The bidimensional decomposition of inequality: A nested Theil approach

MICHELE GIAMMATTEO*
Department of Public Economics
University of Rome "La Sapienza"
m.giammatteo@uniroma1.it

July, 2007

Abstract

In this paper we propose a nested inequality decomposition by income sources and population subgroups derived by the Theil index. We firstly motivate our preference for its associated decomposition by income sources with respect to the axiom-based proposal of Shorrocks (1982) and the Gini-based decomposition of Lerman and Yitzhaki (1985). Then we enhance the set of desirable proprieties able to sustain that choice with the additional requirement of subgroup decomposability. The nested decomposition of the Theil index allows the overall level of inequality to be function of only three types of factors: source-group income and population shares; source-group inequality. Finally, using LIS micro data on incomes, we apply it to the case of geographical disaggregation of inequality in Italy between 1989 and 2000.

Keywords: Inequality; Income sources; Decomposition; Italy;

JEL Classification: D31, O15.

* I would like to express my deep gratitude to my supervisor, Prof. Maurizio Franzini, for his fundamental support throughout my Ph.D. studies. I also thank Joan Esteban for his detailed and constructive comments received during the motivating discussions we had on inequality decomposition issues.
This work is supported by the EU project INEQ (www.criss-ineq.org), FP6-029093.

1. Introduction

In inequality decomposition studies we can distinguish two fundamental approaches. The traditional and larger applied technique is to identify the influence coming from specific population subgroups[1]. A complementary - rather than alternative - approach is to establish how different types of income affect total inequality: an example could be to detect the relative contribution of incomes from financial investments with respect to wages, capital profits, rents or state factors (transfers and taxes).

Our view is that an exhaustive analysis should include a mixture of explanatory factors based on subgroups as well as on income sources. In this paper we propose a nested-Theil decomposition of inequality, which allows us to identify simultaneously income sources and sub-population determinants of the overall level of inequality. We combine into a unique approach the standard decomposition by population subgroups, that separates total inequality in within-group and between-group components, and the decomposition by income sources, which divides overall inequality into proportional factor contributions.

The first crucial theoretical result is given by Shorrocks (1982), in which the objective of dividing total inequality into the partial contribution of each income component is dealt by starting with the definition of few fundamental axioms and then deriving the decomposition rule that satisfies them. The second possibility is suggested by Lerman and Yitzhaki (1985), who derive a decomposition rule that, following mathematically from the disaggregation of the inequality index taken as reference (the Gini coefficient), shows properties which depend strictly on that initial choice.[2]

In what follows, I refer to these alternative methods as those generating *axiom-based* and *natural* decomposition rules, respectively. For the former, I also use the attribute of *desirable*. Moreover, the term *natural* is used not to say that decomposition is the only possible and unanimously acceptable approach, but only because it is obtained by arithmetical derivation from the index of reference. The fundamental difference between the two methods is that, while the axiomatic approach yields a decomposition rule which respects the assumptions directly imposed upon it, in the second case the properties that one can derive must just be accepted, as a consequence of their

---

[1] Decomposition by population sub-groups has been the leading approach followed from many researchers to quantify how education, age, sex and other individual characteristics, affect inequality. The approach consists in dividing a sample into discrete categories (rural and urban residents, individuals with primary or secondary school or higher education, etc) and then calculating the level of inequality within each sub-sample and between the means of the sub-samples. Among the others, one important limitation of this kind of analysis is the lack of control for the endogeneity of some explicative variables that may themselves be partly determined by income patterns. This and other problems have been overpassed adopting regression technique in decomposition analysis (Oaxaca 1973, Fields 1998, Bourgignon *et al.* 1998), which allows continuous variables being permissible, too.

[2] See also Pyatt et. al. (1980), Stark et. al. (1986), Leibbrandt et. al. (1996).

complete dependence on the global inequality index.[3] Despite this, the "natural" decomposition derivable by the Theil index of inequality is shown to be the most suitable and consistent method to implement in empirical application. This is established on the basis of the satisfaction of three very important properties, which allow us to identify such decomposition as the only "well-behaved" method among those considered.

It is worth emphasising that another important technique of decomposition, based on the concept of the Shapley values,[4] is not considered in this paper because of the negation of a very fundamental property: the independence of the level of disaggregation.[5] Despite the great variety of possible applications it could reproduce, our view is that it cannot be included in the following analysis because of our main objective: to seek the most consistent, satisfactory, not ambiguous (or contradictory) decomposition method.

The second, and decisive, step is to enhance the set of desirable proprieties able to sustain the choice of the Theil-based decomposition with the additional requirement of subgroup decomposability. This is done putting in evidence the different advantages coming from the use of the Gini and Theil indices when a two-way decomposition of overall inequality is implemented. Many authors have faced the problem of multidimensional decomposition (see Akita, 2003; Conceição et.al., 2000; Wodon, 1999), but only considering the hierarchical structure of their population attributes of reference (territorial, sectorial, and so on). The only attempt of providing a mixed approach of inequality decomposition by population subgroups *and* income sources is given by Mussard (2004). Considering the Dagum (1997) decomposition of the Gini coefficient and the additional disaggregation of total income into its main components, he proposes a bidimensional approach of study which is shown to possess less properties (and appealing structure) of an equivalent method derivable from the Theil index of inequality.

The paper is organized as follows. In section 2 we introduce in details the Shorrocks's axiom-based approach (1982) as well as two among all the possible natural decompositions of the Gini and Theil indices. Firstly, we emphasize the satisfaction of a fundamental static property (of *uniform addition*, i.e. negative contribution of equally distributed sources); secondly, we propose two new *dynamic*

---

[3] In this respect, Shorrocks (1982) writes: "Using the natural decomposition of the Gini […] could be justified by arguing both that the Gini coefficient should be used as the measure of inequality (which is an acceptable position to take), and that we must choose the decomposition rule that follow naturally from the conventional way in which the Gini formula is written. This latter position is simply untenable". Just a few years later Lerman and Yitzhaki (1985) answer: "The approach based on the Gini is worth pursuing for three reasons. First, its use is desirable, because permits one to form the necessary conditions for (second order) stochastic-dominance. Second, our decomposition yields an intuitive interpretation of the elements making up each source's contribution to inequality. Third, it gives the advantage of examining the marginal changes in the size of an income source on overall inequality."

[4] See Shorrocks (1999) and Sastre and Trannoy (2001) for a detailed analysis of the Shapley-based methodology.

[5] In order to understand the importance of this property, observe that if it were not satisfied the contribution, let us say, of earnings might change if capital incomes were partitioned into rent, interest, and dividends, or transfer payments were split into (private and public) pensions, unemployment subsidies, and so on.

principles which any suitable rule of decomposition should plausibly satisfy. In particular, the following three aspects are highlighted[6]: *i*) different rules of decomposition show divergent responses to uniform source variations, which ultimately depend on the level and spread of the source interested by the change and those corresponding to the overall distribution; *ii*) the positive or negative input of initial contribution and the share of the sources of total income constrain the range of those variations (both in terms of sign and level); *iii*) the Theil-based rule is found to be the only decomposition which fully satisfies both our desirable dynamic principles. In order to underline the importance of such kind of analysis, note that Lerman and Yitzhaki (1985) emphasize: "how percentage changes in particular taxes or transfer influence the distribution of income are important policy issues."[7] This is because many public policies concern directly types of income rather than typologies of individuals: different tax or transfer programmes, minimum wage schemes, or pension reforms can be better evaluated in terms of distribution effects if the mere disaggregation of income by income recipients is overcome.

Section 3 presents a short overview of the Theil (1967) main result about subgroup decomposability, with particular attention devoted to its appealing functional form. In fact, it can be seen as a function of three very simple elements: population and income shares and subgroup inequality.

In section 4 we derive the nested (or bidimensional) decomposition of the Theil index by income sources and population subgroups. Finally, using LIS micro data on incomes, section 5 presents the results of an application of that nested rule of decomposition to the case of geographical disaggregation of inequality in Italy between 1989 and 2000. We are thus able to separate each source contribution to total inequality into two additional terms: the fractions affecting between and within-group regional inequality.

2. Inequality decomposition by income sources

In this section we propose a comparative analysis of three largely applied rule of decomposition by factor components. We analyse the axiomatic derivation of Shorrocks (1982) and two "natural" decompositions derivable from two well known inequality measures: the Gini coefficient and the Theil index[8].

[6] They are analytically derived in Giammatteo (2007).

[7] Lerman and Yitzhaki (1985), Stark et. al. (1986), Leibbrandt et. al. (1996) proposed a similar study (only in the case of the Gini decomposition), providing results on the sign of the global index variation. Se also Paul S. (2004), for an extension over a larger set of inequality indices.

