Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

GRINS FOUNDATION

Discussion Paper Series

Within-group inequality: a comparison of different definitions and a new proposal of decomposition

Discussion paper n. 12/2025

Marc Fleurbaey (1), Peter Lambert Domenico Moramarco (2) and Vito Peragine (3); (1) Paris School of Economics, France, (2,3) University of Bari, Italy.

ISSN 3035-5567

Within-group inequality: a comparison of different definitions and a new proposal of decomposition

Marc Fleurbaey (1), Peter Lambert Domenico Moramarco (2) and Vito Peragine (3); (1) Paris School of Economics, France, (2,3) University of Bari, Italy.

Discussion Paper DP 12/2025

Keywords: inequality, Gini coefficient, decomposition, within-group inequality.

JEL: D63.

In this paper, we compare two alternative procedures for identifying the within-group component of total inequality while decomposing total inequality into a between and a within-group term: the standard additive decomposition, which defines within-group inequality as the weighted sum of the inequality in each group, and the path-independent decomposition, which redefines the concept of within-group inequality by looking at a standardized distribution where groups have the same average income and do not overlap. We show that a decomposition of total inequality based on the former approach others a clean measure of the between-group inequality, which is insensitive to changes in the distribution that do not alter the relative difference between the groups' averages. On the other hand, a decomposition based on the latter approach others an unbiased measure of the within-group inequality, which is independent of the between-group component. Hence, we propose a new decomposition of the Gini index that combines the definition of between-group inequality stemming from the additive decomposition with the measure of within-group inequality at the base of the path-independent decomposition. Finally, we turn to the

decomposition of the partial Lorenz ordering by exploring the implications of Lorenz between-group dominance combined with Lorenz within-group dominance.

We also show the difficulty of defining sufficient conditions for two Lorenz curves not to intersect and suggest an alternative partial order based on concentration curves which do not account for the overlap between groups.

This paper was presented during the Thenth Meeting of the Society for the Study of Economic Inequality (ECINEQ) in Aix-en-Provence; we thank all participants for their comments and suggestions.
Domenico Moramarco gratefully acknowledges ECARES and FNRS for the financial support during the earlier stages of this project. He also acknowledges the GRINS (Growing Resilient, INclusive and Sustainable) foundation; he received funding from the European Union Next-GenerationEU (NRRP - MISSION 4, COMPONENT 2, INVESTMENT 1.3 - D.D. 1558 11/10/2022, PE00000018, Spoke 3)

Within-group inequality: a comparison of different definitions and a new proposal of decomposition*

Marc Fleurbaey†, Peter Lambert Domenico Moramarco‡and Vito Peragine§

September 2, 2024

Abstract

In this paper, we compare two alternative procedures for identifying the within-group component of total inequality while decomposing total inequality into a between and a within-group term: the standard additive decomposition, which defines within-group inequality as the weighted sum of the inequality in each group, and the path-independent decomposition, which redefines the concept of within-group inequality by looking at a standardized distribution where groups have the same average income and do not overlap. We show that a decomposition of total inequality based on the former approach offers a clean measure of the between-group inequality, which is insensitive to changes in the distribution that do not alter the relative difference between the groups' averages. On the other hand, a decomposition based on the latter approach offers an unbiased measure of the within-group inequality, which is independent of the between-group component. Hence, we propose a new decomposition of the Gini index that combines the definition of between-group inequality stemming from the additive decomposition with the measure of within-group inequality at the base of the path-independent decomposition. Finally, we turn to the decomposition of the partial Lorenz ordering by exploring the implications of Lorenz between-group dominance combined with Lorenz within-group dominance. We also show the difficulty of defining sufficient conditions for two Lorenz curves not

*This paper was presented during the Thenth Meeting of the Society for the Study of Economic Inequality (ECINEQ) in Aix-en-Provence; we thank all participants for their comments and suggestions. Domenico Moramarco gratefully acknowledges ECARES and FNRS for the financial support during the earlier stages of this project. He also acknowledges the GRINS (Growing Resilient, INclusive and Sustainable) foundation; he received funding from the European Union Next-GenerationEU (NRRP - MISSION 4, COMPONENT 2, INVESTMENT 1.3 - D.D. 1558 11/10/2022, PE00000018, Spoke 3).

†Paris School of Economics, France. Email: marc.fleurbaey@psemail.eu

‡University of Bari, Italy. Corresponding author. Email: domenico.moramarco@uniba.it.

§University of Bari, Italy. Email: vitorocco.peragine@uniba.it

to intersect and suggest an alternative partial order based on concentration curves which do not account for the overlap between groups.

Keywords: inequality, Gini coefficient, decomposition, within-group inequality.

JEL: D63.

1 Introduction

Many analyses of inequality are concerned with the evolution of inequality over time and whether changes are due to altered inequality within subgroups, demographic shifts, or changes in the level of between-group inequality, somehow defined.

Empirical applications are often focused on determining the share of overall income inequality that can be attributed to differences between population subgroups (defined by age, race, gender, or some other meaningful characteristic). On the other hand, this is not independent from the way one defines and measures inequality within groups.

