

Shapley Inequality Decomposition by Factor Components: some Methodological Issues

Mercedes Sastre* and Alain Trannoy**

*Universidad Complutense de Madrid.

** THEMA, Université de
Cergy-Pontoise

Abstract

The purpose of the paper is to focus on three methodological issues regarding "the Shapley inequality decomposition" by factor components which cannot be solved from a theoretical point of view. First, should we use either *zero income decomposition*, or *equalized income decomposition*? Second, should we favour the Nested-Shapley or the Owen decomposition? Third, can we structure the set of income components in some bliss tree? The empirical evidence displayed by results of several variants of the Shapley decomposition using the LIS database regarding the British and American income distributions help us to propose an answer to these questions: confine the analysis on gross income decomposition and select the equalized Nested Shapley method. In the absence of an ideal tree we propose three trees among which a choice has to be made.

This document was completed when Mercedes Sastre was visiting THEMA . The support of the European Commission through the Contract n ERBFMRXCT980248 is grateful acknowledged. We thank LIS for computing assistance and Frederic Chantreuil and Steve Jenkins for helpful comments. This article is forthcoming in *Living Standards, Inequality and Taxation*. P. Moyes, C.Seidl et A.Shorrocks eds, Kluwer 2000.

1. Introduction

Assessing how different income sources affect income inequality certainly appears to be an important issue for interpreting economic trends and designing economic policy. However, complete agreement does not exist in the literature as to the best way to measure exactly how an income source contributes to inequality. The traditional approach to the problem has consisted in using inequality indices that are decomposable by factor components. In view of the shortcomings of standard methods, Chantreuil-Trannoy (1999) and Shorrocks (1999) have proposed applying some concepts of cooperative games theory. Despite its internal consistency and attractive interpretation in terms of *marginal contribution*, using the Shapley decomposition in empirical applications raises some dilemmas that cannot be solved solely on theoretical grounds. The purpose of this paper is to focus on these methodological issues concerning Shapley decompositions. An inductive approach proves to be very useful in order to reveal the alternatives applied economists face. The empirical evidence provided by the results obtained from several Shapley decomposition variants using British and American income distributions contained in the LIS database have helped us propose a way to solve three dilemmas.

- 1) According to the Shapley decomposition, the contribution of any given source of income to overall inequality can be interpreted as the expected marginal impact of the factor when such an expectation is made over all possible sequences of elimination. This marginalist interpretation of the Shapley rule can be considered in at least two possible ways. Firstly, taking the differences between overall inequality and inequality into account if we eliminate the income component whose effect on aggregate inequality is being assessed in all possible elimination sequences (*zero income decomposition*). The second implies eliminating inequality from that source of income (*equalized income decomposition*). Although the second calculation seems to be more in keeping with inequality measurements, the *a priori* reasons that could favor either one or other approach are not that clear. Empirical analyses can thus help us to clarify the situation and choose between both methods. An important shortcoming of the Shapley decomposition rule is that it does not satisfy the principle of independence of the aggregation level. That is to say, the contribution to inequality of a given income source depends on how the rest of the income components are treated. In this regard, our results obtained by the *zero* approach show that the contribution of income sources to overall inequality is highly volatile. In addition, they are much more dependent on the level of aggregation than those obtained through the *equalized* method.
- 2) Two different approaches derived from cooperative games theory have been proposed in the literature to improve Shapley inequality decompositions regarding the independence requirement. The first is the Nested Shapley (Chantreuil-Trannoy 1999) and the other is the Owen decomposition (Chantreuil-Trannoy 1997 and Shorrocks 1999). Both methods require defining a hierarchical or income level structure in which overall income is made up of a set of primary factors. Each of these is then divided into a group of secondary factors. This approach can of course be extended to a sequence of source subgroup partitions. It is obviously better to assume that such partitions are given exogenously and that they are relevant from an economic viewpoint for the exercise to be meaningful.

In both decompositions, the contribution to overall inequality of a secondary factor is dependent on the treatment given to the primary factor it belongs to. Nevertheless, it is independent of the remaining factors' disaggregation. Although the Nested Shapely and Owen rules satisfy this milder independence requirement, the interpretation of source contributions differs in both models. The Owen contribution needs to consider the coalitions of elementary sources to aggregated sources to which they do not belong. Hence, the Owen contribution

intrinsically considers some subsets whose meaning is doubtful. The Nested-Shapley decomposition therefore seems to be more adequate.

- 3) Once the Nested-Shapley procedure is chosen, a hierarchical income structure (or income tree) that allows a meaningful *marginalist* interpretation of the decomposition rule needs to be defined. The economic relevance of such a tree is of course crucial. We show that the structure of economic links between income sources cannot be easily reduced to a tree. An ideal tree is consequently difficult to imagine. Nevertheless, the differences between the alternatives are not that great. We thus propose three trees to choose from.

