·89	0·45
8	4,240	0·185	2,564	0·52	0·41
9–11	4,830	0·180	1,873	0·38	0·34
12	5,530	0·233	1,134	0·23	0·33
13–15	6,340	0·077	1,430	0·29	0·37
16+	8,260	0·093	740	0·15	0·39
	$\mathbf{m}'\mathbf{p} = 4{,}930$	Total 1·000	$\mathbf{p}'\mathbf{Ep} = 1{,}972$	$G = 0{\cdot}40$	

A variant of equation (7) is obtained by defining

$$\boldsymbol{\pi} = (\mathbf{m}'\mathbf{p})^{-1}\,\hat{\mathbf{m}}\mathbf{p}, \tag{9}$$

so that the ith element of the column vector $\boldsymbol{\pi}$ is the proportion of aggregate

[1] Soltow's analysis is based on U.S. Government (1956).

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

Table 3

Distribution of Income (in Rupees per Month) of Income Receivers in Sri Lanka, by Geographical Areas

Location of income receiver	Income received, $\mathbf{p}'\hat{\mathbf{m}}$	Share of income, $\boldsymbol{\pi}$	$\hat{\mathbf{m}}^{-1}\mathbf{E} = \mathbf{E}^*$ Estates	Rural	Urban	$\mathbf{E}^*\mathbf{p}$
Estates	40·2	0·091	0·37	1·22	1·80	1·17
Rural	296·4	0·671	0·10	0·37	0·61	0·37
Urban	105·4	0·238	0·06	0·23	0·40	0·23
	$\mathbf{m}'\mathbf{p} = 441{\cdot}9$	Total 1·000				$\boldsymbol{\pi}'\mathbf{E}^*\mathbf{p} = 0{\cdot}41$

income which accrues to members of population group i. With this notation (7) becomes

$$G = \boldsymbol{\pi}'\mathbf{E}^*\mathbf{p}, \tag{7a}$$

where

$$\mathbf{E}^* = \hat{\mathbf{m}}^{-1}\mathbf{E}. \tag{10}$$

Table 3 shows the decomposition of G in terms of equation (7*a*) for the Sri Lanka data used in compiling Table 1.

The matrix $\mathbf{E}^*$ defined in equation (10) is a normalisation of the matrix $\mathbf{E}$ obtained by dividing the elements of each row of $\mathbf{E}$ by the mean income for the corresponding population group. Thus the table shows, for example, that the expected gain for a rural household in being given the option of having the income of an urban household within our game framework is equal to 0·61 *times* the average income of a rural household.

A further implication of the definition of $\mathbf{E}^*$ is that its diagonal elements are themselves Gini coefficients: the ith such element is the expected gain for an individual in group i who draws some (other) member of group i within the game framework, expressed as a proportion of average income in group i. This then is the Gini coefficient of inequality for the subpopulation which is in group i. For example, from Table 3, the Gini coefficient of inequality among estate households is 0·37.

An interpretation of the off-diagonal elements of $\mathbf{E}^*$ can be approached by considering inequality as measured simply by differences in the means, m_i, between subgroups. Thus, given the data in the vectors $\mathbf{p}$ and $\mathbf{m}$, and hence $\boldsymbol{\pi}$, and assuming that all individuals in each group have the same income, m_i, an estimate of G can be obtained from first principles, i.e. by plotting the Lorenz curve and calculating the proportion of the area below the 45° line which is above the curve. If the elements of $\mathbf{m}$ are ordered from smallest to largest, and the same ordering adopted for $\mathbf{p}$ and $\boldsymbol{\pi}$, this comes out mathematically as:[1]

$$\tilde{G} = \boldsymbol{\pi}'(\hat{\mathbf{m}}^{-1}\mathbf{A}'\hat{\mathbf{m}} - \mathbf{A}')\mathbf{p}, \tag{11}$$

where $\tilde{G}$ is an approximation to G; $\mathbf{A}$ is a $k \times k$ matrix with all elements on and below the main diagonal equal to one, and all other elements zero. Hence all

[1] It is relatively easy to show from first principles that $G = \boldsymbol{\pi}'(2\mathbf{A}-\mathbf{I})\mathbf{p}-1$. This can be reduced to $\boldsymbol{\pi}'(\mathbf{A}-\mathbf{A}')\mathbf{p}$ from the fact that $\boldsymbol{\pi}'(\mathbf{A}+\mathbf{A}'-\mathbf{I})\mathbf{p} = 1$. The result (11) follows from the equivalence of $\boldsymbol{\pi}'\mathbf{A}\mathbf{p}$ and $\boldsymbol{\pi}'\hat{\mathbf{m}}^{-1}\mathbf{A}'\hat{\mathbf{m}}\mathbf{p}$.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