One can cite the ratio of means and ratio of medians as simple measures of group differences (see Kestenbaum, 1979). There is also a literature on indices of dissimilarity – starting with Duncan & Duncan (1955)'s measure of segregation, in respect of income, education or occupation (as between men and women, or black and white, for example). For the study of occupational discrimination by gender and by race, both the differences in the distributions of the respective groups and the differences in mean earnings between the respective groups are relevant (see Wolff, 1976, p. 152). Disparity curves are generally used for the comparison of distributions (Handcock & Morris, 2006), and in particular for the measurement of discrimination patterns, which occur when one group is systematically at an advantage over another for no ethically acceptable reason - see Le Breton *et al.* (2012) where there is a link with the class of discrimination indices developed by Gastwirth (1975). Social concern is often expressed over demographic disparities between the rates of occurrence of an adverse economic outcome, such as pollution, between the advantaged and the disadvantaged (see Lambert & Subramanian, 2014). The literature on inequality of opportunity has developed a framework in which the inequality between types – which are groups defined on the basis of exogenous circumstances – is interpreted as inequality of opportunity, opposed to inequality due to individual effort (see Checchi & Peragine, 2010). It has also been suggested to base an index of polarization of the income distribution across groups on a comparison of the (assumedly comparable) within-group inequalities with the (typically much larger) between group inequality. Fedorov (2002, p. 499) finds that trends in interregional inequality and polarization were *remarkably similar*. Zhang & Kanbur (2001) propose to measure polarization generally by the ratio of between-group inequality to within-group inequality, but using a decomposable index. Two Gini-based versions of Zhang and Kanbur's measure, one embodying overlap and the other not, are delineated in Lambert & Decoster (2005).

When distinguishing within and between group inequality, it is often desirable to use indices that are decomposable into the sum of these between-group and within-group terms. The interest for decomposability has led many scholars to prefer generalized entropy measures to the Gini or the Atkinson indices of inequality. Despite the tight link with the decomposability of inequality measures, distinguishing within group from between group inequality is a distinct issue. This is the focus of the present paper: we investigate different procedures to identify the between and within-group components of inequality, discuss their relative merits, and propose our preferred definitions,

To address this issue, we base the analysis on the Lorenz curve, which remains the undisputed graphical tool to analyse inequality, and the Gini coefficient, given its tight, and well known, link with the Lorenz curve.

Between group inequality has, since long, been measured as inequality between groups' representative incomes. In particular, while earlier contributions (Bhattacharya & Mahalanobis, 1967; Aronson & Lambert, 1994) favor the use of the arithmetic mean, others (Foster & Shneyerov, 2000; Blackorby *et al.*, 1981) suggest using generalized means which account for within group inequality in the definition of the representative income.

The agreement among economics scholars is less clear when we consider within group inequality, for which two main alternatives have been proposed. The first one follows from the *additive decomposition* of the Gini coefficient first explored by Bhattacharya & Mahalanobis (1967). This is a decomposition of the Gini coefficient into: (i) a weighted sum of subgroup inequality measures, for weights that depend on the population and the income share of the groups, (ii) a between-group measure, capturing the inequality between groups' averages and (iii) a residual term, measuring the overlap between groups. Accordingly, within group inequality is defined as the weighted sum of the inequality within each group, while the between group component corresponds to the inequality in the *smooth* distribution where each income is replaced by the group's arithmetic mean.

This definition has stood alone for many years, but has been challenged by Foster & Shneyerov (2000) who suggest a different smoothing process: one which would remove inequality within subgroups by replacing each person's income by the so-called representative income of his or her subgroup (a generalized mean, in fact). Within-group inequality is defined in such a measurement system as the inequality level of the standardized distribution, which re-scales each group distribution so that the representative incomes of all groups equal the overall representative income, leaving group inequality levels unaltered. Foster & Shneyerov (2000), suggest a *path independent decomposition* of the overall inequality into the sum of these between-group and within-group terms, and proceed with defining a family of indices that satisfy this path independent decomposition. The only representative income functions that are consistent with path independent measures are the generalized means; in this paper, to allow for comparison between the two approaches, we focus on the arithmetic mean as representative income function.

This way of interpreting within-group inequality as the inequality in the standardized income distribution challenges the conventional conceptualization of within group inequality as the weighted sum of the inequalities within each group, for some weighting functions depending on the mean and/or the population size of the different groups. The concept of additive decomposability rests on such definition of within groups inequality (see Shorrocks, 1980). However, the problem with additive decomposable indices is that, unless the weighting functions are independent of the mean incomes, the between group and the within group terms are not independent. In fact, consider a proportional positive income transfer from group i to group j: this should affect the inequality between, but not the inequality within. But this is true only if the weighting functions are independent of the means. On the other hand, path independence is exactly a requirement of independence between the two terms of within and between groups inequality. Hence, path independence is stronger than additive decomposability.

In this paper, we start from the decomposing the Lorenz curve in a way that highlights the two alternative definitions of within group inequality. We derive the corresponding decompositions of the Gini coefficient and highlight that there is indeed a key difference between the two approaches, which has to do with how the between and within components of inequality interact with each other. As also mentioned before, with an additive decomposition of the Gini coefficient, we obtain that the within group component is sensitive to the inequality between groups. Conversely, if the within group inequality is measured as in the path independent decomposition, the resulting between group component becomes sensitive to the within group inequality. This suggests that, at least in the context of the Gini coefficient, researchers should choose with care how to decompose inequality: a study focusing on inequality within groups should decompose the Gini coefficient using the path independent procedure; conversely, if the focus is on inequality between groups, then the additive decomposition would deliver more consistent measures.

