J Econ Inequal (2011) 9:509–528
DOI 10.1007/s10888-011-9176-1

Inequality decompositions—a reconciliation

Frank A. Cowell · Carlo V. Fiorio

Received: 26 January 2010 / Accepted: 29 March 2011 / Published online: 30 April 2011

Abstract We show how classic source-decomposition and subgroup-decomposition methods can be reconciled with regression methodology used in the recent literature. We also highlight some pitfalls that arise from uncritical use of the regression approach. The LIS database is used to compare the approaches using an analysis of the changing contributions to inequality in the United States and Finland.

Keywords Inequality · Decomposition

JEL Classification D63

1 Introduction

What is the point of decomposing income inequality and how should we do it? For some researchers the questions resolve essentially to a series of formal propositions that characterise a particular class of inequality measures. For others the issues are essentially pragmatic: in the same way as one attempts to understand the factors underlying, say, wage discrimination [2] one is also interested in the factors underlying income inequality and it might seem reasonable to use the same sort of applied econometric method of investigation. Clearly, although theorists and pragmatists are both talking about the components of inequality, they could be talking about very different things. We might even wonder whether they are on speaking terms.

F. A. Cowell (✉)
STICERD, London School of Economics, Houghton Street, London WC2A 2AE, UK
e-mail: f.cowell@lse.ac.uk

C. V. Fiorio
DEAS and Econpubblica, University of Milan, via Conservatorio, 7, 20133 Milan, Italy
e-mail: carlo.fiorio@unimi.it

The worry is that the standard theoretical approach, that employs *a priori* reasoning, and recent empirical approaches, that employ an application of regression analysis, are founded upon independent and possibly conflicting bases. Could they therefore provide conflicting messages to researchers and policy makers? However, although the main strands of literature on inequality decomposition have developed separately, this does not mean that they are necessarily inconsistent. It could be the case that at the core of each of the approaches there is an essential common element that can be used to establish a relationship between the principal approaches—the "reconciliation" mentioned in our title. In this paper we show how the two main methods of decomposition analysis (that are often treated as entirely separate) can be developed within a common analytical framework. We investigate regression-based techniques that are commonly used in empirical applications in various fields of economics and show how the methodology required for this can be derived from the *a priori* approach to factor- and source-decomposition. We apply these techniques to data from the Luxembourg Income Study to illustrate how the reconciliation works in practice.

The paper is organised as follows. Section 2 offers an overview of the decomposition literature. Our basic model is developed in Section 3 and this is developed into a treatment of factor-source decomposition and subgroup decomposition in Sections 4 and 5 respectively. Section 6 provides the empirical application, Section 7 discusses related literature and Section 8 concludes.

2 Approaches to decomposition

The two main strands of inequality-decomposition analysis that we mentioned in the introduction could be broadly labelled as "*a priori* approaches" and "regression models."

2.1 *A priori* approaches

Underlying this approach is the essential question "what is meant by inequality decomposition?" The answer to this question is established through an appropriate axiomatisation.

This way of characterising the problem is perhaps most familiar in terms of decomposition by subgroups. A coherent approach to subgroup decomposition essentially requires (1) the specification of a collection of admissible partitions—ways of dividing up the population into mutually exclusive and exhaustive subsets and (2) a concept of representative income for each group. Requirement (1) usually involves taking as a valid partition any arbitrary grouping of population members, although other specifications also make sense [15]; requirement (2) is usually met by taking subgroup-mean income as being representative of the group, although other representative income concepts have been considered [1, 20, 21, 25, 26]. A minimal requirement for an inequality measure to be used for decomposition analysis is that it must satisfy a subgroup consistency or aggregability condition—if inequality in a component subgroup increases then this implies, ceteris paribus, that inequality overall goes up [36, 37]; the "ceteris paribus" clause involves a condition that the subgroup-representative incomes remain unchanged. This allows one to screen out some inequality measures that do not even satisfy the minimal requirement [11], but one can go further. By imposing more structure—i.e., further conditions—on the decomposition method one can derive particular inequality indices with convenient properties [3, 10, 33], a consistent procedure for accounting for inequality trends [23] and an exact decomposition method that can be applied for example to regions [40] or to the world income distribution [30]. By using progressively finer partitions it is possible to apply the subgroup-decomposition approach to a method of accounting for the contributory factors to inequality [13, 16].

The *a priori* approach is also applicable to the other principal type of decomposability—the break-down by factor-source [29, 34, 35, 39]. As we will see the formal requirements for factor-source decomposition are straightforward and the decomposition method in practice has a certain amount in common with decomposition by population subgroups. Furthermore the linear structure of the decomposition (given that income components sum to total income) means that the formal factor-source problem has elements in common with the regression-analysis approach that we review in Section 2.2.

Relatively few attempts have been made to construct a single framework for both principal types of decomposition—by subgroup and by factor source. A notable exception is the Shapley-value decomposition [8, 38], which defines an inequality measure as an aggregation (ideally a sum) of a set of contributory factors, whose marginal effects are accounted eliminating each of them in sequence and computing the average of the marginal contributions in all possible elimination sequences. However, despite its internal consistency and attractive interpretation, the Shapley-value decomposition in empirical applications raises some dilemmas that cannot be solved on purely theoretical grounds. As argued by Sastre and Trannoy [32], if all ambiguities about different possible marginalistic interpretations of the Shapley rule are cleared up, this decomposition is dependent on the aggregation level of remaining income components and is highly nonrobust. Some refinements have been proposed to improve the Shapley inequality decomposition, including the Nested Shapley [8] and the Owen decomposition [38], based on defining a hierarchical structure of incomes. However, these solutions might face some difficulty in finding a sensible economic interpretation and some empirical "solutions" only circumvent the problem without solving it [31, 32]. Charpentier and Mussard [9] have also shown that results derived from the Shapley value are limited in their applicability.

2.2 Regression models

The second analytical strand of analysis that concerns us here derives from a mainstream econometric tradition in applied economics. Perhaps the richest method within this strand is the development of a structural model for inequality decomposition exemplified by Bourguignon et al. [4, 5], in the tradition of the DiNardo et al. [14] approach to analysing the distribution of wages. This method carefully specifies a counterfactual in order to examine the influence of each supposedly causal factor. However, its attractiveness comes at a price: a common criticism is that it is data hungry and, as such, it may be unsuitable in many empirical applications.

Furthermore, the modelling procedure can be cumbersome and is likely to be sensitive to model specification.

A less ambitious version of the regression-model approach is the use of a simple regression model as in Fields [17], Fields and Yoo [18] and Morduch and Sicular [27]. As with the structural models just mentioned, simple regression models enjoy one special advantage over the methods reviewed in Section 2.1. Potential influences on inequality that might require separate modelling as decomposition by groups or by income components can usually be easily and uniformly incorporated within an econometric model by appropriate specification of the independent variables.