Table 4

Inequality in Sri Lanka Due to Differences in Mean Incomes between Geographic Areas

Location of income receiver	Share of income, π	Population proportion, $\mathbf{p}$	$\mathbf{E}_2^* = \hat{\mathbf{m}}^{-1}\mathbf{A}'\hat{\mathbf{m}} - \mathbf{A}'$ Estate	Rural	Urban	$\mathbf{E}_2^*\mathbf{p}$
Estates	0·091	0·177	—	1·01	1·64	0·94
Rural	0·671	0·648	—	—	0·31	0·05
Urban	0·238	0·175	—	—	—	—
	Total 1·000	Total 1·000				$\pi'\mathbf{E}_2^*\mathbf{p} = 0{\cdot}12$

elements of $(\hat{\mathbf{m}}^{-1}\mathbf{A}'\hat{\mathbf{m}} - \mathbf{A}')$ lying below the main diagonal are zero: the expected gain from the opportunity of becoming a member of a poorer group when there is no intra-group variation must be zero. Similarly, the diagonal elements of $(\hat{\mathbf{m}}^{-1}\mathbf{A}\hat{\mathbf{m}}' - \mathbf{A}')$ must be zero. And this result can also be derived as a special case of considering the (i, j)th element of $(\hat{\mathbf{m}}^{-1}\mathbf{A}'\hat{\mathbf{m}} - \mathbf{A}')$ where $i \leqslant j$. Such elements have the form

$$(m_j/m_i) - 1 = (m_j - m_i)/m_i, \tag{12}$$

so they depend simply on mean income differences, expressed as a ratio of group mean incomes. And they must be positive, since by definition group j has a higher average income than group i if $j > i$.

It is important to note at this point that while the result (11) can be derived from first principles as suggested, it can also be obtained directly from equation (7a). If there is no inequality within groups, then the diagonal elements of $\mathbf{E}^*$ (which are the Gini coefficients for within-group inequality) must be zero. Further, elements of $\mathbf{E}$ (and hence of $\mathbf{E}^*$) below the diagonal must be zero since there is no advantage from the chance to have an income from a poorer group. And an element of $\mathbf{E}$ above the diagonal must be of form

$$m_j - m_i > 0, \tag{13}$$

since the gain in the chance to join a richer group must be the difference in group mean incomes in this case. Hence

$$\mathbf{E} = \mathbf{A}'\hat{\mathbf{m}} - \hat{\mathbf{m}}\mathbf{A}' \tag{14}$$

by direct application of the principles for deriving $\mathbf{E}$ in the special case where there is no intra-group variation.

At this point we have an exact disaggregation of G given by (7a); and an expression (11) which depends only on differences in mean incomes, m_i, and which would be exact if there were no intra-group variation. As Table 4 shows, this latter yields a value of $\tilde{G}$ of 0·12 for our Sri Lanka data. Moreover, intra-group variation as measured by the Gini coefficients within groups contributes only 0·18 to the aggregate Gini coefficient of 0·41. Clearly, then, these two components do not account for all inequality as measured by G in this case.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

The third component which is missing depends on whether the distributions for different subgroups of the population overlap. In other words on whether, given $m_j > m_i$, the minimum income within group j is less than the maximum income within group i.

To explore this further we need to consider the exact relationship between the conditional expectations $E(\text{gain}/i \to j)$ as defined in equation (5) and the difference in means $m_i - m_j$. From first principles

$$E(\text{gain}/i \to j) = E(y_j - y_i / y_j > y_i) \Pr(y_j > y_i) \tag{15}$$

and similarly

$$E(\text{gain}/j \to i) = E(y_i - y_j / y_i > y_j) \Pr(y_i > y_j). \tag{16}$$

It is now easily shown[1] that

$$m_i - m_j = E(\text{gain}/j \to i) - E(\text{gain}/i \to j) \tag{17}$$

or, in other words, that the difference between the (j, i)th and (i, j)th elements of the matrix $\mathbf{E}$ is the difference in means $m_j - m_i$. We can therefore write.