[8] See Lerman and Yitzhaki (1985) and Paul (2004), respectively. For the decomposition of the Gini index by income components see also Pyatt et. al. (1980), Stark et. al. (1986), Leibbrandt et. al. (1996).

Our main point is that if an axiomatic approach of study has to be generally preferred because of the possibility of ensuring the satisfaction of desirable (decomposition) properties, on the other hand the existing natural decompositions cannot be ignored, mainly because of the consistencies they may show with respect to the standard inequality theory.

We propose to evaluate the appropriateness of the three decomposition methods said above on the basis of the satisfaction of one important static property (*uniform addition*) and two additional dynamic principles. These are introduced in order to evaluate the decompositions behaviour which follows from a simple variation in the incomes of only one source.

Our opinion is that the dynamic perspective proposed in the following analysis should deserve more attention among researchers. In fact, if the main objective of every decomposition study is that of looking for the inequality determinants over time, the essential requirement should be that of using methodologies which consistently respond to source distribution changes. More precisely, one should expect that a very simple variation in the incomes of a particular source should cause its associated contribution to change without generating *perverse results*. Since income components are usually distributed with different degrees of inequality and different magnitude, every decomposition analysis should be able to determine in a clear and consistent way *how* those variations affect decomposition outcomes and, as consequence, total inequality. This objective is central for policy scheduling: odd decomposition behaviours could imply erroneous identification of the main inequality causes, with the consequent invalidation of analysis interested in estimating the distribution effects of specific policies. In this respect, looking for reasonable relations between source variations and associated contribution effects can be used as a reference framework for the identification of *well-behaved* decomposition rules.

2.1 Axiom-based and "natural" decompositions

Let us denote with $\mathcal{D}^n$ the class of all income distributions Y composed of *n* units (individuals or households). Suppose that total income is also divisible into *M* different income sources, such that

$$Y = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n}\sum_{m=1}^{M} y_i^m = \sum_{m=1}^{M} Y_m \qquad [1]$$

Given a general inequality index $I(Y)$ we can define *the absolute contribution* to total inequality of the *m*-th component of income as the generic function $C_m = C(Y_m;Y)$, with $Y_m \in \mathcal{D}^n$ and $y_i^m \neq 0$ at least for one i[9]. Therefore, we define the m-th *proportional contribution* as $c_m = c\left(Y_m; Y\right) = \frac{C_m}{I(Y)}$.

Hereafter, we identify it also as the generic *rule of decomposition*. Note also that the c_m is usually a function of the total index of reference in the case of natural decompositions; it is not so when the Shorrocks rule is adopted. To conclude this first set of definitions, in what follows we denote the functions C_m and c_m with different notations, more precisely: W_m and w_m in the case of the Shorrocks proposal; G_m and g_m for Gini's natural decomposition ; T_m and t_m for the Theil-based method.

The starting point of Shorrocks (1982) is to suggest six fundamental assumptions (axioms) which every method of decomposition should plausibly satisfy[10]. These allow him to demonstrate that,

$$w_m(I) = \frac{W_m(Y_m;Y)}{I(Y)} = \frac{Cov(Y_m,Y)}{\sigma^2(Y)} \qquad \textit{for all } Y \neq \mu e \qquad [2]$$

is the "only" rule of decomposition satisfying the six axioms and such that the relative importance of different income components with respect to total inequality is "independent of the choice of the measures".[11]

A first method followed by many authors to derive (one of the possible) natural decomposition of the Gini coefficient is based on a mathematical derivation which initially considers this index as defined by

$$G(Y) = \frac{\left(2\operatorname{cov}\left[Y, F(Y)\right]\right)}{\mu(Y)}$$

where $F(Y)$ denotes the cumulative distribution corresponding to the density function $f(y_i)$, defined as the rank of y_i in Y divided by the number of observations. Lerman and Yitzhaki (1985) show that the Gini coefficient can be decomposed as the sum of the absolute contribution coming from every income component as,

[9] Note that the contribution shown by each source Ym is usually different from its own inequality.
[10] See Shorrocks (1982) pp. 196-203.
[11] Note also that it corresponds to that of the "natural" decomposition of the variance.

$$G(Y)=\sum_{m=1}^{M} R_m Gini_m S_m \qquad [3]$$

Every absolute contribution $\left(R_m Gini_m S_m\right)$ is the product of three measures: R_m is the "Gini correlation" between the income component *m* and total income;[12] $Gini_m$ is the relative Gini of component *m*; and S_m is the component *m* share of total income. Using [3] we can also define the *proportional Gini contribution* coming from the income source *m*,

$$g_m=\frac{R_m Gini_m S_m}{G(Y)} \quad \forall m=1,...,M \qquad [4]$$

Several researchers have tried to study more systematically the above decomposition. Among others, Podder (1993) claimed that this interpretation of the Gini decomposition can be shown to be «wrong and totally misleading». To understand why, he takes as an example a constant component of income. In this case the concentration coefficient of the component ($Gini_m$) is zero. This necessarily means that its absolute contribution to the overall level of inequality is also zero. Despite this, Podder emphasizes that «[…] it is reasonable to think the addition of a constant to all incomes leading to a reduction in inequality if we accept relative measures». Note that this undesirable property of zero contribution for equally distributed sources is also satisfied by the Shamrocks' derivation [2]: more specifically, it is directly required by one of his six axioms.

Finally, let us introduce the natural decomposition of the Theil index. His well-known formula is

$$T(Y)=\frac{1}{n\mu}\sum_i\left(\ln\frac{y_i}{\mu}\right)y_i \qquad [5]$$

Simply taking into account the basic relation [1] and applying it to the generic y_i in [5], we can derive the following (natural) decomposition

---

[12] $R_m=\frac{\operatorname{cov}\left(Y_m,F[Y]\right)}{\operatorname{cov}\left(Y_m,F[Y_m]\right)}$, where $F[Y_m]=\left(f(y_1^m),...,f(y_n^m)\right)$, and $f(y_i^m)$ is equal to the rank of y_i^m in Y_m divided by the number of observations.

$$t_m = \frac{T_m(Y)}{T(Y)} = \frac{\sum_i \left(\ln \frac{y_i}{\mu}\right) y_i^m}{\sum_i \left(\ln \frac{y_i}{\mu}\right) y_i} \qquad [6]$$

where $T_m(Y) = \frac{1}{n\mu}\sum_i \left(\ln \frac{y_i}{\mu}\right) y_i^m$ is the generic absolute contribution.[13] Contrary to the Shorrocks proposal [2] and the Gini–based decomposition [4], t_m satisfies a very important property: it is strictly negative for any equally distributed source of income.[14] The importance of this property of *uniform addition* for an inequality index comes from the fact that it is directly implied whenever the *transfer axiom* and the *scale invariance axiom* are satisfied. That said, is it also paramount to extend this requirement to the decomposition ground? To phrase it differently, should the uniform addition property set out a reference axiom which would constrain the *consistent* derivation of a decomposition rule?

Let us underline how for the class of the *natural decompositions* the satisfaction of the property by an inequality index does not guarantee that the associated rule does the same.[15] Thus, even if the corroboration of one property by an overall index does not bind its natural decompositions, could some inconsistencies arise as well? We believe that one useful way to follow could be that of focusing the attention on the *dynamic behaviour* of the existing decomposition rules, ensuring (avoiding) the application of appropriate (improper) methods as a consequence of the satisfaction (negation) of reasonable requisites.

2.2 Uniform variations of incomes and well-behaved decomposition rules

In the previous section we have introduced three rules of decomposition which allow us to distinguish how different income components contribute to the total level of inequality in a given

---

[13] It is not the factor *m*'s Theil, but the pseudo-Theil. In fact, the weights for the *m*-th factor incomes $\left(\frac{1}{n\mu}\log\frac{y_i}{\mu}\right)$ are based on the rank of the total distribution.

[14] If $y_{im} = \mu_m \;\; \forall i$, the numerator of T_m simplifies to μ_m multiplied by the negative of the Theil-L index of overall inequality, $1/n\sum_{i=1}^{n}\ln(\mu/y_i)$.

[15] This is the case, for example, of the Gini-based rule [4] and the natural decomposition of GE(2) given by the Shorrocks Theorem: while these two indices satisfy the property of uniform addition, the corresponding decompositions neglect it: $g_\delta = \frac{R_\delta G_\delta S_\delta}{G(Y)} = 0$ and $s_\delta = \frac{Cov(\delta e, Y)}{Var(Y)} = 0$.

distribution. We have also introduced the important property of uniform addition, which we propose as a first criterion of identification of suitable decomposition.