2.3 An integrated approach?

It is evident that, with some care in modelling and interpretation, the *a priori* method can be developed from an exercise in logic to an economic tool that can be used to address important questions that are relevant to policy making. One can use the subgroup-decomposition method to assign importance to personal, social or other characteristics that may be considered to affect overall inequality. The essential step involves the way that between-group inequality is treated which, in turn, focuses on the types of partition that are considered relevant. One has to be careful: the fact that there is a higher between-group component for decomposition using partition A rather than partition B does not necessarily mean that A has more significance for policy rather than B [24]. However, despite this caveat, it is clear that there should be some connection between the between-group/within-group breakdown in the Section 2.1 approach and the explained/unexplained variation in the Section 2.2 approach.

We want to examine this connection using a fairly basic model.

3 Basic model

To make progress it is necessary focus on the bridge between formal analysis and the appropriate treatment of data. Hence we introduce the idea of data generating process (DGP), i.e., the joint probability distribution that is supposed to characterize the entire population from which the data set has been drawn.

Consider a set of random variables $\mathbf{H}$ with a given joint distribution $F(\mathbf{H})$, where $\mathbf{H}$ is partitioned into $[Y, \mathbf{X}]$ and $\mathbf{X} := (X_1, X_2, ..., X_K)$. Assume that we aim to model Y as a function of $\mathbf{X}$ and a purely random disturbance variable U and that we can write the relation in an explicit form with Y as function of $(\mathbf{X}, U)$

$$Y = f(\mathbf{X}, U|\beta) \tag{1}$$

where $\beta := (\beta_1, ..., \beta_K)'$ is a vector of parameters. For example, we could think of Y as individual income, of $\mathbf{X}$ as a set of observable individual characteristics, such as age, sex, education, and of U as an unobservable random variable such as ability or luck.

For simplicity let us assume that the DGP takes a linear form and that the number of observable characteristics is k. Hence, we can write:

$$Y = \beta_0 + \sum_{k=1}^{K} \beta_k X_k + U \tag{2}$$

Typically one observes a random sample of size n from $F(\mathbf{H})$,

$$\{(y_i, \mathbf{x}_i) = (y_i, x_{1i}, ...x_{ki}), i = 1, ...n\},$$

where the observations are independent over i. One then generates predictions of income for assigned values of individual characteristics using regression methods to compute a vector $\mathbf{b}$, as an estimate of β. The true marginal distribution function of each random variable, which might be either continuous or discrete, is often unknown in economic applications, as data do not come from laboratory experiments, and one only knows the empirical distribution functions (EDF). The sample analogue of model 2 can be written as:

$$y = \beta_0 + \sum_{k=1}^{K} \beta_k x_k + \upsilon,$$

where υ is the residual term. Provided that the standard assumptions such as exogenous covariates and spherical error variance hold, one could use OLS methods to estimate the income model[1] obtaining

$$y = b_0 + \sum_{k=1}^{K} b_k x_k + u, \tag{3}$$

where b_k is the OLS estimate of $\beta_k, k = 0, ...K, u = y - \left(b_0 + \sum_{k=1}^{K} b_k x_k\right)$ is the OLS residual.

Using the upper case letter for denoting a random variable (whose distribution function is not known in typical survey settings) and the lower case letter for denoting a size-n random sample from the same distribution function, the mean and inequality function of Y are denoted by $\mu(Y)$ and $I(Y)$, the mean and the inequality statistics (i.e., functions of the data) with $\mu(y) = \mu(y_1, ..., y_n)$ and $I(y) = I(y_1, ..., y_n)$.

We can analyse the structure of the inequality of y (or of Y) in two different ways

- *Subgroup decomposition*. Suppose that a subset $T \subseteq \{1, ..., K\}$ of the observables consists of discrete variables such that x_k (X_k) can take the values ξ_{kj}, $j = 1, .., t_k$ where $k \in T$ and t_k is the number of values (categories) that can logically be taken by the kth discrete observable. Then in this case we could perform a decomposition by population subgroups, where the subgroups are determined by the t categories, where $t := \prod_{k \in T} t_k$. This decomposition could be informative—what you get from the within-group component is an aggregate of the amount of inequality that is attributable to the dispersion of the unobservable υ (U) and the

[1]We use a standard OLS regression for simplicity of exposition. Other regression methods that employ a distance metric taken from an inequality index could also be used [28].

remaining continuous observables x_k, $k \notin T$ (X_k, $k \notin T$). If all the observables were discrete the within-group component would be an aggregation of $I_{y|x}$ ($I_{Y|X}$) and the between-group component would give the amount of inequality that would arise if there were no variation in υ (U).

- *Factor-source decomposition*. We can also interpret (Eq. 2) as the basis for inequality by factor source expressing $I(Y)$ in terms of component incomes $C_1, ..., C_{K+1}$, where

$$C_k := \beta_k X_k, k = 1, ..., K \tag{4}$$

$$C_{K+1} := U \tag{5}$$

  Notice that the constant term β_0 does not contribute to $I(y)$ and, similarly, if one adds or subtract an arbitrary constant to or from a regressor this will only change the constant with no effect on total inequality. For more details, see Section 4 below.

The application of these decomposition methods has been criticised on a number of grounds. Subgroup decomposition is criticised because it requires partitioning the population into discrete categories although some factors (for example, age) are clearly continuous variables. Moreover, handling more than very few subgroups at the same time can be cumbersome. The factor-source decomposition presented in the Shorrocks [34] form presents the useful property of being invariant to the inequality measure adopted,[2] however it can be criticised as being limited to a natural decomposition rule where total income is the sum of different types of income (for example pension, employment income and capital income). The subgroup and factor-source decomposition methods are sometimes criticised as being purely descriptive rather than analytical and as being irreconcilable one with another. Moreover they are tools which are often not well known in some fields of economics where the main focus is on the determinants of income or the market price of personal characteristics, which are estimated as the OLS coefficient in a Mincer-type wage regression.

The two decomposition methods—by population subgroup and by factor source—can be shown to be related to each other. This can be conveniently done using the model that we have just introduced.

4 Decomposition by factor source

Equation 2 is analogous to the case analysed by Shorrocks [34] where income is the sum of income components (such as labour income, transfers and so on). The inequality of total income, $I(Y)$, can be written using a natural decomposition rule such as:

$$I(Y) = \sum_{k=1}^{K+1} \Theta_k \tag{6}$$

[2] Actually in some situations this might be regarded as a shortcoming, especially when the change of inequality can have a different sign depending on the inequality measure adopted.

where Θ_k depends on C_k and can be regarded as the contribution of factor k to overall income inequality. Define also the proportional contribution of factor k to inequality

$$\theta_k := \frac{\Theta_k}{I(Y)}.$$

Using Eqs. 4 and 5 the results in Shorrocks [34] yield:

$$\theta_k = \frac{\sigma(C_k, Y)}{\sigma^2(Y)} = \frac{\sigma^2(C_k)}{\sigma^2(Y)} + \sum_{j \neq k}^{K+1} \rho(C_k, C_j) \frac{\sigma(C_k)\sigma(C_j)}{\sigma^2(Y)}, k = 1, ..., K+1$$

where $\sigma(X) := \sqrt{\text{var}(X)}$, $\sigma(X, Y) := \text{cov}(X, Y)$ and $\rho(C_i, C_j) := \text{corr}(C_i, C_j)$.