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \tag{18}$$

$$= \mathbf{E}_1 + \mathbf{A}'\hat{\mathbf{m}} - \hat{\mathbf{m}}\mathbf{A}' \tag{19}$$

where $\mathbf{E}_1$ is a symmetric matrix with (i, j)th element given by the minimum of the (i, j)th and (j, i)th elements of $\mathbf{E}$.

Thus $$\mathbf{E}^* = \hat{\mathbf{m}}^{-1}\mathbf{E}_1 + \hat{\mathbf{m}}^{-1}\mathbf{A}'\hat{\mathbf{m}} - \mathbf{A}' \tag{20}$$

and $$G = \boldsymbol{\pi}'[\mathbf{E}_1^* + (\hat{\mathbf{m}}^{-1}\mathbf{A}'\hat{\mathbf{m}} - \mathbf{A}')]\mathbf{p}, \tag{21}$$

where $$\mathbf{E}_1^* = \hat{\mathbf{m}}^{-1}\mathbf{E}_1 \tag{22}$$

and the diagonal elements of $\mathbf{E}_1^*$ are the Gini coefficients of inequality within groups. Moreover, as is apparent from (15) and (16), the off-diagonal elements of $\mathbf{E}_1^*$ will be zero if the distributions within groups do not overlap: without overlaps one or other of the probabilities $\Pr(y_i > y_j)$ and $\Pr(y_j > y_i)$ must be zero, and hence the minimum of (15) and (16) must be zero. In the more general case, when overlaps exist, then by definition some members of the richer group (group j) are poorer than some members of the poorer group (group i). Accordingly, the expected gain for such members of group j from the chance of becoming a member of group i will be positive. Since for all other members of group j the expected gain is zero, it follows that the weighted average for group j as a whole is positive, i.e. that the expression (16) is positive even though $m_j > m_i$. Thus the (j, i)th element of $\mathbf{E}$ is positive and, from (17), must be smaller than the (i, j)th element of $\mathbf{E}$. Accordingly, this positive (j, i)th element of $\mathbf{E}$ is the corresponding (i, j)th (and (j, i)th) element of the symmetric matrix $\mathbf{E}_1$.

[1] Given (15) and (16), the result (17) follows directly from the fact that

$$m_i - m_j = E(y_i - y_j / y_i > y_j) \Pr(y_i > y_j) + E(y_i - y_j / y_i < y_j) \Pr(y_i < y_j)$$

and noting that the second term is the negative of the expression (16). It is interesting to note that the sum of the two expressions (15) and (16) is $E(|y_i - y_j|)$.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

The result (21) is our final disaggregation of the Gini coefficient by population subgroups. It should be noted that all its components are non-negative and that it is essentially in two parts: the first arises from variations within groups while the second depends entirely on differences between group means.[1]

At this point we recall that the data in Table 2 show that the elements of **Ep** need not be monotonically related to group mean incomes. The ith element of **Ep** is the average, taken over all individuals in group i, of the mathematical expectation of the gain for each such individual from having the chance of receiving the income of some other population member. It is natural to anticipate that the poorer group i is, as measured by the average income of the group, m_i, then the more its members might expect to gain from our hypothetical statistical experiment. Thus an inverse relationship between elements of **m** and the corresponding elements of **Ep** is intuitively plausible. Table 2 illustrates the fact that this relationship is subject to exceptions.

To clarify the nature of such exceptions, it can be noted that if there is only one individual in each group, then the (i, j)th element of **E** is simply $\max(0, y_j - y_i)$. Since all elements of **p** are equal to $(1/n)$ in this case, it follows that the ith element of **Ep** is $(1/n) \sum_{j=i+1}^{n} (y_j - y_i)$. Hence the ith element of **Ep** exceeds the $(i+1)$th element by $(1/n)\ (y_{i+1} - y_i)$; which must be positive, since individuals are ordered in terms of increasing income. Thus there can be no counter-intuitive ordering of elements of **Ep** in the absence of grouping: elements of **Ep** decrease as individual incomes increase. By extension, if individual incomes are grouped in such a way that there is no overlap between groups then the monotonicity is maintained. Given such grouping, elements of **Ep** must decline as group mean incomes increase. It follows that exceptions such as that observed in Table 2 must derive from overlaps between groups as opposed to variations within them.