If the acceptance (rejection) of that property cannot constitute the only and decisive principle for evaluating the existing methodologies, what is evident from much of the available empirical literature is the high frequency of divergent results that the applications of different methods produce.[16] The ordinary decomposition of the Gini coefficient, for example, may indicate that wage incomes have a very small positive influence on overall inequality, while that of the Theil index may show the opposite. Moreover, there are no theoretical foundations on the basis of which to explain such conflicting empirical findings.

The objective of this section is to look for a theoretical framework able to support the existing decomposition methods in terms of their *consistent dynamic behaviour*. We propose to follow an approach of study founded on the *marginal variations of income sources* in order to identify *well-behaved* decomposition rules. To this end, we compare the different responses that the rule proposed by Shorrocks, the natural decompositions of the Gini and Theil indices provide as a result of a uniform proportional change in all incomes of one component. Basically, we take as starting points the following two general *dynamic* principles:

PRINCIPLE A (for *inequality increasing sources*). Consider a source of income contributing positively to total inequality (C_m^+). When all the incomes in Y_m^+ increase uniformly, the proportional contribution of the same source has to get bigger, *independently* of the initial share of Y.

PRINCIPLE B (for *inequality decreasing sources*). Consider a source of income contributing negatively to total inequality (C_m^-). When all the incomes in Y_m^- increase uniformly, the proportional contribution of the same source has to get smaller (bigger) *if* its initial share of Y is small (big) enough.

They found on the fact that increasing uniformly a source also increases, but less than proportionally, the aggregate amount of income. Thus, holding constant the other sources in terms of levels and spread, the increased weight of the varied source on total inequality should have different effects, depending on its initial character (inequality increasing or decreasing). More specifically, in the first case it should always reinforce its positive contribution (PRINCIPLE A), while in the second case the effect should not be independent of its initial share (PRINCIPLE B). The idea behind the first principle is that if a component contribution is initially positive, and we increase proportionally its share (driving it closer to the whole distribution), then one would expect

[16] See, for example, Morduch and Sicular (2002) for an empirical application to rural China.

to observe a greater impact on total inequality, since the relative weight of its own inequality with respect to the other sources has risen. On the other hand, if a source initially contributes negatively to total inequality, one would expect that, above a specific level of Y_m^- (with μ_m^- close enough to μ), the character of the component should show a positive change. In other words, the expected behaviour of any source contributing negatively to total inequality should *not* be *monotonic*: its negative contribution should be reinforced when the share is "small", and weakened for shares bigger than a specified threshold.[17] To better understand the point, one can imagine the component Y_m^- increasing (continuously) until it can be considered close enough to *Y*. It seems clear that, accepting the continuity for the function characterising the decomposition rule, almost the overall inequality would be ascribed to the income source *m* subject only to a change of its character from negative to positive influence.

Principles A and B allow us to derive strong conclusions about the appropriateness of the three decompositions described in the previous section. Making explicit the response of [2], [4] and [6] to a uniform variation incomes of just one source[18], it is possible to bind the sign of their marginal behaviours. They result to be function of: *i*) the initial character of the component (inequality increasing or decreasing) subject to the scale transformation; *ii*) the income source (and overall) inequality; *iii*) the source share of total income. Despite this, each decomposition shows very peculiar, and conceptually not expected, behaviour: the *Gini decomposition* constrains the initial contributions to increase (decrease), whenever they are initially positive (negative). This effect disagrees with the expected not monotonic behaviour of inequality decreasing sources (above Principle B). Moreover, the initial share of the altered source does not play any role in defining the sign and magnitude of the proportional contribution. In the *Shorrocks axiom-based proposal* other conditional factors play a crucial role in defining the sign of the marginal variations, even if these are still independent of the source shares. The ratio of partial to total variance makes the prediction about the decomposition behaviour also more complex and, in some cases, unjustifiable. Finally, the Theil-based decomposition is the only one which *perfectly satisfies* both the Principles A and B. As it should be expected, all the inequality increasing sources enhance their positive effect on inequality as a consequence of an increase of their share on total income; conversely, if a source shows a negative effect on total inequality, increasing its relative size can imply different performances dependently of its initial relative weight on total income: the higher the *m*-th source share of *Y*, the more likely is a positive variation (i.e. a fall in its absolute equalising effect)

[17] This threshold cannot be established in general: it will depend on the specific amount, spread and rank of the incomes in *Y* and Y_m.

[18] The corresponding analytical results are explicitly derived in Giammatteo (2007).

The corroboration of the two dynamic principles A and B, in addition to the satisfaction of the property of uniform addition, should suggest the Theil decomposition [6] as the best (among the three considered) *well-behaved*[19] rules of inequality decomposition by income sources It can be consistently applied with the objective of explaining the total patterns of inequality through the distribution of the income components. The Gini-based rule and the Shorrocks axiom-based proposal are two less desirable decompositions to implement. Their use, in fact, should imply the unpleasant occurrence of "obscure" reaction to changes in the source distributions, which in turn could cause the overall inequality to respond perversely to such changes.

3. Theil index and subgroup decomposability

> *«[…] We now find that this measure has a simple interpretation in terms of income shares and population shares; moreover, that it can be aggregated in a straightforward manner. In that respect it is more attractive that most well-known inequality measures such as Gini's concentration ratio.» (Theil, 1967, pp. 95-96)*

The main motivation of decomposing inequality by population subgroups is given by the possibility of examining the relationship between the demographic structure of a population and the associated income distribution. As well known, the Theil (1967) axiomatic measures properly achieve this objective; as a consequence, they are often used in empirical works in order to provide keys of understanding for the observed patterns of inequality. This section briefly summarises the main decomposability results derivable by the Theil index formulation, in order to underline the appealing characteristics also shown by the nested (group and source-based) decomposition of inequality proposed in the following section.

Consider each individual in the total population characterized by the general pair $(y_i, k) = y_{ik}$ of total income $y_i \in Y$ and one attribute $k = 1,...,K$. Suppose that this attribute divides the total population into *K* mutually exclusive and exhaustive groups. Then, we can define

$$n = \sum_{k=1}^{K} n_k \,; \quad \pi_k = \frac{n_k}{n}; \text{ and } \quad Y_k = n_k \mu_k = \sum_{i=1}^{n_k} y_i \,, \qquad k = 1,...,K$$

[19] In pure mathematics, "well-behaved" objects are those that can be proved or analyzed by elegant means to have elegant properties. In both pure and applied mathematics, well-behaved also means not violating any assumptions needed to successfully apply whatever analysis is being discussed.

where n_k and μ_k represent the number of individuals and the *k*-group mean, respectively. The minimum requirement for *population decomposability* is that if inequality increases in a population subgroup then, other things being equal, inequality increases overall (property of *subgroup consistency*). Given the generic index characterization $I(Y)=\Phi(I_1,I_2,...,I_K;\pi,\mu)$, it requires that the function Φ has to be strictly increasing in each of its first *K* arguments $I_k = I(Y_k)$[20].

The "aggregation problem" is solved by Theil (1967) providing a breaking down rule made of two components: the first identifies the distance *between* homogeneous groups of units, while the second incorporates the dispersion *within* each group. In formula, we have

$$\begin{aligned} T(Y) &= \sum_{k=1}^{K}\frac{n_k}{n}\frac{\mu_k}{\mu}\cdot\ln\left(\frac{\mu_k}{\mu}\right) + \sum_{k=1}^{K}\frac{n_k}{n}\frac{\mu_k}{\mu}\cdot\left(\frac{1}{n_k}\sum_{i=1}^{n_k}\frac{y_{ik}}{\mu_k}\cdot\ln\left(\frac{y_{ik}}{\mu_k}\right)\right) \\ &= \sum_{k=1}^{K}\pi_k s_k \ln s_k + \sum_{k=1}^{K}\pi_k s_k\cdot T_k = Tb+Tw \end{aligned} \qquad [7]$$

where $\pi_k s_k = \frac{n_k}{n}\frac{\mu_k}{\mu} = \frac{Y_k}{Y}$ is the total income share held by subpopulation k. The between-group (Tb) and the within-group (Tw) components measure the inequality contribution coming, respectively, from the differences in subgroup means (μ_k) and the income differences inside each population subgroup. Note that the first term contributes nothing only if $s_k = 1,\ \forall k$. In all other cases it will be strictly positive. The second term, which corresponds to the weighed mean of the *K* sub-indices $T_k = \frac{1}{n_k}\sum_{i=1}^{n_k}\frac{y_{ik}}{\mu_k}\ln\left(\frac{y_{ik}}{\mu_k}\right)$, is also never negative and reaches its minimum (zero) in case of equally distributed incomes inside each subpopulation *k*.