Since $\sigma(\beta_k X_k, Y) = \beta_k \sigma(X_k, Y)$ we have:

$$\theta_k = \beta_k^2 \frac{\sigma^2(X_k)}{\sigma^2(Y)} + \sum_{j \neq k}^{K+1} \beta_k \beta_j \frac{\sigma\left(X_k, X_j\right)}{\sigma^2(Y)} + \beta_k \frac{\sigma(X_k, U)}{\sigma^2(Y)} \tag{7}$$

from which we obtain

$$\theta_k = \beta_k^2 \frac{\sigma^2(X_k)}{\sigma^2(Y)} + \sum_{j \neq k}^{K+1} \beta_k \beta_j \rho(X_k, X_j) \frac{\sigma(X_j)\sigma(X_k)}{\sigma^2(Y)} + \beta_k \rho(X_k, U) \frac{\sigma(X_k)\sigma(U)}{\sigma^2(Y)}, \tag{8}$$

for $k = 1, ..., K$ and

$$\theta_{K+1} = \frac{\sigma^2(U)}{\sigma^2(Y)} + \sum_{k=1}^{K} \beta_k \rho(X_k, U) \frac{\sigma(X_k)\sigma(U)}{\sigma^2(Y)}. \tag{9}$$

Replacing β_k by its OLS estimate (b_k), and variances, covariances and correlation by their unbiased sample analogues, the estimate of θ_k, (z_k), can be obtained. A similar approach was followed by Fields [17]. Equations 8 and 9 provide a simple and intuitive interpretation and allow one to discuss the contribution of the value of characteristic k, c_k, to inequality $I(y)$. If we impose more structure on the problem, by assuming that there is no multicollinearity among regressors and all regressors are non-endogenous ($\text{corr}(C_j, C_r) = 0$, $j \neq r$ and $j, r = 1, ..., K, K+1$), then Eq. 7 can be simplified to

$$\theta_k = \begin{cases} \beta_k^2 \dfrac{\sigma^2(X_k)}{\sigma^2(Y)}, k = 1, ..., K \\ \dfrac{\sigma^2(U)}{\sigma^2(Y)}, k = K+1 \end{cases} \tag{10}$$

and it can be estimated as

$$z_k = \begin{cases} b_k^2 \dfrac{\sigma^2(x_k)}{\sigma^2(y)}, k = 1, ..., K \\ \dfrac{\sigma^2(u)}{\sigma^2(y)}, k = K+1 \end{cases} \tag{11}$$

where $\sigma^2(x_k), \sigma^2(y), \sigma^2(u)$ stand for the unbiased sample variance of x_k, y, u, respectively. The sample analogue of the inequality decomposition as in Eq. 6 can be written as:

$$I(y) = \sum_{k=1}^{K+1} Z_k = \sum_{k=1}^{K+1} I(y) z_k = \sum_{k=1}^{K} I(y) b_k^2 \frac{\sigma^2(x_k)}{\sigma^2(y)} + I(y) \frac{\sigma^2(u)}{\sigma^2(y)}. \tag{12}$$

With some simplification, the right-hand-side of Eq. 12 might be interpreted as the sum of the effects of the K characteristics and of the error term, although one should consider it as the sum of the *total value* of the K characteristics, i.e. the product of its "price" of each component as estimated in the income regression ($b_k, k = 1, ..., K$) and its quantity ($x_k, k = 1, ..., K$). One should also notice that the standard errors of Eq. 12 are not trivial to compute as they involve the ratio of variances of random variables coming from a joint distribution and the variance of inequality indices can be rather cumbersome to derive analytically (see for instance [12]). Simulation methods such as the bootstrap are suggested for derivation of standard errors of Eq. 12, although they are not presented for the empirical analysis which follows.

Equation 7 shows that θ_k ($k = 1, ..., K$) can only be negative if

$$\beta_k \left(\sum\nolimits_{j \neq k} \beta_j \sigma \left(X_k, X_j \right) + \sigma \left(X_k, U \right) \right) < -\beta_k^2 \sigma^2 \left(X_k \right), k = 1, ... K$$

for which a necessary condition is that there be either a nonzero correlation among RHS variables or at least one endogenous RHS variable.

It should be noted here that the decomposition (Eq. 6) applies for natural decompositions only, i.e., if the LHS variable can be represented as a sum of factors. In the labour-economics literature it is customary to estimate a log-linear relation, such as

$$\log(y) = b_0 + \sum_{k=1}^{K} b_k x_k + u$$

based on theoretical models of human capital, arguments of better regression fit, or error properties. In this case, the decomposition (Eq. 6) can only be undertaken with $I(\log(y))$ on the LHS.

5 Decomposition by population subgroups

Let us now assume that X_1 is a discrete random variable that can take only a finite number of values $\{j = 1, ..., t_1\}$. Let $X_{k,j} := \iota \cdot X_k$, where ι is an indicator function which is equal to one if $X_1 = j$ and is equal to zero otherwise. Equation 2 can be represented for each sub-group j as:

$$Y_j = \beta_{0,j} + \beta_{1,j} X_{1,j} + \sum_{k=2}^{K} \beta_{k,j} X_{k,j} + U_j \tag{13}$$

Define $P_j = \Pr(X_1 = X_{1,j})$, the proportion of the population for which $X_1 = X_{1,j}$. Then within-group inequality can be written as

$$I_w(Y) = \sum_{j=1}^{t_1} W_j I(Y_j)\,, \tag{14}$$

where t_1 is the number of groups considered, W_j is a weight that is a function of the P_j, and Y_j. The decomposition by population subgroups allows one to write:

$$I(Y) = I_b(Y) + I_w(Y), \tag{15}$$

where I_b is between-group inequality, implicitly defined by Eqs. 14 and 15 as

$$I_b(Y) := I(Y) - \sum_{j=1}^{t_1} W_j I\left(Y_j\right).$$

In the case of the Generalised Entropy (GE) indices we have, for any $\alpha \in (-\infty, \infty)$,

$$W_j = P_j \left[\frac{\mu\left(Y_j\right)}{\mu(Y)}\right]^{\alpha} = R_j^{\alpha} P_j^{1-\alpha}, \tag{16}$$

where $R_j := P_j \mu(Y_j)/\mu(Y)$ is the income share of group j, $\mu(Y_j)$ is mean income for subgroup j, $\mu(Y)$ is mean income for the whole population; we also have

$$I(Y) = \frac{1}{\alpha^2 - \alpha}\left[\int \left[\frac{Y}{\mu(Y)}\right]^{\alpha} dF(Y) - 1\right], \tag{17}$$

from which we obtain

$$I_w(Y) = \frac{1}{\alpha^2 - \alpha}\left[\sum_{j=1}^{t_1} P_j \left[\frac{\mu\left(Y_j\right)}{\mu(Y)}\right]^{\alpha} \int \left[\frac{Y_j}{\mu\left(Y_j\right)}\right]^{\alpha} dF(Y_j) - 1\right] \tag{18}$$

and

$$I_b(Y) = \frac{1}{\alpha^2 - \alpha}\left[\sum_{j=1}^{t_1} P_j \left[\frac{\mu\left(Y_j\right)}{\mu(Y)}\right]^{\alpha} - 1\right]. \tag{19}$$