Given the limitations of the two procedures, we suggest a new decomposition of the Gini coefficient into a between group component that corresponds to the between group inequality in the additive decomposition approach, and a within group inequality component stemming from the path independent decomposition. The new decompositon generates a different residual term, for which we are able to derive an estimator that depends only on the total Gini coefficient, the Gini coefficient in each group, and population and income shares. Our proposal nicely follows the decomposition of the Lorenz curve, from which we identify a clear graphical representation of the (sum of the) between and within group inequality: the concentration curve. Finally, we turn to the decomposition of the partial Lorenz ordering and explore the implications of Lorenz between-group dominance combined with Lorenz within group dominance. We show the difficulty of defining sufficient conditions for two Lorenz curves not to intersect and suggest an alternative partial order based on concentration curves which do not account for the overlap between groups.

The structure of the remainder of the paper is as follows. We start in Section 2 with a decomposition of the the Lorenz curve and derive the alternative decompositions of the Gini coefficient. After highlighting the differences between the two approaches, in Section 3 we show their implications via some illustrative simulations and propose a new decomposition of the Gini coefficient. Section contains a brief discussion and some results on the implication of the two approaches with respect to the decomposition of the Lorenz partial ordering. Section 5 provides some concluding remarks. While in the body of the paper we consider the case of two groups in detail, in the appendix, we sketch some corresponding results for the many group case.

2 Decomposing the Gini index

Consider a distribution of income in a given population represented by an increasingly ordered vector $x \in \mathbb{R}^n_+$ with mean μ. The Lorenz curve of this distribution is the graph of the function

$$L_x(i/n) = \frac{1}{n\mu}\sum_{j=1}^{i} x_j \qquad i = \{1, ..., n\}$$

and the area between the first diagonal and L_x corresponds to the well known Gini coefficient:

$$G = 1 - 2\int_0^1 L_x(t)\ dt$$

Suppose that A and B are two subgroups which make up the population of income recipients. Let the subgroup means be μ_A and μ_B and let the subgroup Gini coefficients be G_A and G_B. Suppose there are n_A income units in subgroup A and n_B income units in subgroup B.

Let $p_j = \frac{n_j}{n}$ be the share of individuals in subgroup $j \in \{A, B\}$ and $q_j = p_j\frac{\mu_j}{\mu}$ denote their relative total income. We assume, without loss of generality, that $\mu_A \geq \mu_B$, so that $q_A \geq p_A$ (if not, we simply relabel the groups). We call *lexicographic parade* the list of individuals obtained by increasingly ordering the income units in B, followed by the increasingly ordered income units in A.

The prominent decompositon approaches in the economic literature - additive and path independent - are based on the idea that between groups inequality should be measured with reference to representative incomes. In the additive decompositon, the representative income of, say, group B is μ_B. In the path independent approach, the representative income of the same group must be a generalize mean of order ρ: $m_B^\rho = \left(n_B^{-1}\sum_{i\in B} x_i^\rho\right)^{1/\rho}$ with $\rho \leq 1$. For reasons of comparability, in this paper we focus on the path independent decomposition with $\rho = 1$. The reference to representative incomes has the great advantage of simplifying the conceptualization of between groups inequality in economic settings. Because of its relevance in the economic literature, it is the focus of the following subsection, and of this paper.

2.1 Representative income decompositons

Following Bhattacharya & Mahalanobis (1967), between group inequality is the inequality in a distribution where each individual income is replaced by his group's mean. We define the between-group Lorenz curve L_{BET} as follows.

$$L_{BET}(r) = \begin{cases} \frac{q_B}{p_B} r & 0 \leq r \leq p_B \\ q_B + \frac{q_A}{p_A}(r - p_B) & p_B \leq r \leq 1 \end{cases} \tag{1}$$

where $r = \frac{i}{n}$ and i is the rank of an income unit in lexicographic income parade. By construction, the income unit r belongs to B if $0 \leq r \leq p_B$ and belongs to A if $p_B \leq r \leq 1$. (See Figure 1(a))

Denote by $\tilde{x}$ the lexicographic parade of x. Formally, $\tilde{x} = \left(x_1^B, ..., x_{n_B}^B, x_1^A, ..., x_{n_A}^A\right)$ with $x_j^T \leq x_{j+1}^T$ for all $j \in \{1, ..., n_T - 1\}$, $T \in \{A, B\}$. Its concentration curve $C_{\tilde{x}}$ is shown in Figure 1(a) and can be defined as

$$C_{\tilde{x}}(r) = \begin{cases} q_B L_x^B\left(\frac{r}{p_B}\right) & 0 \leq r \leq p_B \\ q_B + q_A L_x^A\left(\frac{r-p_B}{p_B}\right) & p_B \leq r \leq 1 \end{cases} \tag{2}$$

where L_x^A and L_x^B are the Lorenz curves of the subgroup distributions.

Pictorially, $C_{\tilde{x}}$ is obtained by drawing the Lorenz curve of B, first, and A treating the piece-wise linear curve L_{BET} as substitute for the 45 degree line. Consequently G_B and G_A can be measured via the area between the L_{BET} and each piece of $C_{\tilde{x}}$. The Lorenz curve of the overall distribution is denoted L_x and represented by the dashed curve in Figure 1(a). As we can see, the area between the first diagonal and L_x is not entirely covered by L_{BET} and $C_{\tilde{x}}$. The area between $C_{\tilde{x}}$ and L_x represents the residual of the Gini decomposition.

