CIRPÉE
Centre interuniversitaire sur le risque, les politiques économiques et l'emploi

Cahier de recherche/Working Paper **06-02**

On the Decomposition of the Gini Coefficient: an Exact Approach, with an Illustration Using Cameroonian Data

Abdelkrim Araar

Janvier/January 2006

Araar: Département d'économique, CIRPÉE & PEP, Pavillon DeSève, Université Laval, Sainte-Foy, Québec, Canada G1K 7P4; Fax: 1-418-656-7798
aabd@ecn.ulaval.ca

Abstract:
Decomposing inequality indices across household groups or income sources is useful in estimating the contribution of each component to total inequality. This can help policy makers draw efficient policies to reduce disparities in the distribution of incomes using targeting tools. Decomposing relative inequality indices, such as the Gini coefficient, is not a simple procedure since, in many cases, the functional form of inequality indices is not additively separable in incomes. More importantly, for some of the indices on which this decomposition can be performed, the interpretation of the decomposition components is often not well founded. In this paper, we use the Shapley value as well as analytical approaches to perform the decomposition of the Gini coefficient and generalize it, in some cases, to the decomposition of other inequality indices. For the analytical approach, our aim is to extend the same interpretation, attributed to the Gini coefficient, to that of the contribution components.

Keywords: Equity, Inequality, Decomposition, Shapley value

JEL Classification: D63, D64

1 Introduction

Assessing and analyzing the inequality phenomenon implied by income distribution is a topic that is getting a lot of attention from researchers and policymakers. Decomposing inequality by components can help shape adequate economic policies that reduce inequality and poverty. Due to its overwhelming popularity, the Gini coefficient is often used to represent inequality in the society. This study aims, via well-founded methods, to review the decomposition of the Gini coefficient by components as well as to propose some new methods of decomposition. In some cases, the decomposition methods proposed can be generalized to other inequality indices. The two main component types that will be explored are the exclusive sub-groups of population such as rural-versus-urban households, and the income sources.

Two main approaches are used to decompose the Gini coefficient. The first one concerns the implementation of the Shapley value approach. The application of this approach in the decomposition of distributive indices was introduced in studies by Shorrocks (1999). The main usefulness property of this decomposition is the additivity of components that implies an exact decomposition, where residues due to the interaction between components are attributed to components by a linear approximation. The second approach concerns the analytical decomposition. This approach was covered in earlier research [1]. Starting with the interpretation of the Gini coefficient as well as the new perception of the intergroup inequality component, this study proposes an exact analytical decomposition of the Gini coefficient. To decompose the Gini coefficient by income components using the analytical approach, this study proposes other forms of decomposition with respect to components that have a natural interpretation.

The plan of this paper is as follows. In the next section, we present the Shapley value approach and we implement it to perform the decomposition of the inequality indices where components are groups. In the third section, we perform the analytical decomposition of the Gini coefficient where the latter is interpreted as the expected relative deprivation normalized by the average of incomes. The other analytical form that we use is that of the single-parameter Gini coefficient as proposed by Donaldson and Weymark (1980). In the fourth section, we review the analytical decomposition of the Gini coefficient by income components and apply the Shapley approach in the decomposition by income components. New decom-

[1] See Bhattacharaya and Mahalanobis (1967), Pyatt (1976) and Silber (1989) for decomposition by subgroup populations, and Rao (1969), Lerman and Yitzhaki (1985), Podder and Chatterjee (2002) for the decomposition by income sources.

position approaches are also proposed in this section. In section five, we illustrate this study's method using data from Cameroon in 2001. Finally, some concluding remarks are made in section six.

2 Decomposition of Inequality Indices According to the Shapley Approach

2.1 The Shapley value

Applied in several scientific domains, the Shapley approach can serve to perform an exact decomposition of the distributive indices, such as the Gini coefficient in this study's case[2]. The Shapley value is a solution concept often employed in the theory of cooperative games. Consider a set N of n players that must divide a given surplus among themselves. The players may form coalitions (these are the subsets S of N) that appropriate themselves a part of the surplus and redistribute it between their members. The function v is assumed to determine the coalition force, i.e., which surplus will be divided without resorting to an agreement with the outsider players (the $n-s-1$ players that are not members of the coalition S). The question to resolve is: How can the surplus be divided between the n players? According to the Shapley approach, introduced by Loyd. (1953), the value or the expected gain of player k, noted by E_k, is shown by the following formula:

$$E_k = \sum_{\substack{s \subset S \\ s \in \{0, n-1\}}} \frac{s!(n-s-1)!}{n!} MV(S,k) \tag{1}$$

$$MV(S,k) = \left(v\left(S \cup \{k\}\right) - v(S)\right) \tag{2}$$

The term $MV(S,k)$ is the marginal value that the player k generates after his adhesion to the coalition S. What will then be the expected marginal contribution of player k, according to the different possible coalitions that can be formed and to which the player can adhere? First, the size of the coalition S is limited to: $s \in \{0, 1, ...n-1\}$. Suppose that the n players are randomly ordered and we note the order by σ, such that:

$$\sigma = \left\{ \underbrace{\sigma^1, \sigma^2, \cdots, \sigma^{i-1}}_{s}, \sigma^i, \underbrace{\sigma^{i+1}, \cdots, \sigma^n}_{n-s-1} \right\} \tag{3}$$

[2]See Shorrocks (1999).