Let us now see how decomposition by population subgroups could be adapted to an approach which uses the estimated DGP. Assuming that all standard OLS conditions are fulfilled, and using a $n-$size random sample $\mathbf{y}, \mathbf{x}_1, .., \mathbf{x}_k$ from the joint distribution function $F(Y, X_1, .., X_k)$ one can estimate Eq. 13 by using dummy variables for identifying different groups obtaining:

$$y_j = b_{0j} + \sum_{k=2}^{K} b_{kj} x_{k,j} + u_j \tag{20}$$

where $b_{0,j}$ are OLS estimates of $\beta_{0,j} + \beta_{1,j}\mu(x_{1,j})$ in subsample j and u_j are the OLS residuals of each group.

Given the OLS assumptions, the unbiasedness property of OLS estimates allows one to write the mean of y_j in Eq. 20 as $\mu(y_j) = b_{0,j} + \sum_{k=2}^{K} b_{k,j}\mu(x_{k,j})$.

The estimated between-group inequality I_b can then be written as:

$$I_{\rm b}(y) = \frac{1}{\alpha^2 - \alpha}\left[\sum_{j=1}^{t_1} p_j \left[\frac{b_{0,j} + \sum_{k=2}^{K} b_k \mu(x_{k,j})}{b_0 + \sum_{k=1}^{K} b_k \mu(x_k)}\right]^{\alpha} - 1\right] \tag{21}$$

where $p_j := n_j/n$ is the population share and n_j is the size of group j. The estimated within-group inequality, using Eq. 12 to decompose $I(y_j)$, is written as:

$$I_{\rm w}(y) = \sum_{j=1}^{t_1} w_j I(y_j) \left(\sum_{k=2}^{K} \frac{b_{k,j}^2 \sigma^2(x_{k,j}) + \sigma^2(u_j)}{\sigma^2(y_j)}\right) \tag{22}$$

where $w_j = (q_j)^{\alpha}(p_j)^{1-\alpha}$ and $q_j := p_j \mu(y_j)/\mu(y)$ is the income share of group j.

In the general case, allowing for the possibility that $\mathrm{corr}(X_{1,j}, X_{k,j}) \neq 0$ and that $\mathrm{corr}(X_{1,j}, U) \neq 0$, decomposition by subgroups is now:

$$I_{\rm b}(y) = \frac{1}{\alpha^2 - \alpha}\left[\sum_{j=1}^{t_1} p_j \left[\frac{b_{0,j} + \sum_{k=2}^{K} b_{k,j} \mu(x_{k,j})}{b_0 + \sum_{k=1}^{K} b_k \mu(x_k)}\right]^{\alpha} - 1\right] \tag{23}$$

where $b_{0,j}$ is now the OLS estimate of $\beta_{0,j} + \beta_{1,j}\mu(x_{1,j}) = \beta_{0,j} + \beta_{1,j} \cdot j$ and

$$\begin{aligned} I_{\rm w}(y) = \sum_{j=1}^{t_1} w_j I(y_j) \Bigg[\Bigg(\sum_{k=2}^{K} b_{k,j}^2 \frac{\sigma^2(x_{k,j})}{\sigma^2(y_j)} + b_{k,j} \sum_{r \neq k} b_{r,j} \rho(x_{r,j}, x_k) \frac{\sigma(x_{r,j})\sigma(x_k)}{\sigma(y_j)} \\ + b_{k,j}\rho(x_{k,j}, u_j) \frac{\sigma(x_{k,j})\sigma(u_j)}{\sigma(y_j)} + b_{k,j} \frac{\sigma\left(x_{k,j}, u_j\right)}{\sigma^2\left(y_j\right)} \Bigg) + \frac{\sigma^2(u_j)}{\sigma^2(y_j)} \Bigg]. \end{aligned} \tag{24}$$

Using a similar notation of that introduced in Section 4, we can write $I(y_j) = \sum_{k=1}^{K+1} \Theta_{jk}$ and $\theta_{jk} := \Theta_{jk}/I(y_j)$ and rewrite the within-group inequality as

$$I_w(y) = \sum_{j=1}^{t_1} w_j \sum_{k=1}^{K+1} \Theta_{jk} = \sum_{j=1}^{t_1} w_j \sum_{k=1}^{K+1} I(y_j)\theta_{jk} \tag{25}$$

6 Empirical application

We applied the method outlined above to the Luxembourg Income Study (LIS) data set,[3] focusing on net disposable income for the United States and Finland in the mid 1980s and in 2004. We chose the United States and Finland as they are two relevant examples of countries belonging to the group of Anglo-Saxon and Nordic

[3]Data are available from http://www.lisproject.org/. For a description of the Luxembourg Income Study, see Gornick and Smeeding [22]. All empirical results can be replicated downloading relevant files from http://fiorio.economia.unimi.it/ftp/proj/ineqdec/cowell_fiorio.zip. The main results are obtained using a modification of the Stata routine ineqrbd [19], which can also be downloaded from Stata typing "ssc install ineqrbd, replace" in the Stata command line.

countries, the first being characterised by higher inequality of after-tax income and a light welfare state, the second being characterised by relatively lower inequality and a substantial welfare state—see for example Brandolini and Smeeding [6, 7]. We focus on inequality computed for equivalised income, using the square-root equivalence scale, so that each individual is given his family's income normalised by the square root of the family size.

We use these data also because they allow us to compare the distribution of a uniformly defined income variable at approximately the same periods. In fact, four data sets are considered: the United States in 1986 and 2004 and Finland in 1987 and 2004. As Table 1 shows equivalised income inequality in mid 1980s Finland was between 42 and 69% smaller than that in the US, according to inequality measures the GE and Gini indices, and between 29 and 59% smaller, using quantile ratios. Nearly twenty years later, inequality of equivalised income increased in both countries, especially for incomes in the upper tail of the income distribution, as GE(2) shows. Although equivalised-income inequality increased relatively more in Finland, it remained consistently lower in Finland with respect to the US.