This paper is organized as follows. Section 2 points out the superiority of the *equalized income decomposition* over the *zero income decomposition* regarding dependence on the degree of disaggregation. We then discuss the relative merits of the Nested Shapely and Owen decompositions in Section 3. Section 4 is dedicated to the problem of choosing a hierarchical income structure. It also offers a discussion on the merits of several trees that can appear to be obvious candidates. The last section summarizes the results obtained and offers some concluding remarks. The Appendix contains a detailed description of the data and some complementary results for British and American income distributions.

2 Equalized versus zero Shapley decomposition

An income distribution among a set of individuals $N=\{1,\dots,i,\dots,n\}$, according to a set of income sources $K=\{1,\dots,j,\dots,k\}$ can be described as a matrix $\mathbf{X}=(x_i^j)_{n \times k}$. In what follows we will not restrain x_i^j to be positive. Let us assume that a specific inequality index $I:R^n \rightarrow [0,1]$ has been selected¹. The Shapley decomposition is able to give the contribution to inequality of any subset of sources. A subset of sources would be denoted S , the set of admissible subsets being the power set 2^K and s the number of elements in S .

Regarding the definition of the Shapley decomposition, at least two calculations seem sensible. They differ in the treatment of components not included in the considered subset. In the first one, defined as the zero income inequality decomposition, the components not included in S are removed. In the second calculation, which leads to what is called *equalized income inequality decomposition* by Chantreuil-Trannoy (1999) the inequality from all components not in S is removed.

More precisely in the first calculation the knowledge of the income distribution helps us to build a distribution of income among subsets of sources $y: 2^K \rightarrow R^n$ such that for all $S \in 2^K, S \neq \emptyset$,

$$y(S) = \left(\sum_{j \in S} x^j_1, \dots, x^j_n \right) \tag{1}$$

and $y(\emptyset)=0$ by convention.

¹ We should notice that the Shapley decomposition calculations require agreement not only with the ordinal meaning of an inequality index but also with its cardinal one.

In the second calculation the distribution of income among subsets of sources is obtained by equalizing complementary sources, i.e. we define: $y^e: 2^K \rightarrow \mathbb{R}^n$ such that $y^e(\emptyset)=0$, and for all $S \in 2^K, S \neq \emptyset$,

$$y^e(S) = \left(\sum_{j \in S} x^{j_1} + \sum_{j \notin S} \mu(x^j), \dots, \sum_{j \in S} x^{j_n} + \sum_{j \notin S} \mu(x^j) \right) \quad (2)$$

where $\mu(x^j)$ is the mean income from source j .

The contribution of source j according to the zero income Shapley decomposition is given by:

$$Sh_j(K, X, I) = \sum_{S \subset K, j \in S} \frac{(s-1)!(k-s)!}{k!} [I(y(S)) - I(y(S - \{j\}))] \quad (3)$$

According to this formula, the contribution of any given factor to overall inequality can be interpreted as the expected marginal impact of the factor when the expectation is taken over all the possible elimination sequences. This decomposition rule satisfies some desirable properties. It leads to a perfect² symmetric (in the sense that the contribution assigned to any factor does not depend on the way in which factors are labeled) decomposition, and it is sensitive to the choice of inequality index. The same comment remains valid for the contribution of source j according to the equalized income Shapley decomposition defined by:

$$Sh_j^e(K, X, I) = \sum_{S \subset K, j \in S} \frac{(s-1)!(k-s)!}{k!} [I(y^e(S)) - I(y^e(S - \{j\}))] \quad (4)$$

The distinction between both calculations seems transparent, and can be appreciated regarding the treatment of an equally distributed component of income. Considering relative inequality indices, eliminating an equally distributed source will increase inequality, and the Shapley contribution calculated according to zero income decomposition will be negative, suggesting an equalizing effect of an equally distributed income source. This can be considered as a reasonable property of the decomposition rule. However, this property is not accomplished by the contribution based on the equalized income decomposition, which assigns a null impact to equally distributed sources³. In spite of the desirable property presented by the former, the latter calculation seems more in the spirit of inequality measurement.

By way of example the values of the relative contribution (the contribution divided by the overall inequality) for the two calculations present marked differences. The empirical exercise has been done using data from the 1995 United Kingdom Family Expenditure Survey⁴, and using the Gini

² The sum of the contributions add up to the amount of inequality.

³ For absolute inequality indices, the contribution of an equally distributed source is always zero. See Chantreuil and Trannoy (1999).

⁴ Available through the Luxembourg Income Study. For more information about this database which allows comparisons of income distributions among different countries see <http://lissy.ceps.lu>. Results for the United States

index as inequality index⁵. The data available allow us to disaggregate United Kingdom households income into three main components: (i) Earnings (**E**): i.e. wages and self-employment income, which represent about 71.8% of overall household income. (ii) Capital income (**C**), representing 11.4% of total income. (iii) Transfers (**T**): including all kind of public and private transfers received for households. The income share of this source is 16.8%. This is only one of all possible partitions, which is very popular in empirical works. According with equation (3) in order to calculate the zero Shapley contribution of the income sources we need first to calculate the inequality for each of the following eight subsets of components S: (E), (C), (T), (E∪C), (E∪T), (C∪T), (∅) and (E∪C∪T)⁶. The following relative contributions have been obtained:

$$\begin{aligned} Sh_E(K,X,I) &= 24.9\% \\ Sh_C(K,X,I) &= 60.9\% \\ Sh_T(K,X,I) &= 14.2\% \end{aligned}$$

to be compared with the following relative contributions in the equalized version:

$$\begin{aligned} Sh_E^e(K,X,I) &= 83.6\% \\ Sh_C^e(K,X,I) &= 11.8\% \\ Sh_T^e(K,X,I) &= 4.60\% \end{aligned}$$

where the relevant subsets are : $[E+\mu(C\cup T)]$, $[C+\mu(E\cup T)]$, $[T+\mu(E\cup C)]$, $[E\cup T+\mu(C)]$, $[C\cup T+\mu(E)]$, $[E\cup C+\mu(T)]$, (∅) and (E∪C∪T).

These results allow to establish a relationship between factor contributions and factors income share, and to use this ratio as a more easily interpretable measure of relative contribution of sources to overall inequality. Thus, in the example, earnings Shapley contribution is substantially higher than its relative weight in the obtention of income, capital income contributes to inequality in a slightly higher percentage than its weight in total income, and transfers contribution is about one fourth of its income share for the equalized version. The results are more amazing in the zero approach, since it tells us that earnings contributes for a little bit than one third to its income share, the great responsibility of inequality bearing on capital incomes. The story is even more dramatic in the US example since the contribution of earnings to overall inequality is negative (see Table 1 in Appendix)! Considering the huge difference between both calculations, it seems that in empirical studies we have to select only one of both.

To decide between the two calculations, the way how they perform with respect to the major shortcoming of the Shapley value would be of particular importance. Indeed the Shapley decomposition presents an important failure: the contribution assigned to a any income source is not independent from the level of disaggregation, i.e. it is sensitive to the way in which other

1994 March Current Population Survey are shown in the Appendix. Computations for the coefficient of variation are available from the authors upon request.

⁵ We are using the distribution of gross income, non adjusted by family size, as it is usual in the empirical studies that decompose inequality by income sources (see Jenkins 1995 for instance). The unit of analysis is the household. See the appendix for a detailed description of the data. The inequality of a source takes into account the fact that for many sources the household income is zero.

⁶ $I(\emptyset)=0$

sources are grouped together⁷. In the following we will explore the intensity of that dependence regarding the calculation selected.

The disaggregation of overall household income presented above is only one of the multiple partitions in which income sources could have been grouped. Let us go back to that example and focus on the contribution of income from transfers. Let us assume that earnings and capital income are combined in a unique source called "market income". In this case, the Shapley contribution of transfers to total inequality jumps from 14,2% to 47,2% in the zero approach while it increases from 4.6% to 5.5% in the equalized version. The following Tables 1 and 2 show with a greater detail, the kind of problems raised by the lack of independence.

**Table 1a: Shapley equalized decomposition contributions. Gini index
United Kingdom 1995 (Household non-adjusted distribution)**

Market Income 94.5%			T 5.5%	
E 83.6%		C 11.8%	T 4.6%	
E 83.1%		C 11.4%	Replac.Tr.3.2%	Other Trans. 2.3%
Wages 65.9%	Self-emp. 18.5%	C 11.5%	T 4.1%	
Wages 66.5%	Self-emp. 17.8%	C 10.4%	Replac.Tr.3.2 %	Other Trans. 2.1%

**Table 2a: Shapley zero decomposition contributions. Gini index
United Kingdom 1995 (Household non-adjusted distribution)**

Market Income 52.8%			T 47.2%	
E 24.9%		C 60.9%	T 14.2%	
E 8.3%		C 42.3%	Replac.Tr.21.8%	Other Tran. 27.5%
Wages 4.8%	Self-emp. 56.9%	C 40.8%	T -2.5%	
Wages -5.1%	Self-emp. 46.6%	C 29.4%	Replac.Tr.11.4%	Other Tran. 17.6%

The results obtained for thinner partitions of household income reflect the extremely high volatility of the zero inequality decomposition contributions, much more dependent from the level of disaggregation than the equalized inequality contributions, and that lead us to favor the equalized approach in the following. An extreme example of this variability is the contribution of transfers. In the case of disaggregating income in only two sources, its contribution is about one half of overall inequality. However, a further disaggregation of factors income makes negative the contribution of transfers to inequality.

Let us comment a little bit more the results for the equalized case. Let us focus, for instance, in the contribution of capital income. With the Shapley inequality decomposition, the contribution of this source is sensitive to the way other factors are treated. Thus, the contributions assigned to capital income are not the same if earnings are treated as a single entity or viewed in terms of two components: wages and self-employment income. Capital contribution to overall inequality is also affected by the way that transfers are grouped together. Indeed, the impact of capital income varies from 11.8% when considering three income sources, to 10.4% when disaggregating overall income into five components. In the same way, transfers contribution to overall inequality varies from 4.1% to 5.5% depending on the disaggregation of market income.

⁷ For a detailed discussion of the properties of the Shapley decomposition see Chantreuil and Trannoy (1999).