For each of the possible permutation of the n players, which equals $n!$, the number of times that the same first s players are located in the subset or coalition S is given by the number of possible permutations of the s players in coalition S (that is $s!$). For every permutation in the coalition S, one finds $(n-s-1)!$ permutations for the players that complement the coalition S. The expected marginal value that player k generates after his adhesion to a coalition S is given by the Shapley value. For every position of the factor k (predetermined cuts of the coalition S), there are several possibilities to form coalitions S from the $n-1$ player (that is the n players without the player k). This number of possibilities is equal to the number of combinations, C_{n-1}^{s}.

How many marginal values would one have to compute to determine the expected marginal contribution of a given factor or player k? Because the order of the players in the coalition S does not affect the contribution of the player k once he has adhered to the coalition, the number of calculations needed for the marginal values is[3]: $\sum_{s=0}^{n-1} C_{n-1}^{s} = 2^{n-1}$. If we do not take into account this simplification, we can write the extended formula of the Shapley Value as follows:

$$E_k = \frac{1}{n!} \sum_{i=1}^{n!} MV(\sigma^i, k) \tag{4}$$

where for each order σ of the $n!$ orders, the players k have only one position that determines the coalition to which he can adhere. The term $MV(\sigma^i, k)$ equals the marginal value of adding the player k to its coalition. The properties of the decomposition of this approach are:

- Symmetry, which ensures that the contribution of each factor is independent of its order of appearance on the list of the factors or the sequence.
- Additivity of components.

2.2 Decomposition of the Gini index by household groups

By supposing that household groups represent factors that contribute to the Gini coefficient, the component of group g according to the Shapley approach is equal to what follows:

[3]See the annex A.

$$E_g^S = \frac{1}{n!}\sum_{i=1}^{n!} MV(\sigma^i, g) \tag{5}$$

where σ^i represents the i_{th} possible order of groups and $MV(\sigma^i, g)$ shows the impact of eliminating group g for the order σ^i on the contribution of the set of groups S. A crucial step for this type of decomposition is to determine accurately the impact of eliminating factors (groups, in this case) on the characteristic function v, which is the Gini coefficient. To clarify better this idea, we analyze this by using the average of incomes. In this example, one needs to look at the decomposition of this average, noted by μ, in components A and B, witch are two groups forming the population. The analytical decomposition of the average is written as:

$$E_A = \phi_A \mu_A \tag{6}$$
$$E_B = \phi_B \mu_B \tag{7}$$

where ϕ_g is the proportion of the population of group g. If we suppose that the elimination of one factor - a group - represents the case where we do not take into account those households that compose the group, the decomposition according to the Shapley approach is as follows:

$$E_A^S = 0.5\left[\mu - \mu_B + \mu_A\right] \tag{8}$$
$$E_B^S = 0.5\left[\mu - \mu_A + \mu_B\right] \tag{9}$$

The necessary condition for reconciling the two approaches, such as $E_F = E_F^S$ ($F = \{A, B\}$), is as follows:

$$\frac{\mu_A}{\mu_B} = \frac{\phi_A}{\phi_B} \tag{10}$$

Hence, when specification of the impact of eliminating factors on the characteristic function is done incorrectly, this can lead to unfounded decomposition results. Now, for the simple example above, if one supposes that the elimination of the group g requires simply the subtraction of $\phi_g \mu_g$, the analytical and Shapley approaches are reconciled.

As proposed in earlier research, decomposing inequality into inter and intragroup is useful to check the importance of each of the two components. A pronounced intergroup inequality reflects income disparities across groups. Inversely, if the intergroup components are marginal, disparity across groups is also marginal. At the first stage of the decomposition, we begin by retaining just these two factors, the intra and intergroup inequality and we express total inequality as follows:

$$I = E^S_{inter} + E^S_{intra} \tag{11}$$

The rules for computing the contribution of each factor are:

- To eliminate the intragroup inequality and to calculate the intergroup inequality, $I(\mu_1..., \mu_g)$, we will use a vector of income where each household has the average income of its group, noted by μ_g;
- To eliminate the intergroup inequality and to calculate the intragroup inequality, we will use a vector of income where each household has its income multiplied by the ratio μ/μ_g. With this new income vector, the average of the incomes of each group equals to μ.
- To illuminate the inter and intragroup inequality simultaneously, we will use simply a vector of incomes where each household has the average of incomes.