The way in which counter-intuitive results may arise, given overlapping groups, can be approached by considering a population divided into two groups. Suppose that the first group has the lower mean and is highly concentrated around it. The second group has the higher mean and wide dispersion. Thus, the first element on the diagonal of **E** is small, while the second element is large. If almost all the population are concentrated in the first group, then only the first column of **E** matters in calculating **Ep**. And the first element of this column can approach zero arbitrarily closely as the concentration in the first group increases. By contrast, the second element of the first

[1] An obvious corollary of the result (21) arises when a Gini co-efficient is calculated from population groups defined by income ranges. Clearly, in this case there are no overlaps between groups, so that the off-diagonal elements of $\mathbf{E}_1^*$ are all zero. If cumulative proportions of income are plotted on a Lorenz diagram against cumulative population proportions, and the points obtained are joined by *straight* lines, then the elementary graphical methods lead to an estimate of G as given by $\bar{G}$ in (11). The straight lines implicitly assume that there is no inequality within groups, i.e. that the diagonal elements of $\mathbf{E}_1^*$ are also zero. By contrast, joining the points on the Lorenz diagram by a smooth curve is one (albeit primitive) method of attempting to capture variations within classes, and hence allows for the contribution of these diagonal elements to the total Gini measure.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

column of $\mathbf{E}$ can be quite large. This element is the expected gain for a group 2 individual who draws a group 1 individual in the statistical game. Most such individuals may expect to gain very little, or nothing, since they will have an income greater than the mean value around which group one is concentrated. However, the wide dispersion of group two will imply a significant number of individuals in that group with income below the central value for group one, especially if the difference in average incomes between the two groups is not large. All individuals in this second category will accordingly have a significant expected gain if given the choice of having a group 1 income. Hence the second element in the first column of $\mathbf{E}$ may not be small, leading to the exceptional result that the first element of $\mathbf{Ep}$ is less than the second.[1]

This informal discussion suggests that three considerations may lead to exceptional ordering of the elements of $\mathbf{Ep}$. If group i is to show a break in monotonicity, then this is most likely when $m_i - m_{i-1}$ is not large; when p_i is small; and when the Gini coefficient for group i is large relative to the Gini coefficients for other groups. Data in Table 2 illustrate the fact that some element of each of these conditions is present for the group defined as having 13–15 years of schooling. This group has $p_i = 0{\cdot}077$ and is the smallest grouping. The mean difference $m_i - m_{i-1}$ of 810 is the second smallest, while the Gini coefficient for the group (0·37) is larger than that for the two groups which rank below it in terms of means.

A result such as (21) is contained in the work of Bhattacharya and Mahalanobis (1967) as noted in the Introduction. However, their analysis does not use the "game" concept developed here or matrix notation. They work entirely in terms of mean absolute differences with the emphasis on the *scalar* contribution of each component to the aggregate Gini coefficient. The result (21) is perhaps, therefore, an extension of theirs if only at the level of psychological novelty.[2]

One obvious implication of (21) is that G cannot be decomposed into a Gini measure of inequality within groups and a Gini measure of inequality between group means; in addition, there are the off-diagonal terms of $\mathbf{E}_1^*$ due to overlaps. This is in effect the point made by Theil (1967) in arguing the superiority of his entropy measure. But we can now suggest that the argument is not

[1] It is, of course, possible to produce an exact mathematical condition for a break in monotonicity to occur. However, none of the various possible arrangements of this condition adds significantly to the interpretation given in the text.

[2] Various other attempts at Gini disaggregation have been made. In the published literature Soltow (1960) working with absolute differences rather than Gini coefficients, separates out what is, in effect, the contribution of the diagonal elements of the matrix $\mathbf{E}_1^*$, i.e. the Gini coefficients of inequality within groups. An unpublished paper by Mehran (1974) goes further, giving an alternative derivation in terms of the three components and extending the decompositon to a two-way population disaggregation. Yet another proof is provided in a private communication from John Fei. Finally, Mangahas (1975) and Rao (1969) provide an alternative formulation which results in separating out the Gini inequality within groups and compounds other effects in an expression of form $(\mathbf{m}'\mathbf{p})^{-1}\,\mathbf{p}'\mathbf{D}\mathbf{p}$, where the matrix $\mathbf{D}$ in the Mangahas version has zeros on and below the main diagonal. This can be reconciled with the present formulation by noting that, since $\mathbf{E}_1$ is a symmetric matrix:

$$\mathbf{p}'\mathbf{E}_1\,\mathbf{p} = \mathbf{p}'\hat{\mathbf{e}}_1\,\mathbf{p} + 2\mathbf{p}'\mathbf{E}_1^{+}\mathbf{p},$$

