International Journal of Economics and Finance; Vol. 10, No. 3; 2018
ISSN 1916-971X E-ISSN 1916-9728
Published by Canadian Center of Science and Education

The Gini Coefficient: An Application to Greece

Augustine C. Arize[1], Paraskevas Bakarezos[2], Ioannis N. Kallianiotis[3], John Malindretos[4] & John Phelan[5]

[1] Department of Economics and Finance, College of Business, Texas A&M University, USA

[2] Agricultural Bank of Greece, Republic of Greece

[3] Department of Economics and Finance, Arthur Kania School of Management, University of Scranton, USA

[4] Department of Economics, Finance and Global Business, Cotsakos College of Business, William Paterson University, Wayne, New Jersey, USA

[5] Department of Economics, College of Business, University of New Haven, USA

Correspondence: John Malindretos, Department of Economics, Finance and Global Business, Cotsakos College of Business, William Paterson University, Wayne, New Jersey, USA. E-mail: MALINDRETOSJ@wpunj.edu

Received: August 15, 2017 Accepted: December 9, 2017 Online Published: February 28, 2018

doi:10.5539/ijef.v10n3p205 URL: https://doi.org/10.5539/ijef.v10n3p205

Abstract

The Gini coefficient is a measure of income inequality. In this study we show that it needs to be adjusted to be a correct measure of income inequality. The result is that decomposition is possible even without the interaction effect. The requirement however, is that there are data on individual incomes. Secondly, the approach is applied to Greece. Third, there is the last section indicating extensions.

Keywords: Decomposition, Gini coefficient, overlapping component of Gini coefficient, Greece

1. Introduction

The measurement of personal income inequality has been for long a matter of interest in economics. There exist several measures of inequality and Stark (1972) gives a comprehensive list. The most widely used measure of inequality is probably the Gini coefficient. The investigation of its various characteristics on both the theoretical and the empirical level has also been long with major contributions by Atkinson (1970) and Sen (1973). An interesting development occurred when Bhattacharya and Mahalanobis (1967) and later Rao (1969) introduced the decomposition approach. According to it the population of income-receiving units is divided into two groups and consequently inequality is distributed to disparities in income between groups. Pyatt (1976) analyzed the same disaggregation in matrix form and his method corresponds directly to that used here.

The purpose of this paper is to present the process of decomposition and its relation to the Gini coefficient with emphasis on the explanation of the nature of the overlapping component (or interaction effect) which arises during decomposition and it is usually treated as an awkward by-product of this process. It is shown that exact decomposition without the presence of the overlapping effect is possible even when group income distributions are overlapping under condition that data on individual incomes are available. The problem is purely mathematical and it arises when decomposition is accompanied by replacement individual income comparisons with differences in mean group income between pairs of groups. In this case, the overlapping component can be completely separated from the other two, the “within” and “between” inequality components. Furthermore, the case of grouped data is considered that does not allow exact separation of the component.

Exact decomposition without the presence of the overlapping effect, in the case of exact data, is the subject matter of the first section. The next section deals with the same case of exact data, the condition under which the overlapping component arises and its mathematical isolation and estimation. The third section considers the more realistic case of non-availability of exact data that makes exact separation of the overlapping component impossible and a combination of estimating techniques is suggested to approximate it. The forth section consists of an illustration of all these results that are applied to data on income in Greece 1962-88. The application is methodologically similar is not to analyze the components and the trend of inequality. The fifth section outlines briefly the necessary extensions of this analysis. The sixth section is a short summary with a presentation of the conclusions of the paper.

In conclusion, this paper was developed and written in the hope of contributing to further understanding of the process and nature of decomposition thus facilitating analysis of income inequality situation.

2. Exact Decomposition of the Gini Coefficient

The most suitable formulation of the Gini coefficient for the purposes of the paper is (Note 1):

$$G = \frac{\sum_{d=1}^{P}\sum_{f=1}^{p}|X_d - X_f|}{2*p^2*\bar{X}} \tag{1}$$

Where p is total number of income-receiving units,

X_d is income from the d^{th} unit,

d,f = 1 … p, and

$\bar{X}$ is average income.

This formulation is therefore based on inter-unit income comparisons and it requires data on individual incomes, called here exact data as opposed to grouped data (Note 2). The numerator in (1) can be considered as a norm of the matrix of absolute differences *A* of dimension (p*p) (Note 3).

It is possible, however, to consider instead of p number of units k number of groups of populations p_1, $\sum_{i=1}^{k} p_1$ =p. In this case, *A* can be partitioned into k^2 sub-matrices $A_{i,j}$, i,j= 1…k. The sub-matrices along the diagonal of *A* are symmetric and of dimension (p_i*p_i), i= 1…k. The rest of the sub-matrices are of dimension (p_i*p_i), i,j= 1…k. i≠j, and depending on the division into groups, none, some, or all can be non-square. All this means that *A* can be decomposed into k^2 sub-matrices each having the following norm (Note 4):

$$A_{\mathrm{ij}} = \sum_{\mathrm{r}=1}^{p_i} = \sum_{\mathrm{w}=1}^{p_j}|X_{ir} - X_{jw}| \tag{2}$$

Where i,j = 1…k

$$\sum_{i=1}^{\mathrm{k}} = \sum_{j=1}^{k} A_{i,j} = \sum_{d=1}^{\mathrm{p}} = \sum_{f=1}^{p}|X_d - X_f|$$