Let x_μ denote the the distribution where everyone has income equal to the average. Following Aronson & Lambert (1994), the additive decomposition of the Lorenz curve can be described by the following steps.

1. $x_\mu \to (\mu_B \mathbf{1}_{n_B}, \mu_A \mathbf{1}_{n_A})$; $L_{x_\mu} \to L_{BET}$: from the equal distribution to the smoothed distribution with only between group inequality;

2. $(\mu_B \mathbf{1}_{n_B}, \mu_A \mathbf{1}_{n_A}) \to \tilde{x}$; $L_{BET} \to C_{\tilde{x}}$: from the smoothed distribution to a distribution in which within group inequality is introduced, but incomes are ordered lexicographically (first by group, then by increasing order within each group);

3. $\tilde{x} \to x$; $C_{\tilde{x}} \to L_x$: from the previous distribution to the final - original - distribution.

Consequently, the additive decomposition of the Gini coefficient is $G = G_{BET}^{Add} + G_{WIT}^{Add} + G_{RES}$, where G is the Gini coefficient, G_{BET}^{Add} and G_{WIT}^{Add} are respectively the the between and the within groups components of the Gini coefficient, and G_{RES} measures groups overlap (see Lambert & Aronson, 1993, for full details).

Formally, G_{BET}^{Add} is twice the area between L_{BET} and the 45 degree line:

$$G_{BET}^{Add} = 1 - 2\int_0^1 L_{BET}(r)dr \ = 1 - 2\left[\frac{p_B q_B}{2} + p_A q_B + \frac{p_A q_A}{2}\right]$$

that is

$$G_{BET}^{Add} = q_A - p_A \tag{3}$$

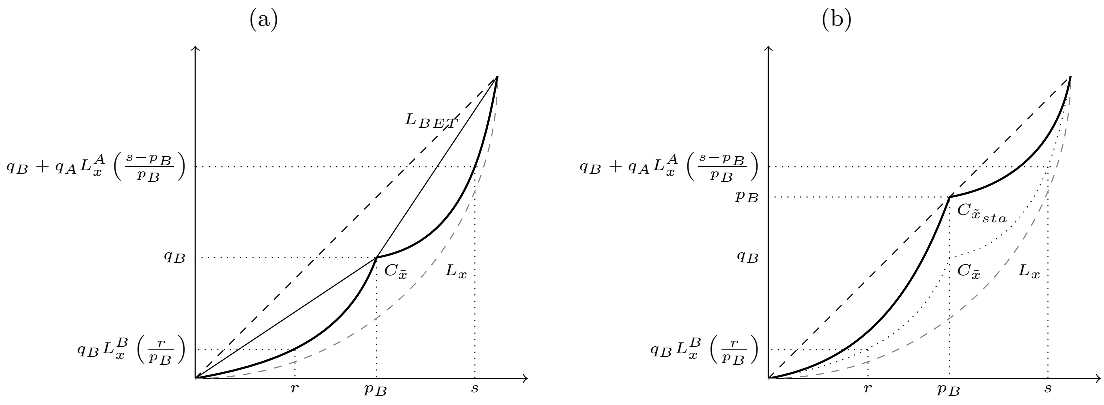
The within component of the additive Gini decomposition is the weighted sum of the Gini coefficients of each group, and corresponds to the area between L_{BET} and $C_{\tilde{x}}$:

$$G_{WIT}^{Add} = \sum_{i=A,B}\left[\frac{n_i}{n}\frac{n_i \mu_i}{n\mu} G_i\right]$$

that is

$$G_{WIT}^{Add} = p_A q_A G_A + p_B q_B G_B \tag{4}$$

Figure 1: Concentration and Lorenz curves



Finally, the residual is measured as twice the area between the concentration curve and the Lorenz curve of the overall distribution:

$$G_{RES} = 2\int_0^1 \Big(C_{\tilde{x}}(r) - L_x(r)\Big)dr$$

The alternative approach we have alluded to goes as follows. First, pool the mean-normalized incomes $\frac{x_A}{\mu_A}$ and $\frac{x_B}{\mu_B}$ in the two groups and sort them into increasing order. Call this the standardized distribution x_{sta}. Each group mean is 1, as is the overall mean: $\mu_{sta} = 1$. Denote by $\tilde{x}_{sta}$ the lexicographic parade of x_{sta} and observe that its concentration curve, $C_{\tilde{x}_{sta}}$, has the value $\frac{1}{n}\sum_{i\in B}\frac{x_i}{\mu_B} = \frac{n_B}{n} = p_B$ at position p_B. In other words, it touches the 45 degree line (See Figure 1(b)). Formally,

$$C_{\tilde{x}_{sta}}(r) = \begin{cases} p_B L_x^B\left(\frac{r}{p_B}\right) & 0 \le r \le p_B \\ p_B + p_A L_x^A\left(\frac{r-p_B}{p_B}\right) & p_B \le r \le 1 \end{cases} \tag{5}$$

Pictorially, $C_{\tilde{x}_{sta}}$ is obtained by drawing the Lorenz curve of the standardized distribution of B, first, and the one for group A after.

The path independent decomposition of the Lorenz curve is then obtained by the following steps. The reader may observe that Foster & Shneyerov (2000) do not decompose the

Lorenz curve, nor the Gini coefficient. Our use of the term path independent is aimed at underlining that, as in Foster & Shneyerov (2000) proposal, within group inequality is defined as inequality in the standardized distribution.