An equivalent expression of [7] possesses particular appeal because of its interpretation as function of income and population shares,

$$T(Y) = \sum_{k=1}^{K}\frac{Y_k}{Y}\ln\left(\frac{Y_k}{Y}\Big/\frac{n_k}{n}\right) + \sum_{k=1}^{K}\frac{Y_k}{Y}\cdot\left[\sum_{i=1}^{n_k}\frac{y_{ik}}{\sum_i y_{ik}}\ln\left(\frac{y_{ik}}{\sum_i y_{ik}}\Big/\frac{1}{n_k}\right)\right] \qquad [8]$$

where $Y = n\mu$. [8] says us that:

[20] See Bourguignon (1979), Cowell (1980), Shorrocks (1980, 1984) for an exhaustive treatment of the subgroups decomposition and the class of additive inequality measure.

i) when the income shares $\frac{Y_k}{Y}$ $\left(\frac{y_{ik}}{\sum_i y_{ik}}\right)$ equal the corresponding population shares $\frac{n_k}{n}$ $\left(\frac{1}{n_k}\right)$ for each $k=1,...,K$ $\left(i=1,...,n_k\right)$, then the between (within) component contributes nothing to overall inequality;

ii) the bigger the discrepancy between group (individual) relative income and group (individual) population shares, the greater the contribution of the between (within) component of inequality;

iii) the Theil index is a function of three simple elements: subgroup income and population shares, and subgroups inequality.

Some works have tried to extend this fundamental one-dimensional result to the case of two (or more) attributes. Akita (2003), for instance, considers the three-level hierarchical structure of a country into regions, provinces and districts in order to derive a *nested geographical decomposition* given by,

$$T_R = Tw_P + Tb_P + Tb_R$$

The overall regional income inequality is thus decomposed into the within-province (Tw_P), the between-province (Tb_P), and the between-region (Tb_R) components. The within-province contribution to inequality is a weighted average of Theil indices at the province level (the weights being the shares of province income in the region), while the overall-between-province component is a weighted average of between-province income inequalities within each region[21].

Conceição, et. al. (2000) also discuss the implications of the Theil index decomposition into a sequence of nested and hierarchic group structures but, differently from the Akita analysis which focuses on the geographical disaggregation, they apply the multilevel decomposition of the Theil index on wages and employment, operating disaggregation by industrial classification. Their analysis presents two interesting peculiarities: the first consists of introducing the multi-sequence decomposition of the Theil index in order to identify the "information gain" of moving towards a higher number of SIC[22] digits in explaining wages evolution; as second, they study the determinants of between-group inequality variation over time as a function of the two main index constituents: income and population shares.

[21] See Akita (2003) for a detailed description of the procedure and the results of his empirical application to China and Indonesia.

[22] Standard Industrial Classification.

Wodon (1999), instead, provides a multidimensional extension of the (one-level) group decomposition of the Gini index proposed by Yitzhaki and Lerman (YL, 1991). A first strategy consists of taking into account *g* mutually exclusive groups obtained by the combination of the two attribute dimensions *k* and *h* (i.e. $g = 1,....,K \cdot H$). Then, he applies the one-dimensional YL decomposition along the *g* categories. A more interesting approach to the bivariate problem is to proceed sequentially (i.e. to operate the disaggregation of the *K* groups into *H* sub-groups). The derived Gini decomposition is thus given by the sum of the following elements: *a*) the within groups component of total inequality[23]; *b*) two terms identifying the first and the second order stratification component[24]; *c*) two other terms identifying the first and second order between-group components of inequality[25].

In what follows we go through the nested decomposition of inequality by income sources *and* population subgroups. In particular, we compare the possibilities which the Gini coefficient provides (Mussard, 2004) with an alternative derivation based on the Theil index of inequality.

4. Nested decomposition rules

Theoretical and empirical literature on inequality decomposition has mainly developed independent analysis for sub-populations and income sources disaggregation.

Given the theoretical results of section 2 about the possible ways of identifying the inequality contribution coming from the various income sources, the appropriateness of the Theil index is reinforced by the possibility of extending its appealing properties to the case of a nested (bidimensional) rule of decomposition. This permits to account "*in a straightforward manner*" for subgroup and income source distribution structure, providing *group-source inequality contributions* which can be expressed as a function of income shares, population shares and group-source specific inequality.

Mussard (2004) proposes an interesting attempt of providing theoretical bases of a unified decomposition approach[26]. His starting point consists of considering the Gini index as the reference measure of inequality. If on one hand this choice has, obviously, robust foundation because the well-known Gini relation with the Lorenz curve and deprivation theory, on the other hand it is not so appropriate when the objective of the analysis is to decompose total inequality. The three main

[23] It is the result of two within group expansions, starting with the dimension *k* and following with dimension *h*.

[24] The first (second) order stratification term measures the (YL) stratification within the overall population (within the group *k*).

[25] The first order between groups term measures the inequality between groups according to dimension *k*, while the second order between group term measure the extent of the inequality, within group *k*, between the households with different characteristic *h*.

[26] See also Rao (1969) for an attempt of providing a unified solution to the decomposition issue.

reasons of this claim are represented by: *i*) the Gini interaction (third) term of its subgroup decomposition[27]; *ii*) on the ground of income source disaggregation, the not satisfaction of the property of uniform addition and of the dynamic principle B treated in section 2; *iii*) its functional final structure, which give clear but not useful (or functioning) indications about the elementary factors driving source and subgroup inequality contributions.

In the rest of this section we will try to validate, instead, the Theil index appropriateness. As noted above: *i*) it implies an easy and suitable inequality interpretation as function of simple income and population shares; *ii*) it also implies a natural decomposition by income sources which satisfies very attractive proprieties; *iii*) it is perfectly decomposable by population subgroups. As specified below, *iv*) the bidimensional (or *nested*) decomposition of the Theil index suggests a very appealing opportunity of linkage between functional and personal income distribution analysis; *v*) it also constitutes a useful tool of policy evaluation: crucial government chooses such as those concerning labour market reforms, transfers and taxes schedule, policy decentralisation, etc., could be properly evaluated in terms of inequality effect.

Consider the total distribution of income *Y* composed of *n* units (individuals or households) receiving income from *M* different sources of income Y_m, such that [1] is still true and $y_i^m \geq 0$ with $y_j^m > 0$ at least for one *j*. Mussard (2004) shows how using one of the possible Gini coefficient decomposition by income sources and the Dagum (1997) disaggregation by population subgroups it is possible to derive the following bidimensional decomposition,

$$G = \sum_{m=1}^{M} \left(\frac{\sum_{k=1}^{K} \left(\sum_{i=1}^{n_k} \sum_{j=1}^{n_k} \left(y_{im,k} + y_{jm,k} - 2y^{*}_{ijm,k} \right) \right)}{2\mu n^2} \right)$$
$$+ \sum_{m=1}^{M} \left(\frac{2\sum_{k=2}^{K} \sum_{h=1}^{k-1} \left(\sum_{i=1}^{nk} \sum_{j=1}^{nh} \left(y_{im,k} + y_{jm,h} - 2y^{*}_{ijm,kh} \right) \right)}{2\mu n^2} \right) \qquad [9]$$

where $y_{im,k}$ is the generic *m*-th income source in group *k* and $y^{*}_{ijm,kh}$ is the *m*-th source of the minimum between $y_{i,k}$ and $y_{j,h}$. From expression [9] is possible to divide the generic *m* source contribution to total inequality *G* into the two components of within and between-group inequality.

---

27 Mussard (2004) uses the Dagum (1997) result about the possibility of decomposing of the Gini ratio into the following three components: 1. the contribution of the within groups income inequalities; 2. the net contribution of the extended Gini inequality between subpopulations taking into account variations in mean, standard deviation, asymmetry; 3. the between groups inequalities of the "transvariazione". See also Lambert and Aronson (1993) for the relation between the Gini residual term and the Lorenz curve.