The result of this procedure is the decomposition of *A* which can be interpreted as follows : The sub-matrices $A_{i,j}$, i=j, contain as elements comparisons between pairs of incomes within each group. The rest of the sub-matrices $A_{i,j}$, i≠j, consist of elements which are the comparisons of pairs of income between units of i^{th} and j^{th} groups. The decomposition of *A* can be used to obtain expressions analogous to (1). These, for the sub-matrices $A_{i,j}$, i=j, can be interpreted as "within" group Gini coefficients:

$$G_{ij} = \frac{A_{i,j}}{2*p_1^2*\overline{X_1}} \quad , \qquad \mathrm{i = j} \tag{3}$$

For the sub-matrices $A_{i,j}$, i≠j, the following expressions can be defined which are interpreted as "between" group Gini coefficients measuring inequality between pairs of income groups (Note 5):

$$G_{ij} = \frac{A_{i,j}}{2*p_1*p_j*\overline{X_{1J}}} \ , \quad \mathrm{i \neq j} \tag{4}$$

Where $\overline{X_{\iota J}} = \frac{\overline{X_\iota} + \overline{X_J}}{2}$

Using the results so far, matrix G can be defined with elements G_{ij} from expression (3) and (4) and of dimension (k*k). Weighting these G_{ij} by group income and population proportions results in an alternative additively decomposable expression for G in (1) as follows:

$$\mathrm{G} = \boldsymbol{\pi}'\mathrm{Gm} \tag{5}$$

Where $\boldsymbol{\pi}'$ is a row vector (1*k) of aggregate group income proportions (Note 6) and m is a column vector (k*1) of group population proportions. Expression (5) is an exact and symmetric decomposition of the Gini coefficient (Note 7, Note 8).

3. Decomposition and Overlapping (Note 9)

The result of the previous section is an exposition of the well-known fact that exact decomposition of the Gini coefficient is possible when exact data are available, and it is included to preserve continuity of the discussion. Exact decomposability, however, is not usually encountered in practice or it might be computationally awkward even when it does. It would be of advantage to replace as many A_{ij} in (2) with a summary measure. One possible replacement is to express A_{ij}, i≠j in terms of differences in group mean incomes, i.e.

$$A_{ij} = \sum_{r=1}^{p_i} = \sum_{w=1}^{p_j}|X_{ir} - X_{jw}| = p_i * p_j|\overline{X_1} - \overline{X_J}| + R,\ \mathrm{i} \neq \mathrm{j} \tag{6}$$

Where R is a remainder and the explanation of its nature is the subject matter of this section.

The investigation of the relationship between the left hand (LHS) and right hand (RHS) side expressions in (6) requires again analysis based on exact data as in the last section but before taking absolute values. The ordering of individual incomes in ascending (or descending) order according to size of income results in skew symmetric sub-matrices A_{ij}, i=j, that is elements above the diagonal are the negatives of the elements below it. The elements of these diagonal sub-matrices are not affected by the RHS replacement in (6). The elements of the sub-matrices A_{ij}, i≠j, are also placed symmetrically along the diagonal of the partitioned A. The difference between A_{ij}, i=j, and A_{ij}, i≠j, is that the elements of the latter might or might not be neatly separated with respect to their signs and that depends on overlapping between pairs of group income distributions. In the case of non-overlapping distributions each A_{ij}, i≠j, above the diagonal consists of elements that are all negative, while each A_{ij}, i≠j, below it (from now on denoted by A_{ij}) consists of the same elements placed symmetrically but all positive (Note 10). When the distribution are overlapping A_{ij} and A_{ij}, i≠j, contain both positive and negative elements not at random but in the following order: The pair-wise comparisons in A_{ij} between members of group i and members of group j for which $X_{ir} < X_{jw}$, come out negative while the rest of the comparison turn out positive since $X_{ir} > X_{jw}$ (and this is the result of overlapping).The corresponding comparisons in pairs in A_{ij} yield the same elements but of opposite sign (Note 11). It is the existence of elements of both sign within each A_{ij}, i≠j, that creates the overlapping effect when LHS expression in (6) is replaced by the summary measure on the RHS combined with the existence of absolute values. The addition during the computation of the double sum on the LHS of (6) is performed after talking absolute values and that means that the outcome of each comparison (or element of A_{ij}) counts positively regardless of its sign. However, the RHS is computed differently since calculation of $\overline{X_1}$ and $\overline{X}_J$ requires addition before taking absolute values and differences of opposite signs off-set against each other (Note 12).

In other words, the "between" group Gini coefficient can be computed using as numerator in (3) either the LHS or the first term in the RHS in (6). Since these two expressions are always equal, because they are not computationally equivalent, it is necessary to make them equal under all conditions. This can be done in 2 ways: firstly, the addition on the LHS expression can be performed before talking absolute signs thus allowing offsetting due to opposite signs on the LHS also and in this case R need not be include in the RHS (Note 13). Secondly, the order of operations during computation on the double sum in (6) remains unchanged while the RHS expression in (6) is increased by the amount $R = 2\sum_{u=1}^{n_{ij}}|L_u|$, u = 1…$n_{ij}$, n_{ji} being the number of differences of opposite sign in A_{ij} and i = 1…k. This amount is, therefore, computed as the absolute value of the sum of all differences of opposite sign (Note 14) multiplied by two (Note 15) thus making all the differences on the RHS expression in (6) of the same sign (negative in our case). This can become clear by focusing temporarily on one comparison only, $X_{ir} – X_{jw}$ = L, $X_{ir} > X_{jw}$, by adding to L the quantity of -2L. Performing this for all the differences within each A_{ij} and separating the sum $2\sum_{u=1}^{n_{ij}}|L_u|$, we allow addition to be performed in the RHS expression in (6) before taking absolute values, while at the same time, we compensate for offsetting, since that is the only way to estimate $\overline{X_1}$, $\overline{X}_J$ (Note 16). In the case of non-overlapping distributions L=0. The first way of making the two expressions in (6) equal by taking absolute values after performing the addition (the equivalent of subtracting the above quantity from the LHS expression) is not correct taking expression (3) into consideration since it underestimates the Gini coefficient by losing the quantity from off-setting due to positive and negative elements. The second way, of adding separately the same quantity to the RHS expression in (6) is in accord with expression (3) since no difference is lost only they are separated. Consideration of all A_{ij} in A results in the following decomposition of the Gini coefficient:

$$G = \pi'[B + X + V]m \tag{7}$$

Where B is a (k*k) diagonal matrix, each diagonal element being a group Gini coefficient calculated according to expression (3). X is a (k*k) symmetric matrix with all diagonal elements equal to zero and all other elements equal to Gini coefficients measuring inequality between all pairs of the k number of groups due to differences in mean group incomes its calculation based on the RHS of expression (6). That is:

$$X_{ij} = \left|\overline{X}_\iota - \overline{X}_J\right| \Big/ \left(\overline{X}_\iota - \overline{X}_J\right) \tag{8}$$

Where V is again a symmetric matrix with all diagonal elements equal to zero while all other elements are Gini coefficients their computation based on overlapping elements as follows:

$$R = V_{ij} = \frac{2\sum_{u=1}^{n_{ij}}|L_u|}{2} * p_i * p_j * \overline{X_{ij}} \tag{9}$$

These coefficients have no special meaning but they are used to make expression (7) equal to (5). In the case of non-overlapping distributions V contains zeros as its elements (Note 17).

4. Decomposition and Grouped Data

It was shown in the previous section that the Gini coefficient can be neatly decomposed into three components as given in (7). However, a number of practical problems may arise with this analysis. The basic difficulty has to do with the fact that exact decomposition is feasible only when data on individual incomes are available. The first problem arises even when data of this kind are available, and it relates to the dimension of matrix A and the difficulty of handling it mathematically even with the aid of computers. One way to avoid this problem, on the empirical level, is to consider average sub-group incomes in the place of individual incomes. In this case the dimension of A is drastically reduced. This formulation creates another well-known problem however, since it underestimates the Gini coefficient. The reason for this is twofold: Firstly, it replaces sub-matrices along the diagonal of A with zeros thus destroying a number of differences and therefore making the numerator in (1) smaller than it would be in case of individual differences for the same data. Secondly, it further underestimates the numerator in (1) since it replaces blocks of elements in each off-diagonal sub-matrices. A_{ij} by their means thus creating an overlapping effect if the income distributions of the sub-groups, which are now being replaced by means multiplied by sub-group populations, are overlapping. The second problem relates to the search of methods to improve estimation in the case of grouped data. Although this is not the objective of this paper, an attempt is made to develop a method with this problem in mind. (See next section for its application). Kakwani and Podder (1973) suggest an efficient method that can be used to estimate Gini coefficients. (The elements of B of the previous section in our case). Then the elements of X can be estimated using mean group incomes. Following this the Gini coefficient for the total population (denoted by RG) can be computed independently using the method suggested by Kakwani and Podder. Then another approximate Gini coefficient for the total population can be calculated using a modified form of (7) that does not include V, that is:

$$MG = \pi'[B + X]m \tag{10}$$

Subtracting MG from RG we obtain an estimate of the total overlapping component as a residual.The final step is to allocate the overlapping component among elements of V in (7). The straightforward way to do this is to compute overlapping elements using mean sub-group incomes in the place of individual incomes and then normalize taking as base the overlapping component computed as a residual (Note 18).

This method results in an exact decomposition of the Gini coefficient but at the same time it "dumps" all other estimation inefficiencies onto the overlapping component since it is estimated as a residual. The practical problem is, therefore, to obtain some idea of magnitude of these inefficiencies relative to the magnitude of the overlapping component. A simple way is to compare the value of the component before and after normalization. This is done with the application in the sections (See Fig. 3) (Note 19).

5. Decomposition of Inequality of Greece 1962-88

All the results of the previous sections were applied to a data base (Note 20) on incomes in Greece for the years 1962-88. The data allowed estimation of mean incomes and populations of sub-groups. The sub-groups were allocated to three groups: The group of pensioner including seventeen sub-groups, the group of wage-salary earners with twenty-three sub-groups and finally the group of entrepreneurs containing thirty sub-groups. Taxes and transfer payments are accounted for in the estimation of incomes. Since these are grouped data, they are subject to all the problems described in the last section. The decomposition of the Gini coefficient is done following the two methods outlined in the previous section of this paper (Note 21). The first method, of using sub-group mean differences, multiplied by the corresponding subgroup populations, in the place of individual incomes, is applied for two reasons: firstly, because it is exact and it serves the purpose of verifying empirically all the results of this paper and secondly, because it can serve as an empirical comparison to the more efficient method described in the last section. Table 1 shows the application of expressions (3) and (4) to the data. This is called "the method of comparison between sub-group means." Table 2 shows the results from the application of expression (7) using again the method of comparison between sub-group means. The lower case letters denote the weight Gini coefficient i.e. g11, g22, g33. Columns (1), (2), (3) are the elements of $\prod'_{=}[B]\,m$ in (7) where B is a (3*3) matrix with its diagonal elements equal to G11, G22, G33, columns (1), (2), (3) of table 1. Columns