We begin by examining the role of two important subgroups, those defined by sex and by education of the household head, where education is coded into four categories (less than high school, high school, college and Master/PhD). One way to investigate these issues is a decomposition by population subgroups of GE indices. Table 2 presents results by education and by sex subgroups: it first gives the measures of inequality computed in each subgroup and then shows the within- and between-subgroup decomposition of inequality for the three GE indices, for United States and then Finland in each period. Given the exact decomposability property of GE indices, the sum of the within and between components is equal to total inequality. One might conclude from Table 2 that, decomposing by education, both the inequality within educational subgroups and the inequality between groups increased in each country. In particular, between-group inequality nearly doubled in both countries, while the trend of within-group inequality was more pronounced in Finland. By contrast, a decomposition by sex of the household head shows roughly the opposite pattern of within and between components: while the former clearly increased in both countries

Table 1 Inequality statistics

	Equivalised disposable income inequality							
	United States			Finland			Finland/US	
	1986	2004	Change (%)	1987	2004	Change (%)	1986–1987 (%)	2004 (%)
p90/p10	5.778	5.380	−7	2.375	2.775	17	−59	−48
p90/p50	2.076	2.080	0	1.482	1.636	10	−29	−21
p50/p10	2.786	2.584	−7	1.603	1.698	6	−42	−34
p75/p25	2.406	2.402	0	1.557	1.687	8	−35	−30
GE(0)	0.212	0.256	21	0.066	0.101	54	−69	−60
GE(1)	0.183	0.244	33	0.063	0.124	96	−65	−49
GE(2)	0.199	0.350	76	0.070	0.315	347	−65	−10
Gini	0.335	0.365	9	0.193	0.240	24	−42	−34

Note: p90 stands for the 90th percentile of the income distribution and similarly, p10, p50, p75 and p25

Table 2 Subgroup inequality decomposition by educational attainment and by sex of the householder

	United States						Finland					
	1986			2004			1987			2004		
	GE(0)	GE(1)	GE(2)	GE(0)	GE(1)	GE(2)	GE(0)	GE(1)	GE(2)	GE(0)	GE(1)	GE(2)
Subgroups by education												
< High school	0.222	0.203	0.230	0.223	0.210	0.308	0.062	0.059	0.061	0.092	0.099	0.131
High school	0.177	0.150	0.156	0.210	0.192	0.262	0.058	0.055	0.061	0.075	0.082	0.193
College	0.135	0.127	0.144	0.185	0.182	0.248	0.051	0.051	0.063	0.102	0.144	0.424
Master/PhD	0.144	0.122	0.124	0.217	0.222	0.306	0.045	0.046	0.048	0.085	0.094	0.121
Within	0.179	0.150	0.165	0.206	0.195	0.298	0.059	0.056	0.062	0.088	0.110	0.300
Between	0.033	0.033	0.034	0.050	0.050	0.052	0.007	0.007	0.008	0.013	0.014	0.014
Subgroups by sex												
Male	0.183	0.162	0.176	0.226	0.225	0.323	0.062	0.060	0.066	0.095	0.116	0.294
Female	0.270	0.246	0.290	0.283	0.263	0.377	0.078	0.079	0.093	0.112	0.141	0.369
Within	0.197	0.170	0.187	0.252	0.241	0.346	0.063	0.061	0.068	0.100	0.122	0.313
Between	0.015	0.013	0.012	0.004	0.003	0.003	0.003	0.003	0.002	0.002	0.002	0.002

Table 3 Factor source decomposition of the within-group component of inequality of equivalised income in the United States using a decomposition by educational attainment

	GE(0)		GE(1)		GE(2)		Factor source decomposition of within inequality (%)	
	1986	2004	1986	2004	1986	2004	1986	2004
Total inequality	0.212	0.256	0.183	0.244	0.199	0.350		
Between inequality	0.033	0.050	0.033	0.050	0.034	0.052		
Less than high school								
Number of earners	0.008	0.007	0.005	0.004	0.004	0.003	14.189	17.912
Num. children < 18	0.006	0.001	0.004	0.001	0.003	0.001	11.053	3.709
Housing rented	0.002	0.001	0.001	0.001	0.001	0.000	4.277	2.712
Age	0.004	0.001	0.003	0.000	0.002	0.000	7.364	1.443
Age squared	−0.002	0.000	−0.002	0.000	−0.001	0.000	−4.461	−0.791
Female	0.002	0.000	0.001	0.000	0.001	0.000	3.319	0.526
Residual	0.035	0.028	0.022	0.015	0.017	0.012	64.259	74.490
High school								
Number of earners	0.008	0.008	0.007	0.007	0.007	0.008	8.844	8.294
Num. children < 18	0.013	0.004	0.010	0.003	0.010	0.004	13.804	3.752
Housing rented	0.004	0.004	0.003	0.003	0.003	0.004	4.127	3.915
Age	0.011	0.006	0.009	0.005	0.009	0.006	12.356	6.051
Age squared	−0.008	−0.004	−0.006	−0.003	−0.006	−0.004	−8.718	−4.191
Female	0.003	0.000	0.003	0.000	0.003	0.000	3.611	0.299
Residual	0.060	0.082	0.049	0.066	0.049	0.079	65.977	81.878
College								
Number of earners	0.000	0.001	0.000	0.001	0.000	0.002	1.201	2.132
Num. children < 18	0.003	0.001	0.003	0.002	0.005	0.003	14.748	2.914
Housing rented	0.000	0.001	0.000	0.002	0.001	0.003	1.530	2.550
Age	0.003	0.003	0.004	0.003	0.005	0.006	15.369	5.668
Age squared	−0.002	−0.002	−0.003	−0.003	−0.004	−0.004	−13.267	−4.299
Female	0.000	0.000	0.000	0.000	0.000	0.000	1.339	0.207
Residual	0.015	0.044	0.018	0.054	0.027	0.091	79.080	90.828
Master/PhD								
Number of earners	0.001	0.000	0.001	0.000	0.001	0.001	3.539	0.770
Num. children < 18	0.002	0.000	0.003	0.001	0.005	0.001	14.908	1.595
Housing rented	0.001	0.000	0.001	0.001	0.001	0.001	3.717	1.498
Age	0.002	0.001	0.003	0.001	0.005	0.003	15.852	3.315
Age squared	−0.002	0.000	−0.002	−0.001	−0.004	−0.001	−12.578	−1.727
Female	0.000	0.000	0.000	0.000	0.001	0.000	2.212	0.269
Residual	0.011	0.019	0.014	0.033	0.022	0.080	72.350	94.280

the latter was roughly stable in absolute value in Finland and clearly decreasing in the United States.[4]

What emerges from this decomposition is that most of the inequality is due to the within component of inequality, but we do not know much about the role of other

[4] A careful analysis of these inequality statistics should also assess the magnitude of the sampling error [12], however in this paper we use the empirical application as an illustration of the methodologies presented in the previous sections. Further discussions about confidence intervals estimation of inequality measures and its decompositions will be presented in Section 7.

household characteristics. From this analysis one cannot disentangle the changed contribution of a demographic characteristic of the population (e.g., education) while controlling for the other (e.g., sex). A possible solution would be to create a finer partition of the sample by interacting education and sex, as proposed in Cowell and Jenkins [13]. However, this method could become cumbersome if one wanted to control for some additional characteristics (e.g., ethnicity, area of residence), would need a discretisation of variables which might reasonably be considered as continuous (e.g., age) and would reduce the sample size in each subgroup, hence the precision of the estimate.