A similar derivation can be obtained for the Theil index. In section 2 we have already shown how to decompose it by income sources, that is

$$T(Y) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\mu}\ln\left(\frac{y_i}{\mu}\right) = \sum_{m=1}^{M}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i^m}{\mu}\ln\left(\frac{y_i}{\mu}\right)\right) = \sum_{m=1}^{M}T(m) \qquad [10]$$

where $T(m)$ is the generic *pseudo*-Theil for the source *m*. The next fundamental step is to implement the source-based decomposition [10] into the subgroup disaggregation of total inequality given by [7]. Keeping in mind the basic income distribution structure [1] and the sub-means additivity,

$$\mu_k = \sum_{m=1}^{M}\mu_k^m \qquad [11]$$

we are able to divide the between-group component of total inequality Tb into *M* source contributions as following,

$$Tb = \sum_{k=1}^{K}\frac{n_k}{n}\frac{\mu_k}{\mu}\cdot\ln\left(\frac{\mu_k}{\mu}\right) = \sum_{k=1}^{K}\frac{n_k}{n}\frac{\left(\sum_{m=1}^{M}\mu_k^m\right)}{\mu}\ln\left(\frac{\mu_k}{\mu}\right)$$

$$= \sum_{m=1}^{M}\left[\sum_{k=1}^{K}\pi_k s_k^m \ln\frac{\mu_k}{\mu}\right] = \sum_{m=1}^{M}Tb(m) \qquad [12]$$

where $Tb(m) = \sum_{k=1}^{K}\pi_k s_k^m \ln\left(\frac{\mu_k}{\mu}\right)$ represents the *m*-th *source contribution to the overall between-group inequality*, and the generic weight $\pi_k s_k^m = \frac{n_k}{n}\frac{\mu_k^m}{\mu}$ is the share of total income held by the *k*-th subgroup and corresponding to the *m* income source. Consistently with the definition of the pseudo-Theil $T(m)$, we can define $Tb(m)$ as the generic "between-group *pseudo*-Theil" (or mean *pseudo*-Theil). Note that $Tb(m)$ is zero whenever Tb is zero (i.e. $s_k = 1, \forall k$) but, differently from the standard subpopulation decomposition, the partial contribution of the *m* source (to the overall between-group inequality) can also be negative.

Following a similar procedure, but considering the individual income relations $y_{ik} = \sum_{m=1}^{M} y_{ik}^{m}$ instead than [11], we can disaggregate by income sources the within-group component of the Theil index as,

$$Tw = \sum_{k=1}^{K} \frac{n_k}{n} \frac{\mu_k}{\mu} \cdot \left(\frac{1}{n_k} \sum_{i=1}^{nk} \frac{y_{ik}}{\mu_k} \cdot \ln\left(\frac{y_{ik}}{\mu_k} \right) \right)$$

$$= \sum_{k=1}^{K} \left[\pi_k s_k \cdot \left(\frac{1}{n_k} \sum_{i=1}^{nk} \frac{\left(\sum_{m=1}^{M} y_{ik}^{m} \right)}{\mu_k} \ln\left(\frac{y_{ik}}{\mu_k} \right) \right) \right]$$

$$= \sum_{m=1}^{M} \left[\sum_{k=1}^{K} \pi_k s_k \cdot \left(\frac{1}{n_k} \sum_{i=1}^{nk} \frac{y_{ik}^{m}}{\mu_k} \ln\left(\frac{y_{ik}}{\mu_k} \right) \right) \right] = \sum_{m=1}^{M} Tw(m) \qquad [13]$$

where $Tw(m) = \sum_{k=1}^{K} \pi_k s_k \cdot \left(\frac{1}{n_k} \sum_{i=1}^{nk} \frac{y_{ik}^{m}}{\mu_k} \ln\left(\frac{y_{ik}}{\mu_k} \right) \right)$ is the weighted sum of the *K* group-pseudo-Theil $T_k(m) = \frac{1}{n_k} \sum_{i=1}^{nk} \frac{y_{ik}^{m}}{\mu_k} \ln\left(\frac{y_{ik}}{\mu_k} \right)$, and represents the *m*-th *source contribution to the overall within-group inequality*. Also in this case, we can define $Tw(m)$ the generic "within-group *pseudo*-Theil".

Expression [12] and [13] allow us to derive the following *subgroup-source* nested *decomposition* of the Theil index,

$$T(Y) = Tb + Tw = \sum_{m=1}^{M} Tb(m) + \sum_{m=1}^{M} Tw(m) \qquad [14]$$

where $Tb(m)$ and $Tw(m)$ represent, respectively, the contribution to between and within-group inequality which can be assigned to the *m*-th income component. The Theil bidimensional decomposition [14] naturally increases the set of inequality determinants, which are not directly observable when subgroup and income source decompositions are implemented separately. Even if this opportunity is also made possible by the existing multidimensional decomposition of Akita (2003), Conceição, et. al. (2000) and Wodon (1999), the chance of introducing income source determinants of overall inequality, must be seen as the major improvement of [14]. It provides the chance of inquiring many economic aspects (basically given by the "functional structure" of total income, separable into wages, profits and rents - as well as state factors) with other important economic and social issues. The gender (or ethnic) discrimination issue, the geographical uneven distribution of resources and the analysis of the impact of demographic structure on the overall level of inequality are only three among the possible frameworks of implementation of a bidimensional decomposition approach. Moreover, the possibility of controlling for fundamental state factors (such as transfers and taxes) provides a useful tool for a more accurate analysis of target policy effects on total inequality.

To conclude, let us note how the bidimensional decomposition [14] permits to achieve these objectives reducing the basic inequality determinants to very few and simple factors: the income and population shares and the group-source inequality. This is clearly not the case of the Gini bidimensional decomposition [9], which shows source-group contributions only partially functional to this role.

5. Empirical analysis

5.1. Data and methodology

In this paper we focus on disposable income, which is obtained by the sum of all work and transfer incomes of all individuals in each household. It is also an "equivalent" measure, being divided by the square root of the number of household member in order to take into account economies of scale. For what concerns outlier observations, the LIS recommendation is to impose 'bottom' and 'top codes' redefinition of incomes, so that all the observations below the 1% of equivalent mean income and all those above 10 times the median unequivalised income are substituted by the respective thresholds. In our case, the sum of all income sources has to equal the corresponding individual (household-based) total income, the redefinition of 'code' incomes into two thresholds would generate inconsistencies between the (adjusted) total income and the sum of its factor components. As a consequence, we have just erased the top and bottom codes of each distribution (0.1% at the bottom and at the top of the disposable income distribution).

Despite the SHIW Bank of Italy survey implemented in the LIS database covers a very long period of time[28], the analysis has been restricted to the period 1989-2000, in order to derive empirical findings able to be more easily extended with those of other countries present in the same database. In the following application we employed a slight different formulation of the nested Theil decomposition [14], because of the need of using weighted procedure which corrects for sample selection. The adjusted weighted formulation[29] of [14] is thus given by

$$T(Y)=\sum_{m=1}^{M}\left[\sum_{k=1}^{K}P_k\frac{\mu_{k(w)}^{m}}{\mu_{(w)}}\ln\frac{\mu_{k(w)}}{\mu_{(w)}}\right]+\sum_{m=1}^{M}\left[\sum_{k=1}^{K}P_k\frac{\mu_{k(w)}^{m}}{\mu_{(w)}}\cdot\left(\sum_{i=1}^{nk}p_i\frac{y_{ik}^{m}}{\mu_{k(w)}}\ln\left(\frac{y_{ik}}{\mu_{k(w)}}\right)\right)\right] \qquad [15]$$

where p_i represent the individual weights normalised for household members, P_k the sum of the k's group simple weights p_i ($i=1,....,n_k$), while $\mu_{(w)}$, $\mu_{k(w)}$, and $\mu_{k(w)}^{m}$ are the weighted means for total, k-th subgroup, and m-th source of the k-th subgroup distributions[30].

5.2. Empirical findings

The study of the relationship between demographic structure and economic inequality has been largely investigated in the recent literature. Many are the dimensions which one can examine in order to look for inequality determinants. Age distribution among population (or more detailed) household members, gender or regional factors, attained education, and worker/job types are only some among the individual characteristics usually taken into account in empirical works. Independently of the decomposition scope, meaning and possible interpretation[31], it is here interesting to note how the fundamental requirement is the possibility of splitting inequality as the sum of between-group-source and within-group-source contributions. This involves in turn the use of measures which support additive decomposability, and as amply verified in the previous sections, the Theil index fully satisfies this property.

In what follows we apply the nested inequality decomposition rule [15] to the Italian income distributions between 1989 and 2000. We take disposable income as the reference economic

[28] The survey started in 1965, even if only in the early Eighties became an important source of Italian households information on income and consumption. Brandolini and D'Alessio (2001), for instance, proposed inequality analysis between 1977 and 1995. After 1989 onward the sampling methodology did not change, the dimension of the sample does not vary much until now

[29] The weights are proportional to the actual population of the strata from which the sample observations are drawn from.

[30] Note that when the standard (no weighted) formulation is adopted we simply have $p_i=1/n_k$ and $P_k=n_k/n$.

[31] See Kanbur (2006) for a discussion of these points.

dimension, and considering the partition of the overall population into four subgroups defined by their geographical location[32].

Brandolini and D'Alessio (2001) examine the effects of demographic structure on the evolution of inequality in Italy between 1977 and 1995 applying mean logarithmic deviation decompositions. Their empirical results can be synthesized in the following three points:

1) inequality in disposable incomes between persons (disaggregated by several dimensions) showed considerable fluctuations but no particular medium-term tendency;

2) in the mid-1990s Italy was, together with the United Kingdom, the EU country with the highest inequality (this result has been amply confirmed, among the others, by the Italian Statistical Institute ISTAT, 2005);

3) Demographic effects on inequality appeared having been small: they played only a secondary role in defining the evolution of inequality in Italy, as well as in explaining the deviations from the levels recorded for other EU countries.