(5), (6), (7) are again the elements of X in (7) weighted by group population proportion and income shares, $\prod'[X]\, m$, while columns (9), (10), (11) show the weighted overlapping elements $\prod'[V]\, m$. Column (4), (8), (12) are the sums of column (1), (2), (3) respectively (the "within" inequality component), (5), (6), (7) (the "between" component), and (9), (10), (11) (the overlapping component). The sum of column (4), (8), (12), in column (13) equals the total (not decomposed) population Gini coefficient, denoted by G, that was calculated using expression (1). The same data are used to decompose the Gini coefficient following the procedure suggested in the last part of section III. This method is called "combined method of estimation, regression and comparison between subgroup means," and all the coefficients relating to it are labeled by the prefix R. Table 3 is analogous to table 2 only the second method is used now. Column (1), (2), (3) show the weighted elements of the "within" component where the elements of *B* are calculated using the method suggested by Kakwani and Podder. It should be noted here that all regression coefficients used to calculate the elements of *B* are statistically significant. Column (5), (6), (7) include the same entries with the corresponding columns of table 2. Total population Gini coefficient, RG, column (14) is computed independently using the method suggested by Kakwani and Podder. Columns (12), (13) show that normalized and estimated overlapping components correspondingly. It is interesting that none of the normalized overlapping elements, column (9), (10), (11) comes out negative and that fact could be considered to be an empirical verification of the discussion in section III.

Figure 1 is a graphic exposition of how the three components from the decomposition vary among one another and also between each other according to the two methods of estimation. Figure 2 shows how the total population Gini coefficients, estimated by two methods, vary together. It is also observed empirically that the first method of estimation generally underestimates inequality (as suggested in section III), that is G < RG, except for year 1968, and it is more volatile. Figure 3 shows how the estimated and normalized overlapping elements vary together.

Table 1. Exact decomposition of the Gini coefficient into within and among group Gini Coefficients for Greece 1962-1988 (method of comparison between sub-group means)

year	G11 (l)	G22 (2)	G33 (3)	Gl2 (4)	Gl3 (5)	G23 (6)
l962	0.526	0.399	0.372	0.523	0.654	0.483
l963	0.459	0.398	0.369	0.5l3	0.662	0.479
1964	0.476	0.389	0.351	0.5 l9	0.664	0.469
1965	0.47l	0.368	0.315	0.509	0.637	0.430
1966	0.474	0.367	0.328	0.515	0.643	0.430
1967	0.454	0.355	0.335	0.493	0.627	0.424
1968	0.459	0.380	0.351	0.505	0.627	0. 435
1969	0.488	0.370	0.369	0.518	0.646	0.441
1970	0.50 l	0.357	0.345	0.514	0.631	0.420
1971	0.507	0.347	0.333	0.526	0.645	0.414
1972	0.510	0.335	0.382	0.54 1	0.094	0.450
1973	0.512	0.292	0.302	0.543	0.687	0.410
1974	0.541	0.273	0.312	0.537	0.673	0.402
1975	0.519	0.295	0.332	0.531	0.662	0.397
1976	0.494	0.268	0.281	0.526	0.660	0.370
1977	0.508	0.293	0.344	0.536	0.663	0.387
1978	0.495	0.249	0.265	0. 501	0.631	0.338
1979	0.518	0.260	0.297	0.484	0.606	0.361
1980	0.486	0.256	0.374	0.478	0.639	0.395
1981	0.447	0.239	0.342	0.444	0.591	0.355
1982	0.435	0.281	0.322	0.474	0.591	0,348
1983	0.449	0.308	0.322	0.465	0.556	0.352
1984	0.431	0.290	0.289	0.446	0.557	0.345
1985	0.444	0.298	0.272	0.444	0.529	0.334
1986	0.43 1	0.290	0.264	0.429	0.536	0.341
1987	0.415	0.270	0.267	0.404	0.539	0.350
1988	0.355	0.265	0.273	0.364	0.528	0.371

Note. G11, G22, G33 are calculated using expression (3) Gl2=G21, Gl3=G31. G23=G32 are calculated using expression All these constitute the elements of G in expression

- Calculation includes primary data that are estimates or subject to revision

Table 2. Decomposition of the Gini coefficient into weighted components of within between and overlapping elements for Greece 1962-1988 (method of comparison between sub-group means)