As a consequence of its dependence from the level of disaggregation, the Shapley decomposition does not guarantee that the contributions assigned to the components in which a given source can be divided, sum up the contribution to inequality of that income source treated as a single unit. One could consider interesting to study the impact on inequality of the public system of pensions and unemployment benefits (what we have called replacement transfers⁸), independently from the rest of public transfers programs. By doing so with the Shapley inequality decomposition, there are two points to be remarked. First, as said above, the contribution of that source will depend on how market income is treated. Secondly, as can be seen in the two last lines of Table 2, the sum of the contributions to inequality of replacement transfers and the rest of transfers is not equal to the contribution of total transfers in the decomposition that treats in the same way market income.

Although the consequences of the dependence from the aggregation level are relatively moderate (no higher than 1.4 percentage points) there is a general trend followed by the contributions of sources regarding the extension of the disaggregation. In general, the results are more dependent from an increase in disaggregation starting from a coarse partition than from a thin one.

On average, the contribution of a given component decreases with successive disaggregation of complementary sources. However, this can not be considered as a property of the Shapley rule, since this pattern is not necessarily true: the relative contribution of wages to overall inequality is slightly higher when transfers are disaggregated into two components than when they are treated as a single source.

An important remark about the interpretation of the Shapley contributions opens the way to a refinement of this decomposition method. As it was mentioned before, the contribution of any given factor to overall inequality can be interpreted as the expected marginal impact of the factor when the expectation is taken over all the possible elimination sequences. Thus, it is important that the elimination sequences or equivalently the subsets of components considered in the calculation have an economic appeal. That is not necessary the case is simply illustrated by taking again the example of the set of sources {earnings, capital, transfers}. For sure, the interpretation of the subset earnings-capital income as market income is straightforward. Nevertheless, the interest to consider the subsets earnings-transfers, and capital-transfers is not that clear, which raises some doubts about the interpretation of Shapley contributions. As long as the disaggregation becomes greater, the risk of having "unnatural" subsets of sources becomes higher, and providing an interpretation of the results may become an uncomfortable task. Introducing an explicit structure on the set of sources can contribute to mitigate this difficulty and it turns out that it can improve the degree of independence of a contribution of a source from the level of disaggregation as well.

3 Nested Shapley *versus* Owen

The introduced structure takes the general form of a tree and the simplest example would be a partition of the set of sources. A partition of the set of income sources K is the set $P_K = \{S_1, \dots, S_l, \dots, S_m\}$ such that for all $S_h, S_l \in P_K$, $S_h \cap S_l = \emptyset$ and

$$\bigcup_{h=1}^m S_h = K$$

(5)

⁸ Replacement income is the sum of unemployment, retirement and other public subsidies related to past economic activity

with $1 < m < k$. For the sake of illustration consider the following example: let us suppose that the data available allow us to disaggregate overall household income (X) into four elementary components, and that this partition is given exogenously and is relevant from an economic view⁹: (i) Earnings (E): i.e. wages and self-employment income. (ii) Capital income (C). (iii) Replacement transfers (R): including unemployment, retirement and other public subsidies related to past economic activity and (iv) Other transfers (O): including means tested and other public transfers not linked to economic activity, as well as private transfers received from other households or institutions. We could consider that these sources group naturally into two main aggregated factors: *Market income*(M), formed by earnings and capital income, $M = E \cup C$, and *Transfers* (T) $T = R \cup O$, including all kinds of transfers.

Consider the contribution of an elementary factor, for instance, capital income. With the Shapley decomposition, we have showed in the previous section that the contribution of this source depends on the number of subgroups considered not only in the disaggregation of market income, but also in the disaggregation of transfers. Therefore, the Shapley value that ignores the additional information about the structure of household income is not an appropriate tool for the decomposition analysis. A desirable requirement would be that the contribution of capital income would be at least independent from the disaggregation of transfers. In this section, we perform an empirical illustration of two different methods, derived from the Shapley value satisfying that milder requirement of independence.

3.1 The Nested-Shapley inequality decomposition

The ingredients of a partitioned decomposition problem is the vector (K, P_K, X, I) . The Nested-Shapley procedure (see Chantreuil and Trannoy (1999)) uses the nested structure of a partition and the calculation procedure can be simply described as a two stage procedure.

In a first stage called the "between stage", the contribution of any subgroup of sources is calculated along the Shapley formula given by (4). In other words, each subset of sources $S_l \in P_K$ is considered as an elementary source in this first stage of the calculation.

$$NSh_{S_l}^e(P_K, P_K, X, I) = Sh_{S_l}^e(P_K, X, I), \forall S_l \in P_K \quad (6)$$

In a second stage called the "within stage" the contribution of any elementary source is calculated along the Shapley procedure taking into account that the contributions of \ the elementary sources belonging to some subset must add up to the Nested-Shapley contribution of that subset.