Table 4 Factor source decomposition of the within-group component of inequality of equivalised income in Finland using a decomposition by educational attainment

	GE(0)		GE(1)		GE(2)		Factor source decomposition of within inequality (%)	
	1986	2004	1986	2004	1986	2004	1986	2004
Total inequality	0.066	0.101	0.063	0.124	0.070	0.315		
Between inequality	0.007	0.013	0.007	0.014	0.008	0.014		
Less than high school								
Number of earners	0.005	0.003	0.004	0.002	0.004	0.003	17.992	13.753
Num. children < 18	0.001	0.000	0.001	0.000	0.001	0.000	4.354	1.538
Housing rented	0.000	0.001	0.000	0.001	0.000	0.001	0.713	3.194
Age	−0.001	−0.001	−0.001	−0.001	−0.001	−0.001	−4.236	−4.186
Age squared	0.002	0.002	0.002	0.001	0.002	0.002	7.354	7.737
Female	0.001	0.000	0.001	0.000	0.001	0.000	2.420	1.595
Residual	0.020	0.015	0.017	0.013	0.016	0.015	71.403	76.369
High school								
Number of earners	0.002	0.001	0.002	0.001	0.002	0.003	8.557	4.382
Num. children < 18	0.001	0.001	0.001	0.001	0.002	0.001	5.873	1.979
Housing rented	0.001	0.001	0.001	0.001	0.001	0.001	3.008	1.907
Age	0.003	0.002	0.003	0.002	0.003	0.003	10.671	4.725
Age squared	−0.002	−0.001	−0.002	−0.001	−0.002	−0.002	−7.746	−3.403
Female	0.001	0.000	0.001	0.000	0.001	0.001	2.267	1.206
Residual	0.019	0.029	0.018	0.029	0.020	0.061	77.369	89.205
College								
Number of earners	0.000	0.000	0.000	0.000	0.000	0.002	0.957	0.844
Num. children < 18	0.000	0.001	0.000	0.001	0.001	0.005	5.871	2.364
Housing rented	0.000	0.000	0.000	0.000	0.000	0.001	2.132	0.393
Age	0.001	0.000	0.001	0.001	0.002	0.002	14.348	0.889
Age squared	−0.001	0.000	−0.001	0.000	−0.001	−0.001	−10.601	−0.667
Female	0.000	0.000	0.000	0.000	0.000	0.001	1.511	0.286
Residual	0.005	0.033	0.007	0.056	0.011	0.200	85.783	95.892
Master/PhD								
Number of earners	0.000	0.000	0.000	0.000	0.000	0.000	4.909	0.919
Num. children < 18	0.000	0.000	0.000	0.000	0.000	0.000	17.572	5.114
Housing rented	0.000	0.000	0.000	0.000	0.000	0.000	0.446	1.534
Age	0.000	0.000	0.000	0.000	0.000	−0.001	8.858	−12.044
Age squared	0.000	0.000	0.000	0.000	0.000	0.001	−1.908	19.319
Female	0.000	0.000	0.000	0.000	0.000	0.000	0.251	0.254
Residual	0.000	0.001	0.000	0.002	0.001	0.004	69.873	84.904

Table 5 Factor source decomposition of thewithin-group component of inequality of equivalised income in the United States using a decomposition by gender

	GE(0)		GE(1)		GE(2)		Factor source decomposition of within inequality (%)	
	1986	2004	1986	2004	1986	2004	1986	2004
Total inequality	0.212	0.256	0.183	0.244	0.199	0.350		
Between inequality	0.015	0.004	0.013	0.003	0.012	0.003		
Male								
Num. of earners	0.009	0.004	0.009	0.004	0.010	0.006	5.834	3.136
Num. < 18	0.019	0.003	0.018	0.004	0.021	0.006	12.273	2.739
Housing rented	0.005	0.003	0.004	0.003	0.005	0.005	2.971	2.248
Age	0.014	0.005	0.013	0.005	0.015	0.008	8.937	4.160
Age squared	−0.010	−0.003	−0.009	−0.004	−0.011	−0.006	−6.231	−2.810
High school	−0.002	−0.002	−0.002	−0.002	−0.002	−0.004	−1.426	−1.883
College	0.007	0.005	0.006	0.005	0.007	0.008	4.372	3.960
Master/PhD	0.016	0.013	0.015	0.014	0.017	0.021	10.263	10.659
Residual	0.098	0.095	0.092	0.101	0.107	0.157	63.008	77.791
Female								
Num. of earners	0.005	0.008	0.003	0.007	0.002	0.009	11.725	6.019
Num. < 18	0.004	0.003	0.002	0.003	0.002	0.004	10.337	2.671
Housing rented	0.001	0.006	0.001	0.005	0.001	0.006	3.405	4.313
Age	0.001	0.005	0.001	0.004	0.000	0.006	2.800	3.814
Age squared	0.000	−0.003	0.000	−0.002	0.000	−0.003	−0.762	−2.250
High school	0.000	−0.002	0.000	−0.001	0.000	−0.002	0.610	−1.288
College	0.002	0.007	0.001	0.006	0.001	0.007	5.951	5.104
Master/PhD	0.002	0.009	0.001	0.008	0.001	0.010	4.797	7.045
Residual	0.025	0.097	0.015	0.083	0.011	0.108	61.138	74.572

What additional insights might a regression-based approach yield? By applying a regression-based factor-source decomposition as discussed in Section 5, we can assess the contribution of (the total value of) each right-hand-side variable to inequality. Our factor-source decomposition of within-group inequality allows us to assess whether one variable contributes uniformly to inequality in each subgroup or has a disproportionate effect across the subgroups. We estimate separate regressions for each subgroup as in Eq. 20 where y_j is the vector of household equivalised incomes of households in group j and as covariates we used, for both countries in both periods, family variables (number of earners, number of children under age 18, whether the family rents or owns its own dwelling) and variables referring to the household head only (age, age squared, sex and four category dummies for education).[5] Clearly this is not a structural model and its specification is unsuitable for a causal interpretation. We deliberately adopted a parsimonious specification, but it is informative about the correlation of some key variables with equivalised household income.

[5]This is a clearly simplified model of equivalised income generation, however available data would not allow the development of a more complex structural model of household income. For further discussion of this issue, see Section 7.