Year	Within Elements G11 (1)	G22 (2)	G33 (3)	Within Component g (4)	Between Elements x12 (5)	x13 (6)	x23 (7)	Between Components x (8)	Overlapping Elements v12 (9)	v13 (10)	v23 (11)	Overlapping Component v (12)	Sum of Component G (13)
1962	0.008	0.261	0.001	0.270	0.070	0.011	0.038	0.119	0.039	0.000	0.011	0.050	0.439
1963	0.007	0.259	0.001	0.267	0.078	0.012	0.038	0.128	0.030	0.000	0.011	0.041	0.436
1964	0.008	0.250	0.001	0.259	0.080	0.013	0.038	0.131	0.031	0.000	0.001	0.042	0.432
1965	0.008	0.234	0.001	0.243	0.084	0.012	0.033	0.129	0.029	0.000	0.011	0.040	0.412
1966	0.008	0.231	0.001	0.240	0.090	0.013	0.033	0.136	0.027	0.000	0.012	0.039	0.415
1967	0.009	0.217	0.001	0.227	0.088	0.014	0.032	0.134	0.031	0.000	0.011	0.042	0.403
1968	0.010	0.228	0.001	0.239	0.090	0.014	0.030	0.134	0.036	0.001	0.014	0.051	0.424
1969	0.012	0.216	0.002	0.230	0.094	0.015	0.032	0.141	0.040	0.001	0.013	0.054	0.425
1970	0.014	0.205	0.001	0.220	0.095	0.015	0.030	0.140	0.042	0.001	0.013	0.050	0.416
1971	0.014	0.196	0.002	0.212	0.104	0.017	0.030	0.151	0.039	0.001	0.012	0.052	0.415
1972	0.013	0.187	0.002	0.202	0.111	0.021	0.041	0.173	0.033	0.001	0.010	0.044	0.419
1973	0.016	0.162	0.002	0.177	0.115	0.021	0.039	0.175	0.030	0.000	0.008	0.038	0.390
1974	0.016	0.148	0.002	0.166	0.107	0.021	0.039	0.167	0.042	0.001	0.006	0.040	0.382
1975	0.016	0.159	0.002	0.177	0.109	0.020	0.036	0.165	0.041	0.001	0.008	0.050	0.392
1976	0.016	0.143	0.001	0.160	0.116	0.021	0.035	0.172	0.035	0.001	0.006	0.042	0.374
1977	0.017	0.155	0.002	0.174	0.120	0.020	0.032	0.172	0.036	0.001	0.010	0.047	0.393
1978	0.017	0.131	0.001	0.149	0.115	0.020	0.032	0.167	0.032	0.001	0.005	0.038	0.354
1979	0.021	0.133	0.001	0.155	0.097	0.019	0.031	0.147	0.051	0.001	0.007	0.050	0.361
1980	0.019	0.129	0.002	0.150	0.106	0.023	0.037	0.166	0.039	0.001	0.007	0.047	0.363
1981	0.019	0.120	0.002	0.141	0.101	0.020	0.030	0.151	0.038	0.001	0.006	0.045	0.337
1982	0.021	0.136	0.001	0.158	0.126	0.020	0.023	0.160	0.032	0.001	0.010	0.043	0.370
1983	0.025	0.146	0.001	0.172	0.117	0.018	0.020	0.155	0.043	0.001	0.011	0.055	0.382
1984	0.026	0.133	0.001	0.160	0.113	0.020	0.023	0.156	0.043	0.001	0.008	0.052	0.368
1985	0.029	0.134	0.001	0.164	0.104	0.018	0.020	0.142	0.055	0.001	0.009	0.005	0.371
1986	0.030	0.128	0.001	0.150	0.100	0.020	0.024	0.144	0.055	0.001	0.006	0.002	0.365
1987	0.032	0.112	0.001	0.145	0.093	0.023	0.028	0.144	0.056	0.000	0.004	0.000	0.349
1988	0.031	0.103	0.001	0.135	0.079	0.025	0.032	0.136	0.057	0.000	0.004	0.001	0.332