1. $x_\mu \rightarrow \tilde{x}_{sta}$; $L_{x_\mu} \rightarrow C_{\tilde{x}_{sta}}$: from equal distribution to standardized distribution where all incomes divided by the group mean, and ordered lexicographically (first by group, then by increasing order within each group), so that only within group inequality appears; [1]

2. $\tilde{x}_{sta} \rightarrow \tilde{x}$; $C_{\tilde{x}_{sta}} \rightarrow C_{\tilde{x}}$: by multiplying each income by the average income of the group, introduce between group inequality;

3. $\tilde{x} \rightarrow x$; $C_{\tilde{x}} \rightarrow L_x$: from the previous distribution to the final - original - distribution.

In the corresponding decomposition of the Gini coefficient, within group inequality is measured by twice the area between this concentration curve - $C_{\tilde{x}_{sta}}$ - and the 45 degree line:

$$G_{WIT}^{Path} = p_A^2 G_A + p_B^2 G_B \tag{6}$$

The above equation mirrors (4) after noticing that with mean-normalized incomes we have $p_A = q_A$. Between group inequality in this measurement system, G_{BET}^{Path}, is obtained as twice the area effect in the Lorenz diagram of moving from the concentration curve for $\tilde{x}_{sta}$ to the concentration curve for $\tilde{x}$ (thus un-normalizing the subgroup incomes again, and still beginning with group B):

$$G_{BET}^{Path} = 1 - 2\int_0^1 \left[C_{\tilde{x}_{sta}} - C_{\tilde{x}}\right] dr$$

[1] One can further decompose this passage in an intermediate one, that is $L_{x_\mu} \rightarrow L_{x_{sta}} \rightarrow C_{\tilde{x}_{sta}}$. In this way, while $L_{x_\mu} \rightarrow L_{x_{sta}}$ introduces within group inequality, $L_{x_{sta}} \rightarrow C_{\tilde{x}_{sta}}$ is a permutation aimed at netting out the groups' overlapping.

The path independent decomposition of the Gini coefficient is then $G = G_{BET}^{Path} + G_{WIT}^{Path} + G_{RES}$ and the fact that the Lorenz curve is not perfectly decomposable into a within and a between components, except in the very special case of non-overlapping subgroups, continues to hold.

As mentioned above, the within group component of the additive decomposition - Eq. (4) - is not independent of the inequality between groups. This is evident if we rewrite Eq. (4) as:[2]

$$G_{WIT}^{Add} = p_A q_A G_A + \left(1 - 2q_A + p_A q_A + G_{BET}^{Add}\right) G_B \tag{7}$$

so that the between group component of the additive decomposition clearly appears in the right hand side.

Observing that, by construction, the sum between and within components in the two alternative compositions must both coincide with twice the area between the 45 degree line and the concentration curve for $\tilde{x}$, it is clear that $G_{BET}^{Add} + G_{WIT}^{Add} = G_{BET}^{Path} + G_{WIT}^{Path}$. Therefore, combining $G_{BET}^{Path} = G_{BET}^{Add} + G_{WIT}^{Add} - G_{WIT}^{Path}$ with Eq. (3), (4) and (6), we can wirte

$$G_{BET}^{Path} = (q_A - p_A)\left[1 + p_A G_A - p_B G_B\right] \tag{8}$$

from which it is evident how the between group component of the path independent decomposition is sensitive to the inequality within groups.

In light of these features of the two approaches, in Section 3 we use simulated data to highlight the consequences of using either one of the two approaches to decompose inequality in between and within components. We then propose an alternative decomposition of the Gini index, based on the idea that between groups inequality should be

[2]Notice that $p_B q_B = (1 - p_A)(1 - q_A) = 1 - q_A - p_A + p_A q_A = 1 - 2q_A + (q_A - p_A) + p_A q_A$

measured with reference to the average groups' income.

Outside the mainstream economic literature, however, an alternative proposal by Dagum (1998) has received attention (see, for example Costa, 2008; Mussini, 2013). Dagum claims that using averages to measure inequality between groups is an oversimplification of the phenomenon. Intuitively, two groups may have the same average but completely different distributions. While using generalized means of order $\rho < 1$ may help accounting for the variability of each group's distribution, using representative incomes we cannot account for the overlap between groups, which is part of Dagum's definition of between group inequality. Before turning to our proposal in Section 3, the following subsection formalizes and discusses Dagum's decomposition.

2.2 Dagum's decomposition

Dagum (1998) proposes an alternative decomposition of the Gini index in which the within group component coincides with the one from the additive approach while between group inequality is not measured with reference to representative incomes. According to Dagum (1998, p. 47) "The use of only means to measure the income inequality between subpopulations is an oversimplification because the different variances and asymmetries of the income distributions and the transvariation of incomes between subpopulations are ignored".

A particularly relevant role in Dagum's decomposition is played by the *transvariation*. This concept, first introduced by Gini (1916), is linked with the groups' overlap and the residual term (G_{RES}) of the Gini decompositons based on representative incomes. In a nutshell, suppose that $\mu_A \geq \mu_B$ as before, then transvariation happens whenever an individual in B has income higher than an individual in A; that is, the two income units violate order between groups, defined in terms of average income.