Table 5.1 shows the Theil index pattern for the period 1989-2000, as well as its disaggregation into the within and between regional components. Total inequality decreased of 7% between 1989 and 1991, but rapidly increased of nearly 40% during the next two years. This rise was followed by a first moderate reduction until 1995. Starting from this year inequality turned to rise until the decade highest level of 0.202 in 1998. Finally, in 2000 it went back to the middle decade level.

Table 5.1 - Disposable income inequality in Italy, 1989-2000: subgroup decomposition by geographical location.

Disposable income	Absolute values			Row %		
	Tb	Tw	Theil	Tb	Tw	Theil
1989	.020	.126	.146	13.4	86.6	100
1991	.017	.119	.136	12.2	87.8	100
1993	.021	.169	.190	11.0	89.0	100
1995	.022	.166	.188	11.5	88.5	100
1998	.022	.180	.202	11.0	89.0	100
2000	.021	.167	.188	11.2	88.8	100

Source: Own calculations on LIS database.

Table 5.1 also points out the contribution of the within and between territorial components. Quite differently from the "commonly idea" of an increasing Italian regional dualism, the within regions component of inequality seems to have more decisively driven the overall trend. More precisely, within-region inequality slightly increased its percentage contribution between 1989 and 1993

[32] The macro-regions composition considered in the elaboration is the following: (*North-West*) Piemeonte, Lombardia, Liguria, Trentino; (*North-East*) Veneto, Friuli Venezia Giulia, Emilia Romagna; (*Centre*) Toscana, Umbria, Marche. Lazio; (*South and Islands*) Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicila, Sardegna.

eroding part of the inequality influence imputable to the between factor. Thereafter, the two relative contributions remained nearly constant around the 12% (between component) and 88% (within component) in explaining total inequality. Despite this, a different key of interpretation arises if one goes through the specific components patterns. In fact, while the diminishing trend of total inequality observed in the first years of the decade can be ascribed to the joined contraction of both between and within components, the two most important inequality upward variations (between 1991-1993 and 1995-1998) are almost fully imputable to the widening within regions contribution.
In sum, Italian inequality resulted to be increased of 28.8% between 1989 and 2000. This overall trend was the result of a moderate increase in between-region inequality (5% over the decade) and a stronger impact of the within component (+33%). Saying it differently, total inequality in 1989 was for a 13.4% due to differences between geographical areas, as for a 86.6% to the level of their own unequal distributions; after 11 years the same percentages were 11% and 89% for the between and within components, respectively.
The territorial analysis of the Italian inequality fails to shed light on the economic determinants underlying overall trend. One could, surely, interprets the results of Tables 5.1 taking into account the regional characteristics of labour and capital markets, their specific productive structures, as well as the decentralised government interventions. All these keys of explanation are not, certainly, worthless. Despite this, the income sources decomposition allows such aspects to be better detected in a clear and consistent framework. In order to emphasize the added value that would come from a joined subpopulations-income sources analysis, let us briefly go through the following questions. Does the population age structure matter because the different source composition of personal incomes (from work, property and transfer components) in the households[33]? Does the standard analysis of the income gender gap hide very different female and male types of income received (for example, from work or capital activities)? How much of the regional determinants on national income inequality could be associated to the different impact of dependent (private or public) jobs autonomous jobs, financial incomes or different impact of state transfers and taxes? In what follows we try to give an answer to this kind of questions.
Before to discuss the results of the nested procedure application, let us propose those of a one-level factor component decomposition. Table 5.2 contains the proportional sources contribution to the Italian disposable income inequality during the period 1989-2000, following by the implementation of the Theil-based decomposition [10].
Some basic facts emerge:

[33] Brandolini and D'Alessio (2001) noted how the incomes of heads of household below the age of 40 worsened between 1977 and 1995, while it improved for heads aged over 65. Could be possible to strictly link this trend to the impact of different income sources within families?

- *Wages and salaries* show a fluctuant contribution around 45% of total inequality until 1995. They lost (in relative term) most of their importance in the second half of the decade, with the minimum of only 27% in 1998;
- Farm *self-employment incomes* seem not playing any substantial role, while the non farm autonomous component emerges as the most important determinant of inequality at the beginning and end of the 1990s;
- *Property incomes* account for around the 20% of total inequality in each year. This average contribution is quite interesting, mainly because of the low share of this component along the decade (6-9% of disposable income). The remarkable value of 29% in 1998 is partially due to its increased share over disposable income (9.4% in 1998 respect to an average of 7.3%);
- Finally, transfers component contributions have drastically reduced their (negative) influence between 1989 and 2000. Especially *Social benefits* diminished (in absolute terms) their impact from a -13.6% in 1989 to only -1.3% in 1995. In the second half of the decade they recovered part of the equalising effect, which in 2000 was slight over 4%.

Note that this evidence appears even worst if one look to the component share, which results continuously increased over the decade (from 13.3% to 19.6% of total income).

Table 5.2 - Source shares and contributions to total inequality in Italy. Disposable income decomposition, 1989-2000 (percentage values).

Inequality Contributions	**1989**	**1991**	**1993**	**1995**	**1998**	**2000**
Wages and salaries	45.6	42.9	47.0	43.8	26.8	35.0
Farm self-employment incomes	-0.9	-0.2	1.2	2.5	0.8	0.7
Non Farm self-employment incomes	54.0	48.4	34.6	29.9	43.0	44.0
Property incomes	17.7	20.1	21.8	20.4	29.2	23.1
Private Pensions	2.5	2.9	4.8	7.3	6.8	3.9
Social benefits	-13.6	-8.9	-5.1	-1.3	-3.8	-4.1
Other social transfers	-3.8	-2.5	-1.7	-2.2	-1.9	-2.0
Meansi & Privati	-1.5	-2.7	-2.5	-1.4	-1.1	-0.8
Cash transfers	.	.	.	1.1	0.3	0.2
Disposable Income (total)	100	100	100	100	100	100
Shares						
Wages and salaries	53.2	51.1	49.3	47.5	47.0	46.6
Farm self-employment incomes	1.1	1.3	0.8	1.1	0.7	0.9
Non Farm self-employment incomes	19.4	17.6	13.7	14.1	16.1	16.8
Property incomes	6.2	6.5	8.4	7.5	9.4	6.9
Private Pensions	3.6	4.6	5.9	6.2	5.7	5.9
Social benefits	13.3	15.4	18.0	19.6	18.0	19.6
Other social transfers	.	.	.	0.4	0.2	0.1
Meansi & Privati	1.9	2.0	1.9	2.0	1.9	1.9
Cash transfers	1.3	1.4	2.0	1.7	1.1	1.3
Disposable Income (total)	100	100	100	100	100	100

Source: Own calculations on LIS database.
Figures may not add up to total because of rounding.

The aim of the next application is to extend the analysis of inequality decomposition by income source also distinguishing among the population subgroups defined by geographical location. As shown by the following Tables 5.3 (a) and (b), the bidimensional decomposition of the Theil index permits to determine the contribution of:

- each component-group contribution to overall inequality;
- the role of each income source in defining the between and within subpopulation components on inequality (Col %);
- to split source contributions into a within and between sub-components (Row %).

In 1989 it emerges how the two sources contributing more to the overall disposable inequality, wages and non-farm autonomous incomes, were more affected by the within-regions differences rather then by the between component: the former represents, in fact, the 81.7% and the 96.2% of the two source contributions. The same conclusion holds for property incomes (with 84.9% of within contribution). For what concerns the social benefits the negative overall effect of -13.6% on total inequality is separated into territorial contributions of different sign! More precisely, while the within component accounts for -118% of the source contribution, indicating a more than proportional containing effect of social benefits within the regions, the between part of the subdecomposition says us that the same source was also accountable for a positive contribution on overall inequality.

In Table 5.2 it has already shown the proportional contribution of *wages and salaries* falling down between 1989 (45.6%) and 1998 (26.8%). Tables 5.3 allow to extend this general evidence. Note, for example, how their contribution to the between-region component strongly decreased until 1993 (passing from 62% to 29%) bud increased almost of two times in the second half of the decade (from 29% to 54.1%). These results are particularly important because of the crucial Italian reform of 1993: the abolition of "scala mobile" (the automatic adjustment of wages to inflation) in favour of the "concertazione" agreement[34].

The nested decomposition analysis says us that if this contributed to contain the relative impact of wages dispersion on disposable income inequality, on the other side it was not so effective in containing income differences between the Italian regions.