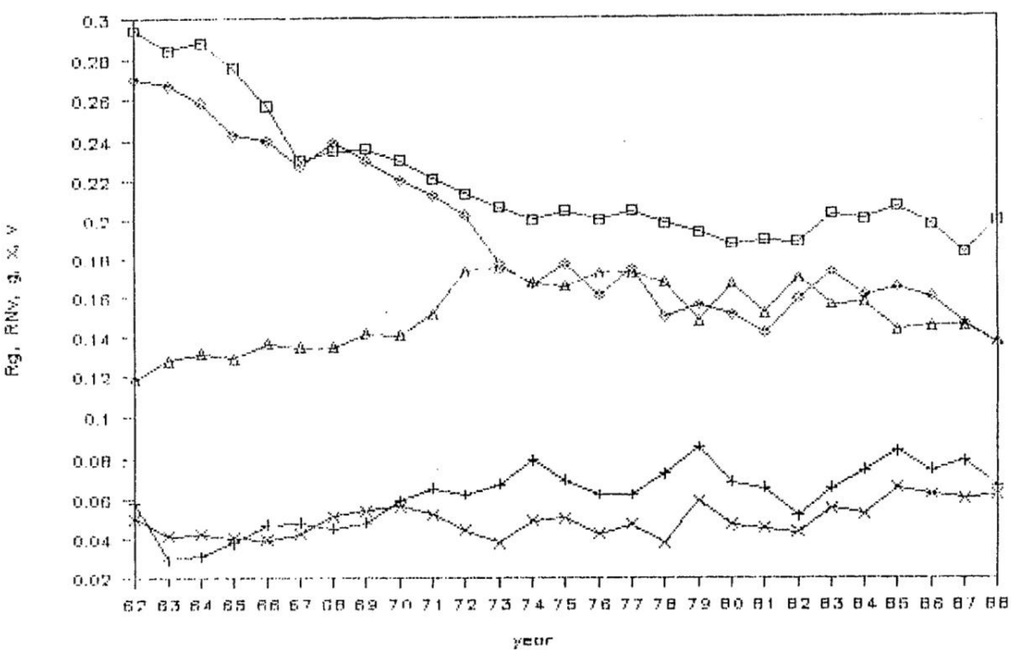

Figure 1. Gini coefficient components

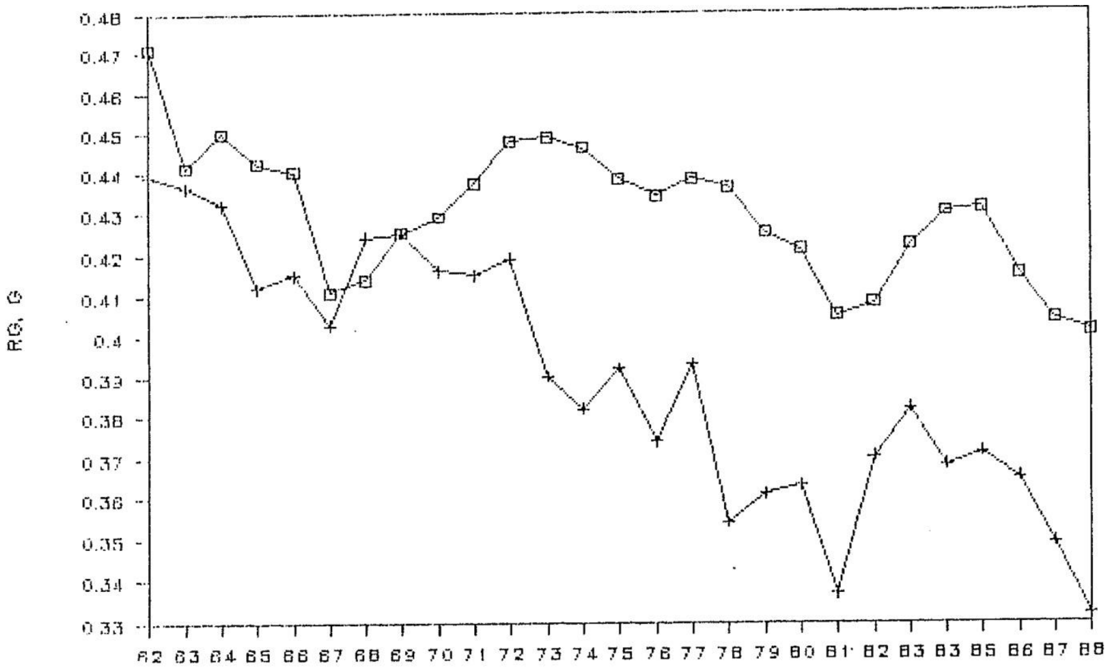

Figure 2. Total population Gini coefficients

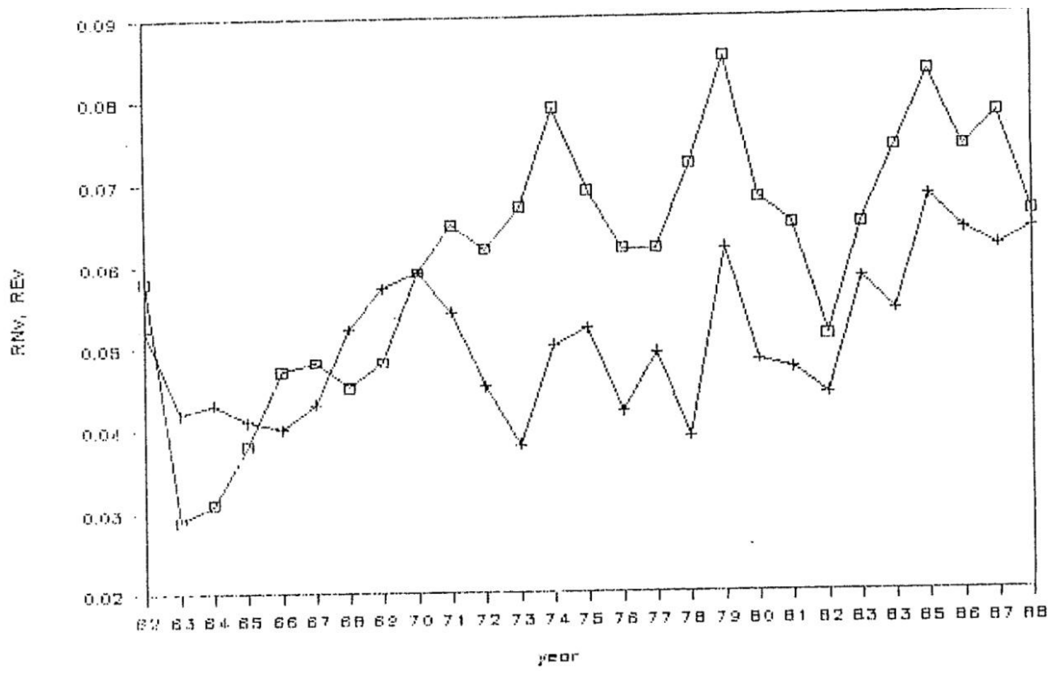

Figure 3. Normolized and estimated overlapping

6. Possible Extension (Note 22)

The purpose of this work is, as already stated, the explanation of the nature of the overlapping component. This required in-depth analysis from the basic level which is the framework used by Bhattacharya and Mahalanobis (1967) as well as that used by Pyatt. The analysis, however, must be brought to higher levels to be in accord with its current status. To this end it can be extended in at least two respects.

The first aspect relates to generalization of the Gini indices as presented in Donaldson and Weymark (1980). They present a class of relative inequality indices characterized by a single parameter and the class of social evaluation functions defined by this class. They also present a corresponding class of absolute indices. The framework of our analysis, in the case of exact decomposition, seems to correspond to their absolute inequality indices (it depends on income differentials). Accordingly, the introduction of "between" inequality, in the case of grouped data, corresponds to their relative indices (only income shares are considered). It seems, therefore, necessary to investigate how the process of generalization is affected by decomposition. This refers to the additional restrictions that need to be placed on the generalization form of the index, in conjunction with the restrictions necessary to satisfy usual conditions an inequality index must satisfy as the principle of population, transfer, and income homogeneity.