where $\hat{\mathbf{e}}_1$ is the matrix formed by setting off-diagonal elements of $\mathbf{E}_1$ equal to zero; and $\mathbf{E}_1^{+}$ is formed by setting to zero all elements of $\mathbf{E}_1$ on and below the main diagonal.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

necessarily persuasive. In expected-gain terms there is a simple dichotomy between the part due to differences within groups and the part due to differences between their means. The fact that the former, given by $\mathbf{p}'\mathbf{E}_1\mathbf{p}$, is not a simple average of its diagonal elements does not seem to be so important. Indeed, there are positive advantages in analytic terms which are discussed in the next section. Meanwhile, the crucial question is whether or not the derived statistics indicated by our disaggregation are interesting. And in these terms it is suggested that a decomposition such as in Tables 1–4 does pass the test, not least because of the interpretation in terms of expected gains via interpersonal comparisons which underlies it.

IV. SOME EXTENSIONS

The main result in this paper is contained in equation (21), where it is shown that the Gini index can be expressed as

$$G = \boldsymbol{\pi}'\mathbf{E}^*\mathbf{p}, \tag{23}$$

where $\boldsymbol{\pi}$ is the vector of income proportions over population subgroups, $\mathbf{p}$ is the vector of population proportions in subgroups, and $\mathbf{E}^*$ is a matrix which can usefully be interpreted and decomposed. The interpretation suggested is that the (i, j)th element of $\mathbf{E}^*$ is the gain to be expected by the typical individual in group i if he (or she) could have the free choice of being a member of population group j. In terms of $\mathbf{E}^*$, this gain is expressed as a proportion of average incomes in group i.

One application we have not developed is that the elements of $\mathbf{E}^*$ are an ingredient in the decision to migrate. Here "migration" should be interpreted in terms of changing population group, and not necessarily as simple geographical migration. Thus, in a classification based on years of schooling, the relevant decision is whether to stay at school longer. This brings up a number of further points we need to develop.

For an individual considering migration, the relevant population grouping is their peer group; the population group to be joined as perceived by the individual; and the rest of the population. If the group to be joined all have higher incomes then no problems of intergroup distribution overlaps arise. If this is not so, then the formulation adopted here assumes that the individual can reject the new situation and revert to the initial one. Another way of putting it is that the migration is perceived as being reversible in the sense that the individual will not be worse off as a result. This may well be closer to human psychology than the assumption that the migrant may reckon *a priori* that he may be stuck with an inferior situation. Thus the model does not assume that the potential migrant thinks it is certain he will be better off, but only that he can revert, and hence be at no risk of being worse off.

Of course, to complete a model of migration would require specification of costs as well as benefits, and some dynamic considerations might also need to be taken into account. Meanwhile, it seems potentially fruitful to point out the link between migration and the Gini disaggregation.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

Going further, the disaggregation applied to migration implies freedom of entry into some other population group. This is qualified if costs, such as acquiring more education, are introduced. By implication, such costs and other restrictions on mobility are a cause of observed inequalities in a society: in the absence of the costs mobility would take place on a greater scale and differences in opportunity would be reduced thereby. In practice, there are other barriers to mobility derived from sex or religion, for example. Such discrimination is a further source of inequality. The framework provided here suggests a way in which the importance of such barriers might be quantified. This would be to replace the matrix $\mathbf{E}^*$ in equation (23) by some other, $\mathbf{E}^*_{\mathrm{D}}$ derived from $\mathbf{E}^*$ by setting to zero all elements of $\mathbf{E}^*$ which correspond to expected gains from changes to a different group which are in fact precluded by discriminatory barriers. Such a procedure can be applied to our Sri Lanka data on the assumption that the Estate population is distinct from that in Rural and Urban areas. The result is a reduction in the total Gini coefficient from 0·41 to 0·29. This implies that 0·12 or 29 % of the aggregate Gini can be associated with immobility between Estate workers and others in Sri Lanka. Clearly some comparative studies are needed to know whether this is a large figure or not, and in any event its significance depends on the extent to which the barriers in a particular case are ones which individuals would like to cross. Meanwhile, the illustration serves as a starting point for quantification and indicates the relevance of the framework to such aspects of inequality.