Great part of the overall inequality came also from the *autonomous incomes*, in particular from non-farm activities. Their contribution to the between-regions component increased in the fist half of the decade (from 15.4% to 34.7% in 1993) returning to go down until 2000 (19%). The within-region contribution of non-farm self-incomes followed, instead, an opposite trend, with a strong decrease in the first half of the period (from 60% to 30.3% in 1995) and an increasing one between 1995 and 2000 (47%).

[34] It found on the adoption of decentralised (at firm and territorial level) incentive mechanisms on wages as well as on their periodical (two years) adjusting with respect to target inflation.

Table 5.3 (a) - Source (percentage) contributions to total (*Theil*), between (*Tb*), and within (*Tw*) group inequality in Italy. Disposable income decomposition, 1989-1993 (*geographical location*).

1989	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***
	Absolute			**Col %**			**Row %**		
Wages and salaries	.012	.055	.067	62.1	43.0	45.6	18.3	81.7	100
Farm self-employment	-.002	.000	-.001	-7.7	0.2	-0.9	-116.0	16.0	-100
Non Farm self-employment	.003	.076	.079	15.4	60.0	54.0	3.8	96.2	100
Property incomes	.004	.022	.026	19.9	17.3	17.7	15.1	84.9	100
Private Pensions	.000	.004	.004	-0.3	2.9	2.5	-1.5	101.5	100
Social benefits	.004	-.023	-.020	18.3	-18.5	-13.6	18.0	-118.0	-100
Other social transfers	-.001	-.004	-.006	-5.4	-3.5	-3.8	-19.1	-80.9	-100
Meansi & Privati	.000	-.002	-.002	-2.4	-1.3	-1.5	-21.4	-78.6	-100
Cash transfers	.	.	.	.	.	.	.	.	.
Disposable Income (total)	.020	.127	.146	100	100	100	13.4	86.6	100
1991	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***
	Absolute			**Col %**			**Row %**		
Wages and salaries	.008	.051	.058	46.6	42.4	42.9	13.3	86.7	100
Farm self-employment	-.002	.002	.000	-14.9	1.8	-0.2	.	.	.
Non Farm self-employment	.005	.061	.066	28.2	51.2	48.4	7.1	92.9	100
Property incomes	.004	.023	.027	24.4	19.5	20.1	14.8	85.2	100
Private Pensions	-.001	.005	.004	-6.3	4.2	2.9	-26.2	126.2	100
Social benefits	.004	-.016	-.012	25.2	-13.7	-8.9	34.4	-134.4	-100
Other social transfers	-.001	-.003	-.003	-3.7	-2.3	-2.5	-18.2	-81.8	-100
Meansi & Privati	.000	-.004	-.004	0.6	-3.1	-2.7	2.6	-102.6	-100
Cash transfers	.	.	.	.	.	.	.	.	.
Disposable Income (total)	.017	.119	.136	100	100	100	12.2	87.8	100
1993	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***
	Absolute			**Col %**			**Row %**		
Wages and salaries	.006	.083	.089	28.9	49.2	47.0	6.8	93.2	100
Farm self-employment	.000	.002	.002	1.4	1.2	1.2	13.5	86.5	100
Non Farm self-employment	.007	.058	.066	34.7	34.6	34.6	11.1	88.9	100
Property incomes	.005	.036	.041	24.8	21.4	21.8	12.6	87.4	100
Private Pensions	.000	.010	.009	-2.3	5.6	4.8	-5.3	105.3	100
Social benefits	.004	-.014	-.010	20.2	-8.3	-5.1	43.5	-143.5	-100
Other social transfers	-.001	-.003	-.003	-3.2	-1.5	-1.7	-20.8	-79.2	-100
Meansi & Privati	-.001	-.004	-.005	-4.6	-2.2	-2.5	-20.7	-79.3	-100
Cash transfers	.	.	.	.	.	.	.	.	.
Disposable Income (total)	.021	.169	.190	100	100	100	11.0	89.0	100

Source: Own calculations on LIS database.

Figures may not add up to total because of rounding.

Table 5.3 (b) - Source (percentage) contributions to total (*T*), between (*Tb*), and within (*Tw*) group inequality in Italy. Disposable income decomposition, 1995-2000 (*geographical location*).

1995	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***
	Absolute			**Col %**			**Row %**		
Wages and salaries	.007	.075	.082	34.6	44.9	43.8	9.1	90.9	100
Farm self-employment	-.001	.005	.005	-2.5	3.1	2.5	-11.6	111.6	100
Non Farm self-employment	.006	.050	.056	27.2	30.3	29.9	10.5	89.5	100
Property incomes	.005	.033	.038	23.3	20.0	20.4	13.2	86.8	100
Private Pensions	.000	.014	.014	0.5	8.2	7.3	0.8	99.2	100
Social benefits	.005	-.008	-.002	24.8	-4.7	-1.3	215.5	-315.5	-100
Other social transfers	-.001	-.003	-.004	-4.7	-1.9	-2.2	-24.5	-75.5	-100
Meansi & Privati	-.001	-.002	-.003	-3.6	-1.1	-1.4	-29.4	-70.6	-100
Cash transfers	.000	.002	.002	0.4	1.2	1.1	3.7	96.3	100
Disposable Income (total)	.022	.166	.188	100	100	100	11.5	88.5	100
1998	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***
	Absolute			**Col %**			**Row %**		
Wages and salaries	.007	.047	.054	31.0	26.3	26.8	12.7	87.3	100
Farm self-employment	.000	.002	.002	-0.6	0.9	0.8	-8.0	108.0	100
Non Farm self-employment	.008	.079	.087	35.8	43.8	43.0	9.2	90.8	100
Property incomes	.007	.052	.059	31.2	29.0	29.2	11.8	88.2	100
Private Pensions	-.001	.014	.014	-3.4	8.0	6.8	-5.6	105.6	100
Social benefits	.003	-.011	-.008	13.9	-6.0	-3.8	39.9	-139.9	-100
Other social transfers	-.002	-.002	-.004	-7.6	-1.2	-1.9	-44.2	-55.8	-100
Meansi & Privati	.000	-.002	-.002	-0.3	-1.2	-1.1	-2.6	-97.4	-100
Cash transfers	.000	.001	.001	0.1	0.3	0.3	3.2	96.8	100
Disposable Income (total)	.022	.179	.202	100	100	100	11.0	89.0	100
2000	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***	***Tb***	***Tw***	***Theil***
	Absolute			**Col %**			**Row %**		
Wages and salaries	.011	.054	.066	54.1	32.6	35.0	17.4	82.6	100
Farm self-employment	.000	.001	.001	-0.5	0.9	0.7	-8.0	108.0	100
Non Farm self-employment	.004	.079	.083	18.9	47.1	44.0	4.8	95.2	100
Property incomes	.006	.038	.043	27.0	22.6	23.1	13.1	86.9	100
Private Pensions	-.001	.008	.007	-4.6	5.0	3.9	-13.1	113.1	100
Social benefits	.003	-.011	-.008	13.7	-6.3	-4.1	37.7	-137.7	-100
Other social transfers	-.001	-.003	-.004	-5.6	-1.5	-2.0	-31.8	-68.2	-100
Meansi & Privati	-.001	-.001	-.002	-3.6	-0.5	-0.8	-48.8	-51.2	-100
Cash transfers	.000	.000	.000	0.6	0.1	0.2	39.4	60.6	100
Disposable Income (total)	.021	.167	.188	100	100	100	11.2	88.8	100

Source: Own calculations on LIS database.
Figures may not add up to total because of rounding.

We have seen in Table 5.2 how *property incomes* considerably increased their contribution on overall inequality between 1989 (17.7%) and 2000 (23.1%), with the highest peak of 29.2% in 1998. While their influence on the within component closely followed that on total income, the impact on the between-regions inequality was in average higher. Despite this, also for this source of income the between component lost part of its relative importance along the decade, passing from the 15.1 of initial period until 11.8% in 1998.

Finally, let us dedicate a detailed discussion to the *social benefits* component. As already noted for the 1989 results, it is characterised by an unexpected conclusion. Its regularly negative overall effect hides dual sub-effects: in each year the negative (equalising) contribution to within-regions inequality is lowered by the positive impact on the between component. On this perspective, the nested decomposition allows a more detailed understanding of the diminishing role of this source in containing total inequality: between 1989 and 1995 the negative influence on the within component of total inequality diminished from -18.5% to -4.7%, while the positive influence on the between fraction increased from 18.3% to 24.8%. In the following two years the overall (negative) effect inverted its trend (from -1.3% to -4.1%) because of the inverted trends observed for its sub-regional components.