Let us recall here the discrete formula of the Gini coefficient of x:

$$G = \frac{1}{2n^2\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|$$

Formally, we can express Dagum (1998)'s decomposition as $G = G_{WIT}^{Add} + G_{gBET}^{Dag}$ where the second element, measuring the *gross* between group inequality is

$$G_{gBET}^{Dag} = \frac{p_B q_A + p_A q_B}{n_B n_A (\mu_B + \mu_A)}\sum_{i=1}^{n_B}\sum_{j=1}^{n_A}|x_{iB} - x_{jA}| \tag{9}$$

As we can see, the gross between group inequality corresponds to all income differences stemming from units of different groups. Clearly, $G_{gBET}^{Dag} = G_{BET}^{Add} + G_{RES}$, that is, the gross between groups inequality in Dagum's decomposition includes everything that is not within group inequality G_{WIT}^{Add}. With reference to Figure 1(a), the reader may observe that Eq. (9) measures the area remaining between L_x and the first diagonal after excluding the area between $C_{\tilde{x}}$ and L_{BET}.

Eq. (9) can be further decomposed after isolating the income differences that respect the order between groups' averages, from those that transgress it. This allow us to define the first order moment of *transvariation*

$$G_{TRV}^{Dag} = -\frac{p_B q_A + p_A q_B}{n_B n_A (\mu_B + \mu_A)}\sum_{i=1}^{n_B}\sum_{j=1}^{n_A}\left[(x_{jA} - x_{iB})\,\mathbb{I}_{(x_{jA} - x_{iB})\leq 0}\right]$$

and the *net* between group inequality

$$G_{nBET}^{Dag} = \frac{p_B q_A + p_A q_B}{n_B n_A (\mu_B + \mu_A)}\sum_{i=1}^{n_B}\sum_{j=1}^{n_A}\left[(x_{jA} - x_{iB})\,\mathbb{I}_{(x_{jA} - x_{iB})\geq 0}\right]$$

given an indicator function $\mathbb{I}_c$ which takes value one if the condition c is true and zero otherwise (see aslo Costa, 2008).[3]

[3]Mussard *et al.* (2005) call G_{nBET}^{Dag} the *gross economic affluence*, a terminology that underlines how

None of the previous terms make explicit reference to a representative income. Nevertheless, it can be shown that $G^{Dag}_{nBET} = G^{Add}_{BET} + G^{Dag}_{TRV}$, which clarifies how Dagum's measure of between group inequality will tend to exceed the standard G^{Add}_{BET} based on average incomes. Moreover, using the notion of transvariation, we can rewrite the additive decomposition as $G = G^{Add}_{WIT} + G^{Add}_{BET} + 2G^{Dag}_{TRV}$, from which we have an explicit expression for the residual term in the previous section: $G_{RES} = 2G^{Dag}_{TRV}$.

The distinctive feature of Dagum's decomposition is that it does not identifies a portion of the Gini coefficient which has no normative relevance. The residual term of the representative income decompositions (G_{RES}) is rarely considered a relevant part of the distributive phenomena. This is not the case in Eq. 9, where the transvariation is part of the inequality between groups. We believe that these considerations call for further exploration of the normative content of the transvariation for inequality measurement. In this paper, however, we leave this question aside to focus on the representative income decompositions and a possible solution to the limitations discussed in Section 2.1.

3 A new representative income decomposition

We have concluded Section 2.1 showing that the path independent decomposition may induce a biased measure of the between group inequality, while in the additive approach biases may concern the within group component. In this section, we begin with an illustrative simulation of these dynamics and conclude by proposing an alternative decomposition of the Gini coefficient based on group averages as representative incomes.

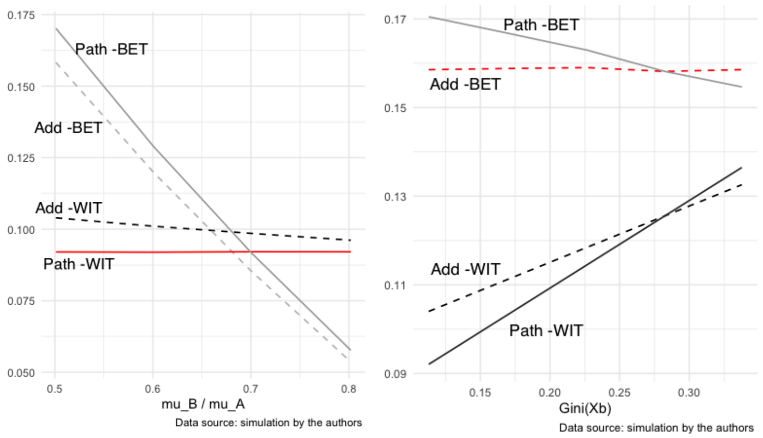
Let x_A be a vector of 50,000 values drawn from a normal distribution with average 15,000 and standard deviation equal to $0.4\mu_A$ the average. Also x_B, which is of dimension 40,000, is drawn from a normal distribution but, as in the previous section, we assume $\mu_A > \mu_B$.

this concept between group inequality is more linked to an idea of advantage of some groups with respect to others.

In Figure 2(a) we show the effect on G^{Add}_{BET}, G^{Add}_{WIT}, G^{Path}_{BET} and G^{Path}_{WIT} of changing μ_B from $0.5\mu_A$ to $0.9\mu_A$. To maintain a stable inequality within x_B, at each iteration we set the standard deviation equal to $0.2\mu_B$.