⁹ We can object that the sources considered are not exogenous from each others (for example transfers income certainly depends of earnings). But any thinner partition will face a problem of endogeneity. In quoting Gottshalk and Smeeding (1997 p 668) "a major drawback of this source decomposition exercise is that they can easily misinterpreted because they do not make a distinction between endogenous and exogenous factors".

$$NSh^e_j(K, P_k, X, I) = \sum_{S \subseteq S_l, j \in S} \left\{ \begin{array}{l} \frac{(s-1)!(s_l-s)!}{s_l!} [I(y^e(S) - I(y^e(S - \{j\}))) + \\ \frac{(s_l-1)!}{s_l!} [NSh_{S_l}^e(P_K, P_K, X, I) - I(y^e(S_l))] \end{array} \right\} \quad (7)$$

where s_l is the dimensionality of S_l . On the within stage this decomposition method only considers subsets of elementary sources belonging to the same aggregated factor. As long as the partition used is relevant from an economic view, the subsets in question will also have an economic interpretation. Here to maintain the notations at the minimum two stages trees have been considered. Obviously this chain calculation will count as many stages as included in the tree.

The Nested-Shapley extension of the Shapley decomposition allow us to think of the equalized income inequality decomposition in two different ways. The first one (equalized), which is the one followed in the above formula calculates the inequality of a subset of sources assuming that inequality for all income components not included in the subset is removed.

However, it is possible to think of a calculation that removes inequality only for income components in the same level of the income structure (what can be called the semi-equalized method). In the within stage the distribution of income among subsets of sources is obtained by equalizing complementary sources belonging to S_l , i.e. we define $y^{se} : 2^{S_l} \rightarrow \mathbb{R}^n$, such that $y^{se}(\emptyset)=0$, and for all $S \in 2^{S_l}, S \neq \emptyset$,

$$y^{se}(S) = \left(\sum_{j \in S} x^j_1 + \sum_{\substack{j \notin S \\ j \in S_l}} \mu(x^j), \dots, \sum_{j \in S} x^j_n + \sum_{\substack{j \notin S \\ j \in S_l}} \mu(x^j) \right) \quad (8)$$

The contributions of the elementary sources for this semi-equalized version of the Nested Shapley decomposition will be given by:

$$NSh^{se}_j(K, P_k, X, I) = \sum_{S \subseteq S_l, j \in S} \left\{ \begin{array}{l} \frac{(s-1)!(s_l-s)!}{s_l!} [I(y^{se}(S) - I(y^{se}(S - \{j\}))) + \\ \frac{(s_l-1)!}{s_l!} [NSh_{S_l}^e(P_K, P_K, X, I) - I(y^{se}(S_l))] \end{array} \right\} \quad (9)$$

With this procedure the contribution of an elementary source does not depend on any data relatively to sources which do not belong to the same aggregated component. However as shown by the below tables the semi-equalized procedure produces a negative contribution of capital income in both American and British example which seems an odd result. For that reason it seems preferable to work with the equalized procedure.

**Table 3a: Equalized and Semi-Equalized Nested-Shapley Decomposition
Gini index. United Kingdom 1995 (Non adjusted distribution)**

Gross Income 100%			
Market Income		All Transfers	
NS=NS ^{se} =94.5%		NS=NS ^{se} =5.5%	
Earnings	Capital	Replac. Transfers	Other Transfers
NS ^e = 89.0%	NS ^e =5.5%	NS ^e =3.8%	NS ^e =1.7%
NS ^{se} =97.4%	NS ^{se} =-2.9%	NS ^{se} =9.1 %	NS ^{se} =-3.6%

**Table 3b: Equalized and Semi-Equalized Nested-Shapley Decomposition
Gini index. United States 1994(Non adjusted distribution)**

Gross Income 100%			
Market Income		All Transfers	
NS=NS ^{se} =90.2%		NS=NS ^{se} =9.9%	
Earnings	Capital	Replac. Transfers	Other Transfers
NS ^e = 89.0%	NS ^e =1.1%	NS ^e =8.2%	NS ^e =1.6%
NS ^{se} =97.2%	NS ^{se} =-7.3%	NS ^{se} =23.4 %	NS ^{se} =-13.3%

3.2 The Owen decomposition

Chantreuil and Trannoy (1997) and Shorrocks (1999) have considered an application of the Owen Value (1977). A direct application of the Owen¹⁰ formula gives:

$$Ow^e_j(K, P_k, X, I) = \sum_{S \in P_k, S_j \notin SG \in S_j, j \notin G} \sum_{g! (s_l - g - 1)! s! (m - s - 1)!} \frac{g! (s_l - g - 1)! s! (m - s - 1)!}{s_l! m!} * [I(y^e(S \cup G \cup \{j\})) - I(y^e(S \cup G))] \quad (10)$$

where g denotes the dimensionality of G.

Since the Owen formula is rather obscure as first glance, the discussion of the differences between the Owen contribution and the Nested-Shapley one will be pursued through the above example. The Owen contribution of earnings (E)} according to this decomposition rule is given by:

$$Ow^e_E(K, P_k, X, I) = \frac{1}{4} * \{y^e(E) + y^e(M) - y^e(C) + y^e(E \cup T) - y^e(T) + y^e(X) - y^e(C \cup T)\}$$

¹⁰ For a generalization of the Owen decomposition that consider situations in which the set of sources is decomposed into a level structure see Chantreuil (1998).