The second aspect relates to the analysis introduced by Shorrocks (1980). The whole analysis there derives the class of additively decomposable inequality measures under the condition that population sub-groups are disjoint. Since our work explains the nature of the overlapping component it would be advantageous to investigate how the restriction imposed by additive decomposability on the form of the index and on satisfaction of the related principles are affected. The analysis in this respect could be further advanced along the lines of the analysis in Shorrocks (1984) to generalization to just decomposable Gini indices as supposed to only additively decomposable ones.

It is obvious now that the final step could be a synthesis of the two lines of analysis referred to with the finding of this work, i.e. the derivation of a generalized additively or just decomposable Gini index allowing division of the populations considered into sub-groups in any possible way and not just disjoint.

7. Summary and Conclusion

A number of results have been derived in the previous sections of this paper, according to its purpose to analyze the nature of the overlapping component that arises during the decomposition of the Gini coefficient. The first step is to clarify the meaning of decomposition and this is done is section I. The conclusion there is that an exact decomposition of the Gini coefficient without the presence of overlapping is possible in the case of exact (non-group) data. The next step is to investigate the condition under which the overlapping component arises as well as its nature. The conclusion in section II is: firstly, the problem of overlapping appears when the "between" (groups) inequality is computed using mean group incomes in the place of interpersonal comparisons. Secondly, the overlapping component appears exactly because of this replacement and it constitutes a mathematical problem since the order of operations using the method of summing-up absolute values of interpersonal comparisons is not compatible with the order of operation during computation of means. The two methods can be made compatible by compensating for summing-up after taking absolute values in the first case and vise-versa for the second and this compensation is the overlapping component. The consideration of certain problems arising from the use of grouped data is the purpose of section III. The use of grouped data makes estimation of the overlapping component non-exact and two methods of decomposition are suggested in the view of this. The first method treats sub-group income means as individual incomes and it is introduced due to its exactness although it generally underestimates the Gini coefficient. The second method uses more efficient techniques and the overlapping component is treated as a normalized residual. All these results are applied to a data base on incomes in Greece for the period 1962-88. The results of this application are presented in section IV.

In conclusion the Gini coefficient might well be the most widely used measure of disparities among incomes and furthermore its decomposition is an interesting and significant step towards better analyzing in equality in incomes. The problem of the non-well defined overlapping component is a disadvantage of this positive development and clarification of its nature strengthens the power of analysis that decomposition offers.

References

Atkinson, A. B. (1970). On the Measurement of Inequality. *Journal of Economic Theory, 2*, 244-63. https://doi.org/10.1016/0022-0531(70)90039-6

Bakarezos, P. (1984). Personal non-Wealth Income Inequality and Tax Evasion in Greece 1962-1975. *Spoudai, 34*, 233-57.

Bhattacharya, N., & Mahalanobis, B. (1967). Regional Disparity in Household Consumption in India. *Journal of the American Statistical Assiciation, 62*, 143-61. https://doi.org/10.1080/01621459.1967.10482896

Donaldson, D., & Weymark, J. A. (1980). A Single-Parameter Generalization of the Gini Indices of Inequality. *Journal of Economic Theory, 22*, 68-86. https://doi.org/10.1016/0022-0531(80)90065-4

Kakwani, N. C., & Podder, N. (1973). On the Estimation of Lorenz Curves from Grouped Observations. *International Economic Review, 14*, 278-292. https://doi.org/10.2307/2525921

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

Mookherjee, D., & Shorrocks, A. (1982). A Decomposition Analysis of the Trend in U.K. Income Inequality. *Economic Journal*, *92*, 886-902. https://doi.org/10.2307/2232673

Pyatt, G. (1976). On the Interpretation and Disaggregation of Gini Coefficients. *Economic Journal, 86*, 243-55. https://doi.org/10.2307/2230745

Rao, V. M. (1969). Two Decompositions of Concentration Retio. *Journal of the Royal Statistical Society, Series A, 132*. https://doi.org/10.2307/2344120

Sen, A. K. (1973). *On Economic Inequality*. Oxford: Claredon Press. https://doi.org/10.1093/0198281935.001.0001

Shorrocks, A. F. (1980). Inequality Decomposition by Population Subgroups. *Econometrica, 52*(6), 1369-385. https://doi.org/10.2307/1913511

Shorrocks, A. F. (1980). The Class of Additively Decomposable Inequality Measures. *Econometrica, 48*(3), 613-25. https://doi.org/10.2307/1913126

Stark, T. (1972). *The Distribution of Personal Income in the United Kingdon 1949-1963*. Cambridge University Press.

Notes

Note 1. See Kendall and Stuart (1963), pp. 48-9.

Note 2. Grouped data convey information in the form of mean incomes for groups and sub-groups of the total income-receiving population along with the size of the population of these groups. "Exact data" refers to the situation that individual incomes that can be attributed to individual income-receiving units exist.

Note 3. The elements of the first row of A consist of the absolute differences from the pair-wise comparisons between income of the first unit and all units, including own income and so on, for the second and subsequent rows up to the p^{th} row. If the units are arranged in ascending (or descending) order according to size of income, *A* has zeros as diagonal elements and it is skewed symmetric.

Note 4. $A_{i,j}$ (boldface) denotes a sub-matrix while $A_{i,j}$ its norm as defined in (2).

Note 5. It is also possible to use for the calculation of G_{ij} and its symmetric counterpart G_{ij}, since $A_{ij} = A_{ji}$, i≠j, instead of $\overline{X_{ij}}$, $\bar{X}_i$, and $\bar{X}_j$ correspondingly but then $G_{ij} \neq G_{ji}$ although $A_{ij} = A_{ji}$. The end result, however, will be the same as in (5) irrespective of whether $\bar{X}_i$ and $\bar{X}_j$ or $\overline{X_{ij}}$ are used. Only in the first case, it is possible for some of the G_{ji}, i≠j, to turn out greater than unity and that means that they cannot be interpreted as Gini coefficients.