Figure 2: Absolute changes in the Gini components

(a) (b)



One would expect, ceteris paribus, a change in the average income of group B to have no effect on the within component of the Gini coefficient. Figure 2(a) shows that this is indeed the case for G^{Path}_{WIT} (solid red line). However, as also highlighted in Eq. (7), under the additive decomposition approach, the within group component of the Gini coefficient (dashed black line) does respond to changes in μ_B/μ_A.[4] This behaviour of G^{Add}_{WIT} makes up for the difference we observe between G^{Add}_{BET} and G^{Path}_{BET} in the same figure.

In Figure 2(b) we fix $\mu_B = 0.5\mu_A$ but increase the income inequality in group B. All things equal, this should only affect the within group component of the Gini coefficient. Our results show that, in line with Eq. (8), this is not the case when between group inequality is defined by G^{Path}_{BET} (solid gray line).[5] We may observe that the graph of the between group component of the additive decomposition (dashed red line) is not completely flat. This is due to small variations of μ_B/μ_A due to sample size.

[4]In particular, G^{Add}_{WIT} is an increasing function of G^{Add}_{BET}.
[5]In particular, G^{Path}_{BET} a decreasing function of G^{Path}_{WIT}.

Therefore, if we are interested in decomposing the Gini index into within and between groups inequality, it does matter which of the two components is of prominent importance in our study, as this can guide us in choosing between these alternative approaches to the Gini decomposition. While G_{BET}^{Add}, which stems from the additive decomposition, is a cleaner measure of the inequality between (average incomes of) groups, the path independent decomposition provides an unbiased measure of the within group inequality: G_{WIT}^{Path}.

As useful illustrative example, we shall consider the problem of measuring equality of opportunity. The literature has propose two approaches to the definition of equal opportunities. The firs, ex ante, approach focuses on groups of individuals with homogeneous characteristics - circumstances at birth, out of individual responsibility - and calls inequality of opportunity the fraction of inequality due to the between group component. The second, ex post, approach considers groups of individuals homogeneous along some dimension of effort or individual responsibility and calls unfair the fraction of total inequality due to the within group component. Following our analysis, it is evident that, to obtain an unbiased measure of inequality of opportunity, one should favor the additive decomposition of the Gini coefficient when implementing the ex ante view, and the path independent decomposition for the ex post approach.

In practice, one may then consider that the best decomposition of the Gini is actually of the type

$$G = G_{BET}^{Add} + G_{WIT}^{Path} + G_{RES}^{*}, \tag{10}$$

where one has

$$\begin{aligned} G_{RES}^{*} &= G_{RES} + G_{BET}^{Path} - G_{BET}^{Add} \\ &= G_{RES} + G_{WIT}^{Add} - G_{WIT}^{Path}. \end{aligned}$$

and, in particular,

$$G^*_{RES} = G - \underbrace{(q_A - p_A)}_{G^{Add}_{BET}} + \underbrace{\sum_{k=A,B} [p_k G_k (q_k - p_k)]}_{G^{Add}_{WIT} - G^{Path}_{WIT}} - \underbrace{\sum_{k=A,B} [p_k q_k G_k]}_{G^{Path}_{WIT}} \tag{11}$$

which is an easy-to-compute expression that depends only on the total Gini coefficient, the Gini coefficient of each group, and population and income shares.

It is noteworthy that, in absence of overlap, one has $G_{RES} = 0$ but in general one still has $G^*_{RES} \neq 0$. Observe that $G^*_{RES} = G_{RES}$ only if $G^{Add}_{WIT} = G^{Path}_{WIT}$, which happens if either there is no between group inequality - $G^{Add}_{BET} = 0$ - or $p_A/p_B = (G_A/G_B)^{-1}$. The dynamics of G^*_{RES} are not straightforward, as it depends on the way between group inequality, within group inequality and transvariation interact. For example, increasing q_A will increase G^*_{RES} of $p_A G_A$, but reduce the overlap between the income distribution of the two groups, hence G_{RES}. We do not consider, however, that this is a strong argument against the proposed decomposition, in particular given the small emphasis so far given to the residual term.

Notably, our proposal can be seen as the proper counterpart, in terms of Gini coefficient, of the decomposition proposed by Foster & Shneyerov (2000). Indeed, as required by the original path independent decomposition, (see Foster & Shneyerov, 2000, p. 204) between group inequality in Eq. (10) is the inequality in the smooth distribution, while between group inequality is measured on the standardized distribution.

4 Partial orderings

We now consider the issue of decomposability of the Lorenz partial ordering into a within and a between component. To illustrate the issue we want to address, consider the following problem. Let $>_L$ define the Lorenz partial ordering. Given a definition of

Lorenz between-group partial ordering $>_{LB}$ and a Lorenz within group partial ordering $>_{LW}$, suppose that, given two distributions x and y, we have $x >_{LB} y$ and $x >_{LW} y$. Can we say that the two conditions above imply that $x >_L y$? If not, is there a partial ordering $>_*$ implied by the partial orderings $>_{LB}$ and $>_{LW}$?