(11)

while for the Nested Shapley the compact formula gives:

$$NSH^e_E(K, P_k, X, I) = \frac{1}{2} * [y^e(E) + y^e(M) + y^e(C)] + \frac{1}{2} * \left\{ \left[\frac{1}{2} * [y^e(M) + y^e(X) - y^e(T)] \right] - y^e(M) \right\}$$

(12)

To understand the general difference between the Nested-Shapley decomposition and the Owen rule, we need to focus on the subgroups of components considered in the calculation. The Nested-Shapley decomposition takes into account only subgroups of sources of the same level, i.e. "subgroups" of elementary or coalitions of aggregated factors. If the partition used to analyze the contribution to inequality of different income sources is assumed to be exogenous and relevant, the interpretation of those subgroups is straightforward. The Owen rule, however, considered also subgroups that mix both elementary and aggregated components of income. It is of particular interest to notice that the Owen decomposition for earnings needs the calculation of the inequality of the subset formed by earnings and all kinds of transfers (EUT), and that formed by capital income and transfers (CUT). These two "coalitions" of sources has a doubtful economic meaning and for an application to inequality decomposition it can be depicted as a weakness. This feature reduces the interest of the latter decomposition rule, and favors the Nested-Shapley approach, that will be use in the following section.

The results in Table 4 show the differences between the values of the Owen and Nested-Shapley contributions for the British and American income distributions. These differences are not insignificant (about 5 percentage points for earnings and capital income) and show the importance of the election of the decomposition method.

**Table 4a: Nested Shapley and Owen equalized decomposition.
Gini index. United Kingdom 1995 (Non adjusted distribution)**

Gross Income 100%			
Market Income		All Transfers	
NS ^e =OW=94.5%		NS ^e =OW=5.5%	
Earnings	Capital	Replac. Transfers	Other Transfers
NS ^e = 89.0%	NS ^e =5.5%	NS ^e =3.8%	NS ^e =1.7%
OW=83.1%	OW=11.4%	OW=3.2 %	OW=2.3%

**Table 4b: Nested Shapley and Owen equalized decomposition.
Gini index. United States 1994(Non adjusted distribution)**

Gross Income 100%			
Market Income		All Transfers	
NS ^e =OW=90.2%		NS ^e =OW=9.9%	
Earnings	Capital	Replac. Transfers	Other Transfers
NS ^e = 89.0%	NS ^e =1.1%	NS ^e =8.2%	NS ^e =1.6%
OW=80.7%	OW=9.5%	OW=6.9 %	OW=3.0%

4 The Choice of a Tree

In order to realize an application of the Nested-Shapley rule on real data, it is necessary to define an income structure which allows to realize a meaningful decomposition of inequality by income components. We first focus on the disaggregation of gross income before tackling the problem of taxes .

4.1 Gross Income decomposition

In most countries microeconomic datasets allow to disaggregate household income in at least eleven elementary sources whose impact on inequality is considered of interest: household head wages and salaries, rest of household wages, self-employment income, private pensions, other capital income, social retirement benefits, unemployment and other replacement transfers, means-tested transfers, other public transfers; private transfers without forgetting direct taxes¹¹. However, there is not a unique partition given exogenously, and depending on the criteria considered to disaggregate income sources, and even the concept of income we want to study (gross versus }net income), somewhat different hierarchical or tree structures can be obtained. This paves the way to differences in results concerning the contribution of income sources to overall inequality. The ideal decomposition would be one taking into account all relevant economic links and not including any non pertinent subset of income sources. The following graph indicates which links between income sources seem relevant. The impossibility of reducing the structure of economic links to a single tree structure emerges quite naturally from the picture.

Table 5: Relevant links between elementary sources

Table 5 around here

Consider for instance the case of replacement incomes. As it makes part of public sector transfer programs, and comprises also a redistributive component at least in some countries, a first trivial solution would be to put it beside purely redistributive transfers (means-tested and other public transfers), in order to assess the contribution of public transfers to inequality. Even if the redistributive feature of replacement incomes cannot be denied, some authors like Bourguignon (1999) prefer to focus on their social insurance roots. Since in most countries social retirement and unemployment benefits are linked to the earnings obtained by workers when active through social

¹¹ A detailed description of each income source is given in the Appendix.

security contributions, they can be viewed as mainly delayed salaries. Moreover the differences in the extent of social security programs among developed countries, along with the substitution between public and private assurance has driven the literature to limit the redistributive analysis to non-contributive social benefits and taxes¹². This line of reasoning leads to a second solution where it is convenient to include replacement income within the Linked to factors income subset. But this option presents some limitations: it would not be possible to assess the joint impact of a subset traditionally analyzed in this kind of studies, namely "market income". The same kind of difficulties arises when analyzing the effect of income from private pensions}. On one hand, it can be seen as replacement or assurance income; on the other hand since it comes from savings, it seems quite natural to include it with the rest of capital income. It is rather difficult to decide between the two alternatives.

These are the kind of decisions we have to face when trying to define an income tree for an empirical decomposition of inequality. The applied analysis requires the careful definition of the subsets consider more relevant to our objectives and the election of the tree structure consistent with them. As there is no an ideal income structure, it is important to assess to which extent the election of the tree affect the results.