The order followed to eliminate factors is arbitrary. To remove this arbitrariness, we use the Shapley approach. This decomposition gives us:

$$E^S_{inter} = 0.5\left[I(y) - I(y(\mu/\mu_g)) + I(\mu_g) - I(\mu)\right] \tag{12}$$
$$E^S_{intra} = 0.5\left[I(y) - I(\mu_g) + I(y(\mu/\mu_g)) - I(\mu)\right] \tag{13}$$

Starting from this decomposition, one can perform a second stage of decomposition. Here the intragroup component is decomposed into specific group components. As we can notice from the equation (13), which defines the contribution of the intragroup inequality, this contribution is based on three inequality indices, since $I(\mu) = 0$. To remove the arbitrariness of the sequence of eliminating the marginal contribution of groups to the total intragroup inequality, we use the Shapley approach. The same rule is used for determining the impact of eliminating the marginal contribution of each group, i.e., the intragroup inequality is eliminated when the income of each household is equal to the average of its group. To clarify better the form of this decomposition, assume that there are only two groups, A and B. Starting with equation (13), one can write the formula as follows:

$$E^S_{intra} = 0.5\left[I(y) - I(\mu_A, \mu_B) + I(y_i^A(\mu/\mu_A), y_i^B(\mu/\mu_B))\right] \tag{14}$$

The contribution of the first group A to the total intragroup inequality can be defined by as[4]:

[4]This decomposition is already programmed in the software DAD 4.4

$$E^S_{intra,A} = 0.25 * \Big[\quad \big[I - I(\mu_A, y_B) + I(y_A, \mu_B) - I(\mu_A, \mu_B)\big] + \\ \big[I(\mu_A, \mu_B) - I(\mu_A, \mu_B) + I(\mu_A, \mu_B) - I(\mu_A, \mu_B)\big] + \\ \big[I(y_i^A(\mu/\mu_A), y_i^B(\mu/\mu_B)) - I(\mu, y_i^B(\mu/\mu_B)) + \\ I(y_i^A(\mu/\mu_A), \mu) - I(\mu, \mu)\big] \quad \Big] \tag{15}$$

One can note here that this approach can be generalized and used to perform the decomposition across groups of any of the usual relative inequality indices.

3 The Analytical Approach

Based on the interpretation of decomposing the Gini coefficient, Pyatt (1976) shows that this coefficient can be expressed as just the mean of expected average gains normalized by the average of incomes. The game for every person consists of randomly drawing a given revenue from the population and accepting such revenue if it exceeds what the person has. The form of this decomposition approach is similar to what Bhattacharaya and Mahalanobis (1967) propose. The decomposition of the Gini coefficient into inter and intragroup components raises a legitimate concern. Indeed, the decomposition of this index can generate a residue that is not simple to interpret. Generally, when we suppose that the intergroup inequality represents inequality where each household has the average income of its group, the algebraic decomposition of the Gini index, noted by I, takes the following form[5]:

$$I = \sum\nolimits_g a_g I_g + \bar{I} + R \tag{16}$$

where I_g is the Gini coefficient for group g, $\bar{I}$ is the intergroup inequality component, a_g is the product of population share and income share going to group g. The component R denotes the residue that exists when incomes overlap between groups. In the same way, Shorrocks (1984) concludes that the class of decomposable inequality indices across groups that can be expressed into size, mean and inequality of each group and respect the scalable invariance axiom are just a transformed form of the generalized entropy index. The specificity of the method we propose resides in the perception of the intergroup inequality. Instead of supposing that this component represents inequality where each person has the average

[5] See also the interpretation of Lambert and Aronson (1993).

income of its group, we continue to use directly personal incomes in the interpretation and measurement of the intergroup inequality component. This new approach allows an exact decomposition of the Gini coefficient to be in the following form:

$$I = \sum_g a_g I_g + \tilde{I} \tag{17}$$

where $\tilde{I}$ represents the intergroup inequality component.

3.1 The Gini coefficient and relative deprivation

According to Runciman (1966), the magnitude of relative deprivation is the difference between the desired situation and the actual situation of a person. We define the relative deprivation of household i compared to j as follows [6]:

$$\delta_{i,j} = (y_j - y_i)_+ = \begin{cases} y_j - y_i & if \quad y_i < y_j \\ 0 & otherwise. \end{cases} \tag{18}$$

The expected deprivation of household i equals to:

$$\bar{\delta}_i = \frac{\sum_{j=1}^{N} (y_j - y_i)_+}{N} \tag{19}$$

The Gini coefficient can be written in the following form:

$$I = \sum_{i=1}^{N} \frac{\bar{\delta}_i}{\mu_y N} = \frac{\bar{\delta}}{\mu} \tag{20}$$

This functional form of the Gini coefficient shows that this coefficient is the ratio between the average expected relative deprivation $\bar{\delta}$, and the average of incomes, μ [7]. This simple functional form gives a justifiable interpretation to the contribution of each household to total inequality. Starting from this, the contribution of each household to total inequality depends on its expected relative deprivation. When household k belongs to group g, one can rewrite its average relative deprivation as follows:

$$\bar{\delta}_k = \phi_g \bar{\delta}_{k,g} + \tilde{\delta}_{k,g} \tag{21}$$

[6] See also Yitzhaki (1979) and Hey and Lambert (1980).