Table 6 Factor source decomposition of the within-group component of inequality of equivalised income in Finland using a decomposition by gender

	GE(0)		GE(1)		GE(2)		Factor source decomposition of within inequality (%)	
	1986	2004	1986	2004	1986	2004	1986	2004
Total inequality	0.066	0.101	0.063	0.124	0.070	0.315		
Between inequality	0.003	0.002	0.003	0.002	0.002	0.002		
Male								
Num. of earners	0.003	0.002	0.004	0.002	0.004	0.005	8.005	2.372
Num. < 18	0.002	0.001	0.003	0.002	0.003	0.005	5.050	2.108
Housing rented	0.001	0.001	0.001	0.001	0.001	0.003	1.368	1.195
Age	0.001	0.002	0.002	0.002	0.002	0.006	2.996	2.466
Age squared	−0.001	−0.001	−0.001	−0.001	−0.001	−0.003	−1.183	−1.439
High school	0.000	0.000	0.000	0.000	0.000	−0.001	−0.461	−0.283
College	0.005	0.002	0.006	0.003	0.005	0.007	10.458	3.141
Master/PhD	0.001	0.001	0.001	0.001	0.001	0.002	1.418	0.875
Residual	0.032	0.060	0.040	0.076	0.037	0.201	72.349	89.566
Female								
Num. of earners	0.003	0.001	0.001	0.001	0.003	0.003	14.199	3.358
Num. < 18	0.000	0.000	0.000	0.000	0.000	0.001	−0.111	0.729
Housing rented	0.001	0.000	0.000	0.000	0.001	0.001	2.602	0.622
Age	−0.002	0.000	0.000	0.000	−0.002	0.000	−7.151	−0.345
Age squared	0.003	0.000	0.001	0.000	0.003	0.001	13.315	0.845
High school	0.000	0.000	0.000	0.000	0.000	0.000	−0.043	−0.011
College	0.002	0.001	0.001	0.001	0.002	0.003	9.694	3.809
Master/PhD	0.000	0.000	0.000	0.000	0.000	0.001	0.253	1.247
Residual	0.015	0.029	0.004	0.033	0.015	0.079	67.243	89.745

Inequality decomposition estimates are presented for education subgroups in Tables 3 and 4, and for gender subgroups in Tables 5 and 6.[6] All these tables have the same structure: the first line reports the total inequality using GE(0), GE(1) and GE(2) for each of the two years considered and the second line reports the between inequality. In the following lines a decomposition of within-group inequality is provided, accounting for the contribution of each covariate in each subgroup to within-group inequality. The contribution of each covariate in each subgroup is obtained as in Eq. 25, by multiplying the factor-source decomposition of inequality in each group ($I(y_j)$) by its weight (w_j). The factor-source decomposition of the inequality in each subgroup is reported in percentage terms in the last two column for each of the years considered. As this inequality decomposition enjoys the same properties as the factor-source decomposition suggested in Shorrocks [34], namely the fact that it is invariant to the inequality measure used, we used these factors to decompose the within components of the GE(0), GE(1) and GE(2).

Table 3 shows that in the US female-headed households and households with young children accounted for a decreasing share of within-group inequality, while the number of earners in the household accounted for a relatively stable share of

[6]Tables of results are presented omitting the OLS coefficient estimates and their significance, which could however be obtained from the authors upon request.

within-group inequality. This decomposition shows that the largest contribution to within-group inequality is due to the number of earners and the number of children younger than 18 and that the rented household accounts for a relatively large share of inequality in the high school educated household, while it is less important in the less and the most educated households. In Finland the number of young children is much less relevant to account for within-group inequality except for the group of college educated households, possibly due to a larger welfare system. The negligible contribution to within-group inequality of the most educated group reflects the relatively small share of population in this groups (less than 1.3% in 2004) and shows that within-group inequality is mostly due to the group of high school or less educated households (Table 4).

Looking at gender subgroups, Table 5 shows that the large increase of within-group inequality in the US as measured by the GE(2) index between the two years considered is accounted for by the female subgroup and in particular by the number of earners, the number of young children and by high level of education. This trend is instead much less evident in Finland (Table 6).

Finally, it should be pointed out that the proposed inequality decomposition is exact only if the contribution of the residual is not ignored. Indeed, Tables 3–6 show that, after controlling for a set of individual and family characteristics, the residual within each subgroup still accounts for a proportion between 61% and 94% of total inequality within subgroups and that the residual accounts for an increasing share of within-group inequality over time. This suggests that a simple linear model such as the one we have suggested for illustrative purposes should be enriched either by including more controls, when available, or by specifying a richer model.

7 Discussion

Clearly any empirical methodology should come with a set of warnings about implementation: so too with the techniques illustrated in Section 6.

First, although the computation of standard errors is sometimes treated as a trivial problem (as in Morduch and Sicular [27]), this is not so; the main reason for the complexity is that the inequality index computed from a random sample is itself a random variable and cannot be treated as deterministic in the calculation of standard errors (see Section 4); moreover, $I(y)$ often appears at the denominator of these decompositions making theoretical computation of standard errors cumbersome. A viable way to assess the robustness of estimates is to provide different specifications of the regression models, assessing the effects of the inclusion or exclusions of some independent variables and the significance of results could be assessed by computing standard errors using the bootstrap.

Second, a single-equation model, such as that developed above, should only be interpreted as a descriptive model, showing correlations rather than causal relationships. Could we have done better by opting for a richer model such as the Bourguignon et al. [4, 5] simultaneous-equation extension of the Blinder-Oaxaca decomposition? Their interest is in the change across time of the full distribution of income and related statistics. The components of their model are an earnings equation for each household member (linking individual characteristics to their remuneration), a labour supply equation (modelling the decision of the individual and of other household's members) and a household income equation (aggregating the individuals' contributions to household income formation). The estimation of such an econometric model at two different dates allows one to disentangle: (i) a "price effect" (people with given characteristics and the same occupation get a different income because the remuneration structure has changed) (ii) a "participation" or "occupation effect" (individuals with given characteristics do not make the same choices as for entering the labour force because their household may have changed) and (iii) a "population effect" (individual and household incomes change because socio-demographic characteristics of population of households and individuals change). The main merit of such an approach is that it builds a comprehensive model of how decisions regarding income formation are taken, including the individual decision of entering the labour force and wage formation mechanism, into a household-based decision process, extracting part of the information left in the residuals of single-equation linear models as the one used in this paper. Bourguignon et al. [5] used this methodology to argue persuasively that the apparent stability of Taiwan's income inequality was just due to the offsetting of different forces. However, the rich structural model comes at the expense of increasing the complication of the estimation process and of introducing additional and perhaps questionable assumptions. Among the most important limitations of the Bourguignon et al. approach are: the robustness of the estimates of some coefficients, the problem of simultaneity between household members' labour-supply decisions, the issue of understanding what is left in the residuals of the labour supply equations and the counterfactual wage equations, the path-dependence problem (i.e., which counterfactual is computed first) is also a problem.[7] In sum, the full structural model approach for inequality analysis can be cumbersome and is likely to be sensitive to model specification.

8 Conclusion

At the beginning we raised the question of whether the main approaches to inequality decomposition were on speaking terms. The *a priori* approach and the regression-model approach outlined in Section 2 might appear at first glance to be incompatible. However, they can be made to "talk to each other." The key to the translation lies in an appropriate application and interpretation of the factor-source decomposition method. Our approach to reconciling the different strands of inequality-decomposition analysis is based on a single-equation regression, builds on the Shorrocks [34] methodology and is aimed at providing a tool for understanding inequality, especially when the data are not sufficiently detailed to allow a structural model specification. It shares some features with the approach suggested by Fields [17],[8] but improves on it by including in the analysis the decomposition by subgroups and in showing how this might also be useful to identify differences in determinants of inequality.

[7] To get some idea of the magnitude of the path-dependence problem the authors computed all possible evaluations of price, participation and population effects, although the complex problem of computing proper confidence intervals for the structural model is not tackled. The problem has something in common with that of the Shapley-value method discussed in Section 2.1.