V. SUMMARY AND CONCLUSIONS

This paper shows, first, that the Gini-coefficient measure of income inequality can be interpreted in terms of a simple statistical game. In this game each individual draws an income at random from the distribution of incomes which obtains in the society. If the income so selected is higher than the actual income of the individual, then this higher income is recorded. Otherwise the individual's actual income is recorded. Thus for all individuals except the richest there is a mathematical expectation of a positive gain of recorded income over actual income: for the richest person the expected gain is zero. In section II of the paper it is shown that if these expected gains are averaged over all individuals, then the result, expressed as a proportion of average income, is equal to the Gini coefficient. Hence the latter can be interpreted in terms of the average expected gain from having the option of receiving the income of someone else selected at random.

On the basis of this result an expression is obtained, in section III, for the Gini coefficient within a population which is sub-divided into groups by some criterion, such as location (but which could be income itself). This expression is the sum of three parts, none of which can be negative. The simplest is due to differences in mean incomes between groups, and would be the only component if there was no variation in income within groups. The second and third terms both depend on variation within groups. The distinction between them is this. If there are no overlaps between the income ranges in different groups

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

then the third term is zero: otherwise the third term is positive, reflecting the fact that there is now a positive expectation of gain in the statistical game for poorer members of a rich group who draw richer members of a poorer group as a result of the random sampling. Meanwhile, irrespective of whether groups overlap, variation within them contributes to overall inequality and leads to the second term in the decomposition which depends on the Gini coefficients of inequality within each group.

This decomposition is partially illustrated by reference to the work of Soltow (1960) on earnings according to years of formal education. A full illustration is provided by analysing some income distribution data for Sri Lanka in which households are disaggregated by geographical location.

Finally, in section IV, it is suggested that the disaggregation may have particular relevance to studies of migration and discrimination. Here migration is to be understood as movement from one group to another and therefore includes non-geographical mobility, such as a change in educational status. If such mobility is based on a difference between the costs and benefits of a shift from one group to another, then it may be reasonable to equate the benefits with the expected gain for individuals in the former group who draw an individual from the latter group in the statistical game. This leaves outstanding the specification of costs, but meanwhile suggests that the costs represent barriers to mobility which would otherwise take place. If these barriers take the extreme form of total discrimination, whereby mobility from some groups to others is precluded, then a measure of their importance can be obtained by setting to zero all expected gains which arise within the game framework from comparisons between individuals who are in fact separated by such total discrimination.

GRAHAM PYATT

Development Research Center, World Bank:
on leave from University of Warwick

Date of receipt of final typescript: October 1975

REFERENCES

Atkinson, A. B. (1970). "On the Measurement of Inequality." *Journal of Economic Theory*, vol. 2.

Bhattacharya, N. and Mahalanobis, B. (1967). "Regional Disparities in Household Consumption in India." *Journal of the American Statistical Association*, vol. 62, no. 317 (March).

Central Bank of Ceylon (1974). Department of Economic Research, *Survey of Sri Lanka's Consumer Finances* 1973. Colombo.

Karunatilake, H. N. S. (1974). "Changes in Income Distribution in Sri Lanka." The Japan Economic Research Centre, Seminar on Income Distribution, Employment and Economic Development in Southeast and East Asia, December (Mimeograph).

Kendal, M. G. and Stuart, A. (1963). *The Advanced Theory of Statistics*. Vol. I. *Distribution Theory*, 2nd ed. London: Griffin.

Mangahas, M. (1975). "Income Inequality in the Philippines: A Decomposition Analysis." *World Employment Programme Research Working Papers*. Population and Employment Working Paper, no. 12, I.L.O., Geneva.

Mehran, F. (1974). "Decomposition of the Gini Index: A Statistical Analysis of Income Inequality." I.L.O., Geneva, September (Mimeograph).

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms

Rao, V. M. (1969). "Two Decompositions of Concentration Ratio." *Journal of the Royal Statistical Society*, series A, vol. 132, part 3.

Sen, A. K. (1973). *On Economic Inequality*. Oxford: Clarendon Press.

Soltow, Lee (1960). "The Distribution of Income Related to Changes in the Distributions of Education, Age, and Occupation." *Review of Economics and Statistics*, vol. XLII.

Theil, H. (1967). *Economics and Information Theory*. Amsterdam: North-Holland.

U.S. Government (1956). "Income of Families and Persons in the United States: 1956." *Current Population Reports*, series P-60, no. 27.

This content downloaded from
201.110.122.183 on Sun, 14 Aug 2022 03:12:19 UTC
All use subject to https://about.jstor.org/terms