Note 6. It should be noted that $\pi' * \mathrm{m} = 1$

Note 7. In the context of this section "exact decomposition" allows additive recomposition of the Gini coefficient from the "within" and "between" components to the original value, i.e. expression (5) equal expression (1). In the next section this definition is expanded to include the overlapping component as well, i.e. expression (7) equals expression (1).

Note 8. Pyatt decomposes the Gini coefficient in a non-symmetric way similar to that suggested in footnote 5, while his method is to replace interpersonal comparisons with a statistical game. The end result is the same as in (5) only certain of his equivalent to "between" group Gini coefficients are greater than unity.

Note 9. Overlapping between two group or sub-group distributions is defined as follows: Given $\bar{X}_i$, $\bar{X}_j$, the highest observed income(s) in group i is greater than the lowest income(s) observed in group j. Two distributions are non-overlapping when the highest observed income(s) in group i is less than the lowest income(s) in group j. Besides these two cases, it is also possible to have the highest income(s) in group i equal the lowest income(s) in group j. Then the result, in terms of exact decomposition, is the same as in the case of non-overlapping distributions. The final consideration refers to the extreme case which occurs when all individual incomes in group i are equal and also equal to all incomes in group j (see footnote 17).

Note 10. The elements of each A_{ij} can be interpreted as the result of comparisons made by members in the i_{th} group "looking" at income of members in the j_{th} group, $\bar{X}_i < \bar{X}_j$. The elements of the corresponding A_{ij} are the differences of the opposite comparison.

Note 11. The satiation described here sheds light on the relationship between the statistical game suggested by Pyatt and each pair of A_{ij}, A_{ji}. He seems to consider only positive elements of A_{ij} and only positive elements of A_{ji}, since it is possible to work only with the elements of A above or below the diagonal correcting expression (1) by eliminating division by two. The same result can be obtained by considering the opposite statistical game (elements with negative signs would be considered). In the case of non-overlapping distributions the game results in A_{ij} with all elements equal to zero and all positive elements in A_{ji}. In the case of the opposite statistical game situation reverses itself.

Note 12. Computation if the group income means relates to calculation of the double sum in the way shown in (6).

Note 13. That is:

$$\left|\sum_{r=1}^{p_i} = \sum_{w=1}^{p_j}\left(X_{ir} - X_{jw}\right)\right| = p_i * p_j\left|\overline{X_1} - \overline{X_J}\right|$$

Note 14. In our case (of ascending ordering in incomes) differences termed as of opposite sign are positive for each A_{ij} and negative for each A_{ji}. This definition would be reversed in the case of descending ordering.

Note 15. In our case of ascending ordering of incomes the problem is to have:

$$\sum_{r=1}^{p_i} = \sum_{w=1}^{p_j}\left|X_{ir} - X_{jw}\right| = p_i * p_j\left|\overline{X_1} - \overline{X_J}\right| = \left|p_j \sum_{r=1}^{p_i} X_{ir} - p_i \sum_{w=1}^{p_j} X_{jw} - \right|$$

This can be done by adding to RHS expression the quantity R = $2\sum_{u=1}^{n_{ij}}|L_u|$, where $X_{ir} - X_{jw} = L$, $X_{ir} > E_{jw}$ considering elements of A_{ij} (that is a difference of opposite sign), and n_{ij} is the number of differences of opposite signs in A_{ij} (positive in this case). In other words, absolute signs are taken after addition is performed in $\sum X_{ir}$ and $\sum X_{jw}$ since that is the only order of operations that allows estimation of $\overline{X_1}$, $\overline{X_J}$.

Note 16. It is the case, when we have to use mean group income and not individual incomes, for one reason or another, that creates the whole problem. If we were free to use individual incomes, we would use decomposition as given in (5).

Note 17. When the distributions are overlapping X_{ij} in expression (8) will be zero when $\overline{X_1} = \overline{X_J}$, i≠j i.e. "within" inequality is zero since group income means are equal. In this case, this LHS if expression (6) equals R (if we allowed addition in the LHS to be performed before taking signs then inclusion of R is not receive the same amount of income.

Note 18. There is no warranty that the estimated value of the overlapping element will be positive. Nevertheless the procedure of normalization can still be carried out and negative overlapping elements have the same meaning as positive ones, that is they correct the elements of X due to consideration of mean instead of individual incomes.

Note 19. The main result of this section which should be emphasized is that it is not possible to decompose the coefficient exactly in the case of grouped data and the best that can be done is to use some kind of approximation as done here. This problem is not a consequence of the process of decomposition but of the use of grouped data which affects estimation of the coefficient even in the form of expression (1). On the other hand, it is clear that there exists some relation, albeit approximate, between the cases of exact and grouped data since the elements of B in (7) are calculated using some efficient method (the one suggested by Kakwani and Podder in this case) and the elements of X in (7) are the same during approximation, when $\overline{X_1} \neq \overline{X_J}$, .

Note 20. This data base was originally complied by author from a wide variety of sources for the year 1962-75, (1984). The base was re-calculated using an even wider variety of sources an expanded up to 1988 for the purposes of this study.

Note 21. A Personal Computer was used for the calculations. No computer programs for the required work existed so they were developed by the author using LOTUS 123 and its macro commands (LOTUS 123 is a trademark of Lotus Development Corporation).

Note 22. This is the content of my current research.