[7] See also Araar and Duclos (2003) for the new interpretation of the Gini coefficient.

$$\tilde{\delta}_{k,g} = \sum_{\substack{j=1 \\ j \notin g}}^{N-K_g} \frac{(y_k - y_j)_+}{N} \tag{22}$$

where ϕ_g is the population's share of group g, K_g is the number of households that belong to the group g, $\bar{\delta}_{k,g}$ is the expected relative deprivation of household k at the level of group g and $\tilde{\delta}_{k,g}$ is the expected relative deprivation of household k at the level of its complement group. By rewriting the Gini coefficient, we find:

$$\begin{aligned} I &= \sum_{g=1}^{G} \sum_{k=1}^{K_g} \left[\frac{\phi_g \bar{\delta}_{k,g} + \tilde{\delta}_{k,g}}{\mu N} \right] && (23) \\ &= \sum_{g=1}^{G} \left[\phi_g^2 \frac{\mu_g}{\mu} \frac{\sum_{k=1}^{K_g} \bar{\delta}_{k,g}}{\mu_g K_g} \right] + \sum_{g=1}^{G} \sum_{k=1}^{K_g} \frac{\tilde{\delta}_k}{\mu N} && (24) \\ &= \sum_{g=1}^{G} \phi_g \psi_g I_g + \tilde{I} && (25) \end{aligned}$$

where G is the number of groups and $\tilde{I}$ is equal to the Gini coefficient where the relative deprivation within the group is ignored and ψ_g is the income share of group g. By supposing that the component $\tilde{I}$ represents the intergroup inequality we give a new definition of what represents this component. Here, this component expresses the expected intergroup deprivation normalized by the average of incomes. Without group income overlap, one can write the decomposition as follows[8]:

$$G = \sum_{g=1}^{G} \phi_g \psi_g I_g + I(\mu_g) \tag{26}$$

In this first analytical decomposition approach, we focus on the fact that the interpretation of the Gini coefficient can be based on the relative deprivation. Each components of this decomposition continues to have the same interpretation that

[8] In the case of distribution without overlap, the relative deprivation of a given member of the poor group compared to other m members of the rich group is equivalent to the m differences between mean of the rich group and its income.

the Gini coefficient has. In the following section, this coefficient is reformulated and written in a form that takes into account the rank or the classification of households according to income.

3.2 The single-parameter Gini coefficient

Donaldson and Weymark (1980) propose to generalize the Gini coefficient of inequality. The single-parameter Gini coefficient depends on the ethical parameter, denoted by ρ, that expresses the level of social aversion to inequality. By supposing that incomes are ranked such that, $y_1 \geq y_2 \geq \cdots \geq y_i \geq \cdots \geq y_N$, this coefficient takes the following form:

$$I_\rho = 1 - \frac{\xi_\rho}{\mu_y} \tag{27}$$

where

$$\xi_\rho = \sum_{i=1}^{N} p_{i,\rho} y_i \quad \text{and} \quad p_{i,\rho} = \frac{i^\rho - (i-1)^\rho}{N^\rho} \tag{28}$$

For the ordinary Gini coefficient, the parameter ρ equals 2 and $p_{i,2} = (2i - 1)/N^2$. Note here that, when $\rho > 1$, the weight $p_{i,\rho}$ decreases sharply when the household rank increases. In other words, the weight attributed to the poorest household is relatively higher than the one attributed to the richest household. Despite the fact that the weight $p_{i,\rho}$ depends on the rank of household i, the social welfare function is additively separable on incomes. Hence, we can rewrite this function by using the notation at group level, such that:

$$\xi_\rho = \sum_{g=1}^{G} \sum_{k=1}^{K_g} p_{k,\rho} y_k = \sum_{g=1}^{G} \xi^*_{g,\rho} \tag{29}$$

Where $\xi^*_{g,\rho}$ is the contribution of group g to the social welfare ξ_ρ. By rewriting the single-parameter Gini coefficient, we find:

$$I_\rho = \sum_{g=1}^{G} \left[\psi_g - \frac{\xi^*_{g,\rho}}{\mu} \right] \tag{30}$$

where ψ_g is the income share of group g. With this first analytical form of decomposition, the contribution of group g is given by:

$$E_g = \psi_g - \frac{\xi^*_{g,\rho}}{\mu} \tag{31}$$

It is clear that, according to equation (28), the weight $p_{i,\rho}$, attributed to household i to compute for the inequality at the population level, will be different from its attributed weight when computing for inequality at the group level. By rewriting the contribution of group g to social welfare ξ_ρ, we find:

$$\begin{aligned}
\xi^*_{g,\rho} &= \sum_{k=1}^{K_g} p_{k,\rho} y_k && (32)\\
&= \sum_{k=1}^{K_g} \left({\phi_g}^\rho \pi_{k,\rho} + \tau_{k,\rho} \right) y_k && (33)\\
&= {\phi_g}^\rho \xi_{g,\rho} + \tilde{\xi}_{g,\rho} && (34)
\end{aligned}$$

where $\pi_{k,\rho}$ is the weight attributed to household k for the social welfare function used to compute inequality at group level (i.e. $I_{g,\rho} = 1 - \frac{\xi_{g,\rho}}{\mu_g}$), $\tau_{k,\rho}$ represents the re-ranking impact on weight by including the others groups and $\tilde{\xi}_{g,\rho} = \sum_k \tau_{k,\rho} y_k$. By using the last equation, one can write:

$$\begin{aligned}
I_\rho &= \sum_{g=1}^{G} \left[\frac{\mu_g}{\mu} \left(\phi_g - \frac{\xi^*_{g,\rho}}{\mu_g} \right) \right] && (35)\\
&= \sum_{g=1}^{G} \left[\frac{\mu_g}{\mu} {\phi_g}^\rho \left(1 - \frac{\xi_{g,\rho}}{\mu_g} \right) \right] + 1 - \frac{\sum_{g=1}^{G} \left(\tilde{\xi}_{g,\rho} + \mu_g {\phi_g}^\rho \right)}{\mu} && (36)\\
&= \sum_{g=1}^{G} {\phi_g}^\rho \frac{\mu_g}{\mu} I_{g,\rho} + \tilde{I} && (37)
\end{aligned}$$