6. Concluding remarks

If one of the objectives of introducing inequality decomposition by income sources is to control for the impact of component shares and spreads on total inequality, then the natural choice should be that of using the approach which better provides this chance. In particular, we propose the Theil index of inequality as the measure able (better than others) to account for this objective. This is mainly due to: *i*) its property of perfect decomposability by population subgroups, *ii*) the desirable proprieties that its decomposition by income sources possesses (*uniform addition* and the two new *dynamic principles* A *and* B); *iii*) its very attractive formulation as function of income and population shares. This last point, in particular, should be further investigated, because the implications it implies in favour of two fundamental economic issues: *a*) the necessity of a linkage between functional and personal income distribution; *b*) the important role it can assume for policy evaluation (or prediction) analysis.

The relevance of our analysis lies on the following three main aspects:

- by a methodological point of view, it stimulates the recently "drowsy" debate on the decomposition problem, which has been abundantly animated around the early Eighties as much as excessively neglected (except that in the innumerable empirical applications) during the last fifteen years;
- it suggests (or simply acts as an initial attempt of proposing) possible extensions of the standard analysis, also as "crossing" element of connection between different theories and empirical approaches.
- finally, it provides new suggestions under a political economy perspective, because of its natural predisposition to be used as a political decision instrument. Decentralization, fiscal federalism and other fundamental policy issues could be properly investigated in relation to the unequal distribution over the space of activities and resources. Country specific inquiries

as well as European integration of markets and institutions could give rise to interesting (and unexplored) lines of research.

With respect to this last point, could a government decision be more effective in containing total inequality, acting on specific income sources (wages, capital incomes, transfers and/or taxes) and at the same time "targeting" its political decisions in order to confine inequality increasing on the components playing the major influence? Could an "optimal" combination of the two inequality dimensions (subgroups and income sources) lead to an "optimal" choice? Given a welfare improvement target, could a specific combination of policies imply lower public spending?

We think that one among the best way of answering to these plausible questions is given by the Theil index of inequality, by its desirable proprieties of subgroup and income source decomposition, as well as by the possibility of linkage with crucial and directly observable exogenous variables: the source-subgroup income and population shares; the source-subgroup income inequality.

The twofold explication provided by the nested decomposition of inequality seems to be very important under a political point of view as a useful instrument for policy makers. Politics might be well assisted in planning economic (targeting) intervention able to contain overall inequality through multi-directional optimal setting of their options. Note also that similar studies could be naturally extended to other contexts, such as the analysis of the determinants of the income gender gap, or the role played by the household composition (number of earners and types of income received), and so on.

A proper context of analysis would be that of decomposition overall European inequality by factor components and territorial dimension (the groups being its own member countries). The empirical findings could furnish new evidences about the between/within European countries determinants of inequality coming from market and state redistribution of income.

References

AKITA T. (2003) "Decomposing regional income inequality in China and Indonesia using two-stage nested Theil decomposition method," *The Annals of Regional Science*. 37 (1), pp. 55-77.

BOURGUIGNON F. (1979) "Decomposable income inequality measures," *Econometrica*, 47 (4), pp. 901-920.

BRANDOLINI A., AND G. D'ALESSIO (2001) "Household structure and income inequality," *Luxemburg Income Study* Working Paper No. 254.

CONCEIÇÃO P., J.K. GALBRAITH, AND P. BRADFORD (2000) "The Theil index in sequences of nested and hierarchic grouping structures: implications for the measurement of inequality through time with data aggregated at different levels of industrial classification," UTIP Working Paper No. 15.

COWELL F.A. (1980) "On the structure of additive inequality measures," *Review of Economic Studies* 47, pp. 521-531.

DAGUM C. (1997) "A new approach to the decomposition of the Gini income inequality ratio," *Empirical Economics*, 22, pp. 515-531.

DEUTSCH J., AND J. SILBER (2004) "Measuring the impact of various income sources on the link between inequality and development: implication for the Kuznets curve," *Review of Development Economics* 8 (1), 110-127.

FEI L.C.H., G. RANIS, AND S.W.Y. KUO (1978) "Growth and family distribution of income by factor components," *Quarterly Journal of Economics*, 92, pp. 17-53.

FIELDS G.S. (2002) "Income inequality in urban Colombia: a decomposition analysis," *Review of Income and Wealth*, 25, pp. 327-341.

FIORIO C.V. (2002) "Effects of earning and pension income disparities on Italian household income distribution," XIV Siep Conference.

GIAMMATTEO M. (2007), Inequality decomposition by income sources: a theoretical and empirical analysis, Ph.D. dissertation, Università degli studi di Roma "La Sapienza".

ISTAT (2006) "Rapporto annuale sulla situazione del paese."

KANBUR R. (2006) "The policy significance of inequality decompositions," *Journal of Economic Inequality*, 4 (3), pp. 367-374.

LAMBERT P.J. AND R.J. ARONSON (1993) "Inequality Decomposition Analysis and the Gini Coefficient revisited," *Economic Journal*, 103, pp. 1221-1227.

LEIBBRANDT M.V., C.D.WOOLARD, I.D. WOOLARD (1996) "The contribution of income components to income inequality in South Africa," World Bank LSMS WP n.125.

LERMAN R.I. (1999) "How do Income Sources affect Income Inequality?" in J. Silber (ed.) *Handbook of Income Inequality Measurement*, Boston: Kluwer Academic Press, pp. 341-362.

LERMAN R.I., AND S. YITZHAKI (1985) "Income inequality effects by income source. A new approach and applications to the US," *Review of Economics and Statistics* 67, pp. 151-156.

LERMAN R.I., AND S. YITZHAKI (1991) "Income stratification and income inequality," *Review of income and Wealth* 37, pp. 313-329.

LERMAN R.I., AND S. YITZHAKI (1994) "Effects of marginal changes in income source on U.S. income inequality," *Public Finance Quarterly* 22 (4), pp. 403-417.

MORDUCH J., AND T. SICULAR (2002) "Rethinking inequality decomposition with evidence from rural China," *The Economic Journal*, 112, pp. 93-106.

MUSSARD S. (2004) "The bidimensional decomposition of the Gini index. A case study: Italy," *Applied Economics Letters*, 11, pp. 503-05.

MUSSARD S., F. SEYTE, AND M. TERRAZA, (2003) "Decomposition of Gini and the generalized entropy inequality measures," *Economics Bulletin*, 4 (7), pp. 1−6.

PAUL S. (2004) "Income source effects on inequality," *Journal of Development Economics* 73, pp. 435-451.

PODDER N. (1993) "The disaggregation of the Gini coefficient by factor components and its applications to Australia," *Review of Income and Wealth* 39 (1), pp. 51–61.

PODDER N. AND S. CHATTERJEE (2002) "Income inequality trends in terms of income sources," *Journal of Public Economics* 86, pp. 1–27.

PODDER N., P. MUKHOPADHAYA (2001) "The changing pattern of sources of income and its impact on inequality: the method and its application to Australia, 1975-94," *Economic Record* 77, n.238, pp. 242–251.

PYATT G., C. CHEN, AND J. FEI (1980) "The distribution of income by factor components", *Quarterly Journal of Economics* 94, pp. 451-473.

RAO V.M. (1969) "Two decomposition of concentration ratio," *Journal of the Royal Statistical Society*, 132, pp. 418-425.

SASTRE M., AND A. TRANNOY (2001) "Shapley inequality decomposition by factors components," Papeles de Trabajo del Instituto de Estudios Fiscales, Madrid.

SHORROCKS A.F. (1980) "The class of additively decomposable inequality measures," *Econometrica* 48, pp. 613-625.

SHORROCKS A.F. (1982) "Inequality decomposition by factor components," *Econometrica* 50 (1), pp. 193-212.

SHORROCKS A.F. (1983) "The impact of income components on the distribution of family incomes," *Quarterly Journal of Economics* 98 (2), 311-326.

SHORROCKS A.F. (1984) "Inequality decomposition by population sub-groups," Econometrica 52, 1369.1385.

SHORROCKS A.F. (1999) "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value," Unpublished manuscript. Department of Economics, University of Essex.

SILBER J. (1993) "Inequality decomposition by income source: a note," *Review of Economics and Statistics* 75 (3), pp. 545-547.

STARK O., J.E. TAYLOR, S. YITZHAKI (1986) "Remittances and inequality," *Economic Journal* 96, pp. 722-740.

THEIL, H. (1967) "Economics and information theory," Amsterdam: North Holland.

WAN G.H. (2001) "Changes in regional inequality in rural China: decomposing the Gini index by income sources," *Australian Journal of Agricultural and Resource Economics* 45 (3), pp. 361-381.

WODON Q.T. (1999) "Between group and target transfers," *Review of Income and Wealth* 45 (1) pp. 21-39.