[8] See also Fields and Yoo [18] and Morduch and Sicular [27].

Our approach is fairly robust, providing an improvement on other methods; it also provides results consistent with other decomposition methods. The simple specification makes no claims about causality but enables one to distinguish clearly between methods of accounting for inequality that rely solely on a breakdown of the factors that underlie predicted income and the breakdown of inequality of observed income.

Acknowledgements We are very grateful for comments from conference and seminar participants at the Brucchi Luchino workshop, CEPS, Cornell, LAGV Marseille, Leuven and the EEA Barcelona. We also want to thank the referees of this Journal for their constructive comments.

References

1. Blackorby, C., Donaldson, D., Auersperg, M.: A new procedure for the measurement of inequality within and among population subgroups. Can. J. Econ. **14**, 665–685 (1981)
2. Blinder, A.S.: Wage discrimination: reduced form and structural estimates. J. Hum. Resour. **8**, 436–455 (1973)
3. Bourguignon, F.: Decomposable income inequality measures. Econometrica **47**, 901–920 (1979)
4. Bourguignon, F., Ferreira, F.H.G., Leite, P.G.: Beyond Oaxaca-Blinder: accounting for differences in household income distributions. J. Econ. Inequal. **6**, 117–148 (2008)
5. Bourguignon, F., Fournier, M., Gurgand, M.: Fast development with a stable income distribution: Taiwan, 1979–94. Rev. Income Wealth **47**, 139–163 (2001)
6. Brandolini, A., Smeeding, T.M.: Income inequality in richer and OECD countries. In: Salverda, W., Nolan, N., Smeeding, T.M. (eds.) Oxford Handbook on Economic Inequality, chapter 4. Oxford University Press, Oxford (2008)
7. Brandolini, A., Smeeding, T.M.: Inequality patterns in western-type democracies: cross-country differences and time changes. In: Democracy, I., Representation, P.B., Anderson, R.S.F.C.J. (eds.) Democracy, Inequality and Representation. Russell Sage Foundation, New York (2008)
8. Chantreuil, F., Trannoy, A.: Inequality decomposition values: the trade-off between marginality and consistency. Working Papers 99-24, THEMA, Université de Cergy-Pontoise (1999)
9. Charpentier, A., Mussard, S.: Income inequality games. J. Econ. Inequal. (2011, to appear)
10. Cowell, F.A.: On the structure of additive inequality measures. Rev. Econ. Stud. **47**, 521–531 (1980)
11. Cowell, F.A.: Inequality decomposition—three bad measures. Bull. Econ. Res. **40**, 309–312 (1988)
12. Cowell, F.A.: Sampling variance and decomposable inequality measures. J. Econom. **42**, 27–41 (1989)
13. Cowell, F.A., Jenkins, S.P.: How much inequality can we explain? A methodology and an application to the USA. Econ. J. **105**, 421–430 (1995)
14. DiNardo, J., Fortin, N.M., Lemieux, T.: Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. Econometrica **64**, 1001–1044 (1996)
15. Ebert, U.: On the decomposition of inequality: Partitions into nonoverlapping sub-groups. In: Eichhorn, W. (ed.) Measurement in Economics. Physica Verlag, Heidelberg (1988)
16. Elbers, C., Lanjouw, P., Mistiaen, J.A., Özler, B.: Reinterpreting between-group inequality. J. Econ. Inequal. **6**, 231–245 (2008)
17. Fields, G.S.: Accounting for income inequality and its change: a new method with application to distribution of earnings in the United States. Res. Labor Econ. **22**, 1–38 (2003)
18. Fields, G.S., Yoo, G.: Falling labor income inequality in Korea's economic growth: patterns and underlying causes. Rev. Income Wealth **46**, 139–159 (2000)
19. Fiorio, C., Jenkins, S.P.: Regression-based inequality decomposition, following Fields (2003). UK Stata User Group meeting, 10 September 2007
20. Foster, J.E., Shneyerov, A.A.: A general class of additively decomposable inequality measures. Econ. Theory **14**, 89–111 (1999)
21. Foster, J.E., Shneyerov, A.A.: Path independent inequality measures. J. Econ. Theory **91**, 199–222 (2000)

22. Gornick, J.C., Smeeding, T.M.: The Luxembourg income study. In: Darity, W.A.J. (ed.) International Encyclopedia of the Social Sciences, 2nd edn., pp. 419–422. Macmillan, Detroit (2008)
23. Jenkins, S.P.: Accounting for inequality trends: Decomposition analyses for the UK. Economica **62**, 29–64 (1995)
24. Kanbur, S.M.N.: The policy significance of decompositions. J. Econ. Inequal. **4**, 367–374 (2006)
25. Lasso de la Vega, C., Urrutia, A.: Path independent multiplicatively decomposable inequality measures. Investig. Econ. **29**, 379–387 (2005)
26. Lasso de la Vega, C., Urrutia, A.: The extended Atkinson family: the class of multiplicatively decomposable inequality measures, and some new graphical procedures for analysts. J. Econ. Inequal. **6**, 211–225 (2008)
27. Morduch, J., Sicular, T.: Rethinking inequality decomposition, with evidence from rural China. Econ. J. **112**, 93–106 (2002)
28. Olkin, I., Yitzhaki, S.: Gini regression analysis. Int. Stat. Rev. **60**, 185–196 (1992)
29. Paul, S.: Income sources effects on inequality. J. Dev. Econ. **73**, 435–451 (2004)
30. Sala-i-Martin, X.: The world distribution of income: falling poverty and ... convergence, period. Q. J. Econ. **121**, 351–397 (2006)
31. Sastre, M., Trannoy, A.: Changing income inequality in advanced countries: a nested marginalist decomposition analysis. mimeo (2000)
32. Sastre, M., Trannoy, A.: Shapley inequality decomposition by factor components: some methodological issues. J. Econom. **Supplement 9**, 51–90 (2002)
33. Shorrocks, A.F.: The class of additively decomposable inequality measures. Econometrica **48**, 613–625 (1980)
34. Shorrocks, A.F.: Inequality decomposition by factor components. Econometrica **50**(1), 193–211 (1982)
35. Shorrocks, A.F.: The impact of income components on the distribution of family income. Q. J. Econ. **98**, 311–326 (1983)
36. Shorrocks, A.F.: Inequality decomposition by population subgroups. Econometrica **52**, 1369–1385 (1984)
37. Shorrocks, A.F.: Aggregation issues in inequality measurement. In: Eichhorn, W. (ed.) Measurement in Economics. Physica Verlag, Heidelberg (1988)
38. Shorrocks, A.F.: Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value. mimeo, Department of Economics, University of Essex (1999)
39. Theil, H.: The measurement of inequality by components of income. Econ. Lett. **2**, 197–199 (1979)
40. Yu, L., Luo, R., Zhan, L.: Decomposing income inequality and policy implications in rural China. China World Econ. **15**, 44–58 (2007)