For the Gini coefficient, the decomposition can be written as follows:

$$I = \sum_{g=1}^{G} \phi_g \psi_g I_g + \tilde{I} \tag{38}$$

The variable $\tilde{I}$ can be perceived as the component that captures the intergroup inequality. Notice here that the residue does not appear, which is due to our interpretation of the intergroup inequality. With this interpretation, the intergroup component is highly linked to the re-ranking impact of switching from the group to the population level by including the complement group. Again, one can notice that this decomposition is similar to the first decomposition given by equation (25).

Recall that, with the standard analytical approach, the intergroup component concerns inequality when each household has the average income of its group. On the other hand, with our new approach, the perception of the intergroup component is based directly on individual incomes. The following example illustrates this idea. In this example, assume that the two exclusive groups, A and B, compose the total population. Also, suppose that each group is composed of two households and B' represents a potential income distribution for group B. Based on the standard definition of the intergroup component, the intergroup inequality is the same for cases B and B'. However, with the new approach, the intergroup inequality is not the same. This is can be explained and defended by the fact that any feeling of deprivation concerns directly the household instead of the group entity.

Table 1: Illustrative Example I

Household	A	B	B'
1	3	9	1
2	5	11	19

4 Decomposition of the Gini Coefficient by Income Components

4.1 Analytical decomposition

The decomposition of the Gini coefficient by income components is also interesting. This decomposition allows to have a clear idea on how each component contributes to the total inequality. First, one supposes that the sum of K components equals the total income and the amount of component k, noted by s_k, equals or is greater than zero. The analytical decomposition of the single-parameter Gini coefficient can be done by using equation (27) as follows:

$$I_\rho = 1 - \frac{\sum_{k=1}^{K} \sum_{i=1}^{N} p_{i,\rho} s_{k,i}}{\mu} \quad (39)$$

$$= 1 - \frac{\sum_{k=1}^{K} \xi_{k,\rho}^*}{\mu} = \sum_{k=1}^{K} \left(\frac{\mu_k}{\mu} - \frac{\xi_{k,\rho}^*}{\mu} \right) \quad (40)$$

$$= \sum_{k=1}^{K} \psi_k C_{k,\rho} \quad (41)$$

where ψ_k is the income share of component k, $s_{k,i}$ is the level of component k for household i and $C_{k,\rho}$ is the single-parameter concentration coefficient of component k. One can recall here that this straightforward result was found by Rao (1969). Again, one can find easily the same result represented by the equation (41) when the Gini coefficient is expressed by relative deprivation such that:

$$\bar{\delta}_i = \frac{\sum_{j=1}^{N} \left(\sum_{k=1}^{K} s_{k,j} - \sum_{k=1}^{K} s_{k,i} \right)_+}{N} \quad (42)$$

$$= \sum_{k=1}^{K} \frac{\sum_{j=1}^{N} (s_{k,j} - s_{k,i}) * I(y_j > y_i)}{N} = \sum_{k=1}^{K} \bar{d}_{i,k} \quad (43)$$

where $I(y_j > y_i) = 1$ if $y_j > y_i$ and 0 otherwise. By using equation (20), we have that:

$$I = \sum_{k=1}^{K} \frac{\sum_{i=1}^{N} \bar{d}_{i,k}}{N\mu} = \sum_{k=1}^{K} \psi_k C_k \quad (44)$$

From these results, one can conclude that the concentration coefficient and the contribution are positively linked. This direct conclusion can be wrong as reported by Podder and Chatterjee (2002). The simple example to show such error is to consider a component with a constant amount for all households. While this component should have a negative contribution on inequality, its concentration coefficient equals to zero. To remedy this misinterpretation, these authors propose to transform equation (41) and to check the importance of the following:

$$\psi_k(C_k - I) \text{ where } \sum_{k=1}^{K} \psi_k(C_k - I) = 0 \quad (45)$$

Unfortunately, this proposal is again open to criticism since the sum of contributions still equals zero. Comparing the concentration and Gini coefficients just gives us the direction of the contribution. Shorrocks (1988) try to establish four definitions for the contribution of the k_{th} component. These are:

1. The percentage of inequality due to component k alone;
2. The reduction in inequality that would result if this component was eliminated;
3. The percentage of inequality that would be observed if this was the only source of differences in incomes and all other components were allocated evenly; and
4. The reduction in inequality that would follow from eliminating differences in component k.

It is clear that the interpretation of the concentration coefficient is not concordant with any of these definitions. Instead of addressing the question of how this component contributes to total inequality, one can addresses a complementary question, which is; How the component contributes to reduce inequality? It is clear that the constant component does not explain differences in incomes or inequality, and its concentration coefficient equals zero. Furthermore, its contribution resides in decreasing the relative importance of differences. The illusion here is that this constant component does not explain the inequality that exists, but contributes to reduce it. To illustrate this, suppose that there are only two components, where the first one is constant. We have that:

$$I = \frac{\mu_2}{\mu} I_2 \tag{46}$$

Furthermore, one should be careful in interpretating the results of decomposition. Rao (1969) approach allows us to catch the source of the inequality. The comparison between concentration and the Gini coefficients makes it possible to know the direction of marginal contribution even if there is a constant component.