The Nested-Shapley inequality contributions to households gross income inequality for ten elementary sources according to three appealing income structures have been calculated. The first tree follows the traditional approach which divides household income in two main sources: income from production factors or market income and all kinds of transfers. Then, market income is subdivided in earnings and capital income, and we have distinguished between replacement transfers}, formed by social retirement benefits and unemployment and other replacement transfers; and redistributive transfers, } which are also subdivided into public and private (see Tables 6a and 6b in appendix). Both other trees proposed to limit the redistributive analysis to non-contributive benefits, and considered replacement transfers as incomes linked to labor. They differ regarding the treatment given to private pensions. In the second tree income from pensions funds is considered beside replacement income (see Tables 7a and 7b in appendix), while the third tree emphasized the linked between that source and the rest of capital income(see Tables 8a and 8b in appendix).

Nested-Shapley contributions according this three trees exhibit a moderate volatility. As it was expected, the sources more affected are those which treatment differs in the alternative trees and those directly related to them. Thus, the Nested-Shapley contributions of the different types of earnings are very little affected by the disaggregation of the rest of factors. However, the impact is higher for replacement transfers, redistributive transfers and capital income. For instance, replacement transfers contribution to inequality varies from -1.8% when those transfers are considered as income linked to production factors, to 3.8% when including with the rest of transfers. The contribution of "other capital income" inequality contribution oscillates between 2.3 and 0.6%, and private pension contributions go from 7.6% to 4.9%. Even if the magnitude of the changes is not huge in absolute terms, the picture is less favorable in relative terms since components in question presents a rather small income share. We conclude that the importance of the choice of a tree can be hardly overestimated.

We notice also that Nested-Shapley contributions are not restricted to be positive, and that certain factors contributions, namely "unemployment and other replacement transfers"; and private transfers, present a negative sign, suggesting an absolute equalizing effect.

¹² More on that in Bourguignon (1999)

4.2 Introducing Taxes: a difficulty

In all the examples considered until now, we have defined income in gross terms, i.e. without taking into account taxes, whose role in income inequality is of considerable interest for assessing fiscal policy redistributive effect. What are the consequences of including taxes in the analysis? Do the contributions of different sources are affected by the fact of considering net income instead of gross income?. The results for the decomposition of net income are shown in Tables 9a and 9b.

**Table 9a: Equalized Nested Shapley Decomposition.
Gini index. United Kingdom 1995 (Non adjusted distribution)**

Net Income 100%			
Gross Income			Taxes
NS=98.8%			NS=1.1%
Market Income		All Transfers	
NS=109.1%		NS=-10.3%	
Earnings	Capital	Repl. Transfers	Other Transfers
NS=111.8%	NS=-2.7%	NS=-3.7%	NS=-6.6%

**Table 9b: Equalized Nested Shapley Decomposition.
Gini index. United States 1994 (Non adjusted distribution)**

Net Income 100%			
Gross Income			Taxes
NS=100.8%			NS=-0.8%
Market Income		All Transfers	
NS=106%		NS=-5.1%	
Earnings	Capital	Repl. Transfers	Other Transfers
NS=115.0%	NS=-9.0%	NS=2.1%	NS=-7.2%

First of all the fact of including taxes influences the magnitude and the sign of relative contributions. For example the comparison with the results in Table 2 teach us that the relative contribution of transfers becomes negative, while the relative contribution of markets income exceeds one. The inspection of formula (7) allows us understand why. Before introduction of taxes both market income and transfer contributions decompose the inequality of gross income. After taxes, the object to decompose has changed. Now market income and transfer contributions decompose the nested Shapley contribution of gross income. The change in results is a by product of the lack of independence of the Shapley approach from the aggregation level. More specifically it tells us that the Nested Shapley contribution is not independent from adding an additional starting stage in the tree of components.

Second the fact of including taxes affects the results concerning the contribution of capital income in an unpleasant way. The contribution of capital income becomes negative, a rather uncomfortable result to shoulder. To understand why, the following remark is helpful.

Remark 1: *Adding an additional starting stage will affect the absolute Nested Shapley contributions of the previous components of the "between stage " by the same constant.*

An immediate calculation establishes that the constant in question is the composed of the difference of two terms. The first one is the inequality of net income minus the inequality of net income if the tax system was replaced by a poll tax. If taxes are a little bit progressive this term is negative. The second term is the inequality of net income in the hypothetical situation of an equalized gross income. Due to the inequality of taxes this second term is undoubtedly positive which gives a negative constant. If this constant is negative, then it can occur that adding an additional starting stage will change the sign of the Nested Shapley contribution of some previous aggregated components. In particular it will be the case if the former contributions were small. The former contribution (see Table 2) of transfers was weakly positive and the contribution of capital as well. Hence the puzzle of a negative contribution for capital income is deciphered.

Nevertheless one can have the feeling that the property encapsulated in the remark is rather inappropriate at least for an application to inequality issues. It will maybe better to have a decomposition method for which the ratio of relative contributions at some stage are independent of what happens on previous stages of the tree.

5 Concluding comments

As the exhibited results have fully shown, applying mechanically the Shapley decomposition can deliver odd results (For a rough application to french Data see Auvray-Trannoy (1992)). The major shortcoming comes from the lack of independence from the level of disaggregation. Some empirical solutions can circumvent the problem not to cure the illness. Among the solutions are avoiding the use of the zero calculation, taking care in the definition of the tree, applying the Nested Shapley. For the decomposition of net income it remains a difficulty which can be solved whether a more structural model of income decomposition of sources is introduced before applying Shapley ideas (for more details see Sastre-Trannoy 2000). Once these principles have been respected, one can hope to have sensible results which can be compared to those obtained with "natural" decompositions. In a companion paper (the above quoted paper) we apply the lessons gathered here and compare the Shapley decomposition results to those obtained with more standard methods.

References

- Auvray C. and A.Trannoy (1992): "Décomposition par source de l'inégalité des revenus à l'aide de la valeur de Shapley", Journées de Microéconomies Appliquées, Sfax.
- Bourguignon F. (1999): "Fiscalité et Redistribution". *Rapport au Conseil d'Analyse Economique* vol 11.
- Chantreuil F. (1998): "Axiomatics of Level Structure Values" in *"Power Indices and Coalition Formation"* (eds. Holler M.J.J. and G. Owen). Forthcoming.
- Chantreuil F., and A. Trannoy (1997): *Inequality Decomposition Values*, Mimeo. Université de Cergy-Pontoise, presented in Bordeaux TMR Network Distribution and Redistribution of Income Conference. Bordeaux December 1997.
- Chantreuil F., and A. Trannoy (1999): *"Inequality Decomposition Values: The trade-off between marginality and consistency"*. DP 9924 THEMA.
- Gottsschalk P. and T. Smeeding (1997): "Cross-National Comparisons of Earnings and Income Inequality". *Journal of Economic Literature*, 35, pp 633-687.
- Jenkins S.P. (1995): "Accounting for inequality trends: decomposition analyzes for the UK, 1971-86". *Economica* 62 pp. 29-64.
- Lerman R. (1999): "How income sources affect income inequality?" in (J.Silber) *Handbook on Income Inequality Measurement*. Kluwer Academic Publishers (Boston, Dordrecht, London).
- Owen,G. (1977):"Values of games with priori unions", in R.Heim and O.Moeschlin (eds). *Essays in Mathematical Economics and Game Theory*. New York: Springer Verlag.
- Sastre M and A. Trannoy (2000): *"A Marginalist approach to Inequality Decomposition By Factor Components: An Application to OECD Countries using the LIS database"*.Mimeo THEMA.
- Shapley L.S. (1953): "A value for n-person games", in *Contributions to the Theory of Games*, Vol. 2, *Annals of Mathematics Studies* 28. Ed by H.W.Kuhn and A.W.Tucker. Princeton University Press, pp307-317.
- Shorrocks A.F. (1999):*"Decomposition procedures for distributional analysis: A unified framework based on the Shapley value"*, mimeo, University of Essex.

APPENDIX

**Table 1b: Shapley equalized decomposition contributions. Gini index
United States 1994(Household non-adjusted distribution)**

Market Income 90.2%			T 9.8%	
E 81.2%		C 10.0%	T 8.8%	
E 80.8%		C 9.8%	Replac.Tr.6.5%	Other Trans. 2.9%
Wages 74.7%	Self-emp. 7.1%	C 9.8%	T 8.4%	
Wages 74.8%	Self-emp. 6.7%	C 9.3%	Replac.Tr.6.4 %	Other Trans. 2.8%

**Table 2b: Shapley zero decomposition contributions. Gini index
United States 1994(Household non-adjusted distribution)}**

Market Income 30.9%			T 69.1%	
E -0.08%		C 67.9%	T 32.2%	
E -15.4%		C 49.7%	Replac.Tr.34.3%	Other Tran. 31.4%
Wages -16.9%	Self-emp. 56.6%	C 48.0%	T 12.3%	
Wages -25.0%	Self-emp. 45.6%	C 37.5%	Replac.Tr.21.9%	Other Tran. 20.1%

Table 6a around here

**Table 6a :Nested Shapley equalized decomposition.Gini index.
United Kingdom 1995 (Household non-adjusted distribution)**

Table 6b around here

**Table 6b:Nested Shapley equalized decomposition.Gini index.
United States 1994 (Household non-adjusted distribution)**

Table 7a around here

**Table 7a :Nested Shapley equalized decomposition.Gini index.
United Kingdom 1995 (Household non-adjusted distribution)**

Table 7b around here

**Table 7b :Nested Shapley equalized decomposition.Gini index.
United States 1994 (Household non-adjusted distribution)**

Table 8a around here

**Table 8a :Nested Shapley equalized decomposition.Gini index.
United Kingdom 1995 (Household non-adjusted distribution)**

Table 8b around here

**Table 8a :Nested Shapley equalized decomposition.Gini index.
United States 1994 (Household non-adjusted distribution)**