4.2 Extracting the contribution of the constant effect

To show clearly the impact of the constant part of each component, we propose the following decomposition of the Gini coefficient:

$$I = \sum_{k=1}^{K} \left(\underbrace{\psi_k^* C_k^*}_{VE\text{: Variation Effect}} + \underbrace{-(c_k/\mu) I^*}_{CE\text{: Constant Effect}} \right) \tag{47}$$

where definitions of symbols with $(*)$ are similar to those defined above except that we use the translated income components instead of the usual components, i.e. $s_{k,i}^* = s_{k,i} - c_k \ \forall i$ and $c_k = min(s_{k,1}, ..., s_{k,i}, ..., s_{k,N})$.

In the following example, we discuss about the virtue of this form where we suppose that the population is composed of three households and where the total income is composed of four components.

Table 2: Illustrative Example II

Components	$i = 1$	$i = 2$	$i = 3$	VE^a	CE^b	Contribution
A	11	12	13	0.0370	-0.1314	-0.0944
B	0	10	20	0.3703	0	0.3704
C	6	6	6	0	-0.0717	-0.0717
D	4	3	2	-0.0370	-0.0239	-0.0609
Total	21	31	41	0.3703	-0.2270	0.1434

[a]: Variation Effect

[b]: Constant Effect

Example: $A^* = [0, 1, 2]$

- A: With this component all households have a constant amount of 11. This contributes to a decrease the relative variation observed in total income. The constant effect is higher than the variation effect. This can be explained by the fact that the maximal (absolute) variation of component A is 2.
- B: This component explains the main variation of total income and has a pure variation effect.
- C: This component simply decreases the relative variation of incomes.
- D: The two effects of this component contribute to reduce inequality.

One can note here that this decomposition is similar to Rao's decomposition if $c_k = 0 \ \forall k$.

4.3 Showing the ranking effect

Starting from equation (41), the use of the concentration coefficient instead of the Gini coefficient for each component is implied by the interaction effect between components. To clarify this, one can write what follows:

$$I = \sum_{k=1}^{K} [\psi_k I_k + \psi_k (C_k - I_k)] \tag{48}$$

In the case where each component gives the same rank of households as total income, the ranking effect $\psi_k(C_k - I_k)$ equals zero and we can write:

$$I = \sum_{k=1}^{K} \psi_k I_k \tag{49}$$

Generally, the importance of the interaction effect can be estimated by the ratio, $IE = \frac{\sum_k \psi_k |C_k - I_k|}{I}$.

4.4 Interpreting the marginal contribution of income components

At this stage, we propose to shed light again on the marginal contribution of each component to the Gini coefficient. For this purpose, we assume that the marginal contribution represents the variation in the Gini coefficient implied by adding the k_{th} component to the set of complement components. Based on equation (41), we can write:

$$I - \bar{I}_k = \Delta_k = \bar{\psi}_k (\bar{C}_k - \bar{I}_k) + \psi_k (C_k - \bar{I}_k) \tag{50}$$

where:

- $\bar{I}_k$: Gini coefficient excluding component k.
- $\bar{C}_k$: Concentration coefficient excluding component k.
- $\bar{\psi}_k = 1 - \psi_k$: income share of complement components.

This explicit form gives us more information on the nature of each contribution. To better exhibit the advantages of this form, the following cases are presented.

Constant component: If the component k, is constant, we have that:

$$\Delta_k = -\psi_k \bar{I}_k \quad (51)$$

This implies that the total impact is just the mean effect(ME), that depends on the importance of the income share of this constant component.

Ranked component: If component k has the same power of ranking households as the complement income, then:

$$\Delta_k = \psi_k(I_k - \bar{I}_k) \quad (52)$$

Two mechanisms that explain the impact of adding component k to the complement part on the Gini coefficient:

1- Mean Effect : $-\psi_k \bar{I}_k < 0$
2- Inequality Effect: $\psi_k I_k > 0$

Hence, for this special case, we have that $\Delta_k \leq 0$ if $I_k \leq \bar{I}_k$. The inverse conclusion is also true.

Non-ranked component: This is the usual case. To check the direction of the impact, we can write the following condition:

$$\Delta_k > 0 \Rightarrow \frac{\psi_k(C_k - \bar{I}_k)}{\bar{\psi}_k(\bar{I}_k - \bar{C}_k)} > 1 \quad (53)$$

If the main part of the interaction effect with the complement, expressed by $(\bar{C}_k - \bar{I}_k)$ in equation (50), is less important, the difference $(C_k - \bar{I}_k)$ determines the importance of the impact.

4.5 Shapley decomposition

One can use the Shapley approach to estimate the contribution of each source. As we mentioned previously, a crucial step in carrying out this decomposition is to determine the impact of the elimination of components on the Gini coefficient.

Proposal 1:
Replace the component k by μ_k for each household if the former component is

eliminated (this refers to the elimination of the inequality of component k).

The analytical and the Shapley approaches give the same results when the ranking of households based on each component and the ranking based on total income are the same[9].

$$E_k = \psi_k I_k \quad (54)$$

As a general rule, this first proposal is most appropriate when interaction effect between components is null. Otherwise, this proposal can be seriously criticized. This is because the proposed procedure for eliminating components does not take into account the rank of this component conditional to that of the total income. That is, the re-ranking power of the component for the complement is always neglected. This can give conflicting results. The following example illustrates this clearly .

Table 3: Illustrative Example III

Components	Household_1	Household_2	Household_3	Absolute Contribution
A	10	20	30	0.1818
B	3	2	1	0.0000
Total	13	22	31	0.1818

While component B should have a negative contribution since it contributes to reduce relative deprivation of total incomes, the results show that this component does not contribute to explaining the total inequality.

Proposal 2:
Replace the component k by zero for each household if this component is eliminated.

The results of this proposal can be criticized. Basing on the last illustrative example, the contributions of the two components are equal. From the conclusions of these two proposals, it appears that the analytical approach remains the most

[9]See the annex B when the marginal contribution of component is the same whatever the set of coalition S or the order σ.

convincing in explaining the contribution of income components. As a general rule, when the interaction between factors represents the main part of the characteristic function (the Gini coefficient in this case), the Shapley decomposition is not the most appropriate decomposition method.

5 Illustration Using Cameroonian Household Survey Data

To illustrate how these proposed approaches perform the decomposition of the Gini coefficient by groups and by income components, we use the Cameroonian Household Survey (ECAM II: Enquête Camerounaise Auprès des Ménages) conducted by in the National Institute of Statistics in 2001. This is a national survey with a sample of about 11,000 households selected randomly using two stages in the urban areas and three for the rural areas. We use total expenditures per-adult equivalent as the indicator of well-being at the household level. This indicator is the total expenditures of a household divided by the equivalence scale, which is 1 for each adult and 0.5 for each child[10]. The three groups that we retained are households who live in urban area, in semi-urban area and in rural area respectively.

From results exposed in table (4), one can remark that inequality decreases from the urban to the rural areas. This result is not surprising since, as a general rule, the decrease in the absolute variability of income is higher than that of the average. The other remark concerns the importance of the intergroup inequality where this component represents about 64 percent of total inequality. In table (5), we expose the same decomposition where the standard analytical approach is used and where the residue that emerged from the overlap is maintained.

By comparing results of table (4) and table (5), one can notices that the overlap part is significant and represents about 7 percent of total inequality. In table (6), we expose this decomposition with the Shapley approach. Moreover, the importance of intra and intergroup differs from what was found using the analytical approach. Nevertheless, the relative importance of the intragroup inequality for each group remains practically the same compared to what was found using the analytical approach.

To illustrate the decomposition of the Gini coefficient by income components,

[10]This represents the conventional equivalence scale that was used widely in others studies concerning the Cameroonian data.

the total expenditures per adult equivalent on food and nonfood components are used. We start by performing this decomposition via the Rao's approach. As showed in table (7), the nonfood component explains approximately two-thirds of the total inequality. This result is not surprising since the poorest household should increase its share of expenditures on food. This implies a decrease in variability or deprivation for this component. On the other hand, the Shapley decomposition is used again and the results are presented in table (8). Here one notes that these results are close to the ones obtained using the analytical approach. This is explained by the low impact of interaction between components.

Table 4: Analytical Decomposition

Group	S-Gini	Population Share	Income Share	Absolute Contribution	Relative Contribution
Intragroup	0.1488	—	—	0.1488	36.15%
Intergroup	0.2628	—	—	0.2628	63.85%
Uuban	0.4122	0.3484	0.5281	0.0759	18.43%
Semi-urban	0.3322	0.0819	0.0852	0.0023	00.56%
Rural	0.3204	0.5697	0.3867	0.0706	17.15%
Total	**0.4115**	**1.0000**	**1.0000**	**0.4115**	**100.00%**

Table 5: Standard Analytical Decomposition Approach

Component	Absolute Contribution	Relative Contribution
Intragroup	0.1488	36.15%
Intergroup	0.1965	47.78%
Residue (Overlap)	0.0663	16.07%
Total	**0.4115**	**100.00%**

Table 6: Shapley Decomposition Approach

Group	Absolute Contribution	Relative Contribution
Urban	0.1268	30.81%
Semi-urban	0.0209	5.07%
Rural	0.1373	33.37%
Intragroup	0.2850	69.25%
Intergroup	0.1265	30.75 %
Total	**0.4115**	**100.00%**

Table 7: Decomposition by Expenditure Components (Rao's approach)

Component	Concentration Coefficient	Income Share	Absolute Contribution	Relative Contribution
Food exp.	0.3046	0.4373	0.1332	32.36%
Non food exp.	0.4946	0.5627	0.2783	67.64 %
Total	**—**	**1.0000**	**0.4115**	**100.00%**

Food exp: Total expenditures on food by equivalent adult

Non food exp: Total expenditures on non food by equivalent adult

Table 8: Decomposition by Expenditure Components (Shapley's Proposal_1)

Component	Absolute Contribution	Relative Contribution	Ranking Effect
Food exp.	0.1366	31.19%	-0.0231
Non food exp.	0.2749	66.81 %	-0.0163
Total	**0.4115**	**100.00%**	**-0.0394**

Food exp: Total expenditures on food by equivalent adult

Non food exp: Total expenditures on non food by equivalent adult

Ranking Effect: $\frac{\mu_k}{\mu}(C_k - I_k)$ see section 4.3

6 Conclusion

The decomposition of the Gini coefficient by groups or income components continues to be an attractive exercise for researchers. Exploring determinants of inequality by showing the importance of each component is the challenge in the pursuit to create well-founded policies against inequality. Based on the results in this study, one can conclude that the analytical approach can give convincing results on the contribution of components if the these are well interpreted. In general, the Shapley approach can be used to assign rationally the interaction effect to components. This allocation has a linear form and implies an exact decomposition. When the interaction effect is less important, this decomposition does not interfere with the main analytical results. The illustrative examples of the decomposition of the Gini coefficient by groups show that the Cameroonian rural areas contribute less than the urban areas to total inequality in Cameroon. For the decomposition by expenditure components, the nonfood component explains about two-third of the total inequality.

References

ARAAR, A. AND J.-Y. DUCLOS (2003): "An Atkinson-Gini family of social evaluation functions," *Economics Bulletin*, 3, 1–16.

BHATTACHARAYA, N. AND B. MAHALANOBIS (1967): "Regional Disparities in Household Consumption in India," *Journal of the American Statistical Association*, 62.

DONALDSON, D. AND J. WEYMARK (1980): "A Single Parameter Generalization of the Gini Indices of Inequality," *Journal of Economic Theory*, 22, 67–86.

HEY, J. D. AND P. J. LAMBERT (1980): "Relative Deprivation and the Gini Coefficient: Comment," *Quarterly Journal of Economics*, 95, 567–573.

LAMBERT, P. J. AND J. R. ARONSON (1993): "Inequality Decomposition Analysis and the Gini Coefficient Revisited," *Economic Journal*, 103, 1221–27.

LERMAN, R. AND S. YITZHAKI (1985): "Income Inequality Effects by Income Source: A New Approach and Applications to the United States," *Review of Economics and Statistics*, 67, 151–156.

LOYD., S. (1953): *A Value for n-Person Games, in Contributions to the Theory of Games II (Annals of Mathematics Studies)*, Princeton University Press.

PODDER, N. AND S. CHATTERJEE (2002): "Sharing the National Cake in Post Reform New Zealand: Income Inequality Trends in Terms of Income Sources," *Journal of Public Economics*, 86, 1–27.

PYATT, G. (1976): "On the interpretation and disaggregation of Gini coefficients," *Economics Journal*, 86, 243–255.

RAO, V. M. (1969): "Two Decompositions of Concentration Ratio," *Journal of the Royal Statistical Society*, 132, 418–25.

RUNCIMAN, W. G. (1966): *Relative Deprivation and Social Justice: A Study of Attitudes to Social Inequality in Twentieth-Century England*, Berkeley and Los Angeles: University of California Press.

SHORROCKS, A. F. (1984): "Inequality Decomposition by Population Subgroups," *Econometrica*, 52, 1369–1385.

——— (1988): *Aggregation Issues in Inequality Measurement, in Measurement Economics*, Heidelberg: Physica-Verlag.

——— (1999): "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value," Tech. rep., Mimeo. Department of Economics, University of Essex.

SILBER, J. (1989): "Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality," *The Review of Economics and Statistics*, 71, 107–115.

YITZHAKI, S. (1979): "Relative Deprivation and the Gini Coefficient," *Quarterly Journal of Economics*, 93, 321–324.

ANNEX A: Binomial Theorem of Newton

Newton discovered a formula for $(a + b)^n$ that would work for all values of n, including fractions and negatives:

$$(a+b)^n = \sum_{s=0}^{n} C_n^s a^{n-1} b^n \quad \forall (a,b) \in \Re, n \in N \tag{A.1}$$

Raising $(a + b)$ to the power n is equivalent to multiplying n identical binomials $(a + b)$. The result is a sum where every element is the product of n factors of type a or b. The terms are thus of the form $a^{n-p}b^p$. Each of these terms is obtained a number of times equal to C_n^p, which is how many times one can choose p elements among n. When $a = b = 1$, one will have:

$$(1+1)^n = \sum_{s=0}^{n} C_n^s = 2^n \tag{A.2}$$

Hence, one can conclude that:

$$\sum_{s=0}^{n-1} C_{n-1}^s = 2^{n-1} \tag{A.3}$$

ANNEX B: Decomposition of the Total Index According to the Shapley Approach

When the marginal contribution of the factor k, $MV(S, k) = \bar{x}$, is constant for any order or coalition S, the Shapley value of factor k is as follows**:**

$$\begin{aligned} E_k &= \sum_{\substack{s \subset S \\ s \in \{0, n-1\}}} \frac{s!(n-s-1)!}{n!} \bar{x} \\ &= \sum_{s=0}^{n-1} C_{n-1}^s \frac{s!(n-s-1)!}{n!} \bar{x} \\ &= \sum_{s=0}^{n-1} \frac{(n-1)!}{s!(n-s-1)!} \frac{s!(n-s-1)!}{n!} \bar{x} \\ &= \sum_{s=0}^{n-1} \frac{1}{n} \bar{x} = \bar{x} \end{aligned} \tag{B.1}$$