We consider the orderings of income distributions for $A \cup B$, with $\mu_A \geq \mu_B$, induced by the concentration curve $C_{\tilde{x}}$ and between-groups Lorenz curve L_{BET} of Figure 1(a), and by the concentration curve $C_{\tilde{x}_{sta}}$ of Figure 1(b). Each of these involves a lexicographic ranking of income units in which incomes are listed by magnitude within groups, and with group B appearing first (the lexicographic parade). Let r_j denote the rank of an individual within the income distribution of his type $j \in \{A, B\}$, so that $r_B = \frac{r}{p_B}$ if $0 \leq r \leq p_B$, and let $r_A = \frac{r-p_B}{p_A}$ if $p_B \leq r \leq 1$. Each of these runs from 0 to 1, and marks the position of the income unit at overall position $r = \frac{i}{n}$ in its group's income parade. In these terms, the within group component of the additive decomposition is given by

$$C_{\tilde{x}}(r) - L_{BET}(r) = \begin{cases} q_B \left(L_x^B(r_B) - r_B \right) & 0 \leq r \leq p_B \\ q_A \left(L_x^A(r_A) - r_A \right) & p_B \leq r \leq 1 \end{cases} \tag{12}$$

while in the path independent decomposition we have

$$C_{\tilde{x}_{sta}}(r) = \begin{cases} p_B L_x^B(r_B) & 0 \leq r \leq p_B \\ p_B + p_A L_x^A(r_A) & p_B \leq r \leq 1 \end{cases} \tag{13}$$

Now we take another income distribution y over $A \cup B$ in which the subgroup means have the same relationship as in x: $\mu_A^y \geq \mu_B^y$. The same mathematics applies, and so, taking the difference,

$$C_{\tilde{x}}(r) - C_{\tilde{y}}(r) = \left[L^x_{BET} - L^y_{BET}\right] + \begin{cases} q^x_B\left[L^B_x(r_B) - r_B\right] - q^y_B\left[L^B_y(r_B) - r_B\right] & 0 \le r \le p_B \\ q^x_A\left[L^A_x(r_A) - r_A\right] - q^y_A\left[L^A_y(r_A) - r_A\right] & p_B \le r \le 1 \end{cases} \tag{14}$$

and

$$C_{\tilde{x}_{sta}}(r) - C_{\tilde{y}_{sta}}(r) = \begin{cases} p_B\left[L^B_x(r_B) - L^B_y(r_B)\right] & 0 \le r \le p_B \\ p_B + p_A\left[L^A_x(r_A) - L^A_y(r_A)\right] & p_B \le r \le 1 \end{cases} \tag{15}$$

Let us define a strong within-group partial ordering $\succ_{WIT}$ such that $x \succ_{WIT} y$ if and only if $L^A_x > L^A_y$ and $L^B_x > L^B_y$. From (15), we have the following result.

Proposition 1. *$C_{\tilde{x}_{sta}} > C_{\tilde{y}_{sta}}$ if and only if $x \succ_{WIT} x$*

In words, dominance in terms of concentration curves of the standardized income distribution is equivalent to Lorenz dominance in each group. This reflects and strengthen the idea of G^{Path}_{WIT} being an adequate measure of within group inequality. If the means in y are ordered in reverse, i.e. if $\mu^y_A < \mu^y_B$, results (14) and (15) cannot be sustained, and Proposition 1 of course fails. If, however, $p_A = 0.5$ and we admit anonymity as normative principle, then a simple relabelling of the groups would allow us to recover the condition $\mu^y_B \le \mu^y_A$.

For the between-group partial ordering, $\succ_{BET}$ we rely on the conventional approach to define $x \succ_{BET} y$ if and only if $L^x_{BET} > L^y_{BET}$. It is easy to see that $x \sim_{BET} y$ if and only if $q^x_A = q^y_A$. Consequently, if $x \sim_{BET} y$ and $x \succ_{WIT} y$, then $C_{\tilde{x}} > C_{\tilde{y}}$. The latter is a consequence of the following result

Proposition 2. *If $x \succ_{BET} y$ and $x \succ_{WIT} y$, then $C_{\tilde{x}} > C_{\tilde{y}}$*

Proof. Let $x \succ_{BET} y$ and $x \succ_{WIT} y$, we need to show that

$$C_{\tilde{x}} - C_{\tilde{y}} = \begin{cases} q_B^x L_x^B(r_B) & -q_B^y L_y^B(r_B) \\ q_B^x + q_A^x L_x^A(r_A) & -q_B^y - q_A^y L_y^A(r_A) \end{cases} \geq 0 \tag{16}$$

By $x \succ_{WIT} y$, $L_x^B(r_B) \geq L_y^B(r_B)$ and $L_x^A(r_A) \geq L_y^A(r_A)$. Moreover, from the first line of equation (1), $x \succ_{BET} y$ implies $\frac{q_B^x}{p_B} \geq \frac{q_B^y}{p_B}$, that is $q_A^x \leq q_A^y$.

Let us consider the first line of (16). Sufficient conditions for

$$q_B^x L_x^B(r_B) \geq q_B^y L_y^B(r_B)$$

are $L_x^B \geq L_y^B$ and $q_B^x \geq q_B^y$: both following by assumption. Let $q_A^x = q_A^y - \epsilon$; by assumption $\epsilon \geq 0$. Let us rewrite the second line of (16) as

$$\begin{aligned} q_B^x + q_A^x L_x^A(r_A) &\geq (q_B^x - \epsilon) + (q_A^x + \epsilon) L_y^A(r_A) \\ q_A^x L_x^A(r_A) &\geq -\epsilon + q_A^x L_y^A(r_A) + \epsilon L_y^A(r_A) \\ q_A^x \left(L_x^A(r_A) - L_y^A(r_A)\right) + \epsilon &\geq \epsilon L_y^A(r_A) \end{aligned}$$

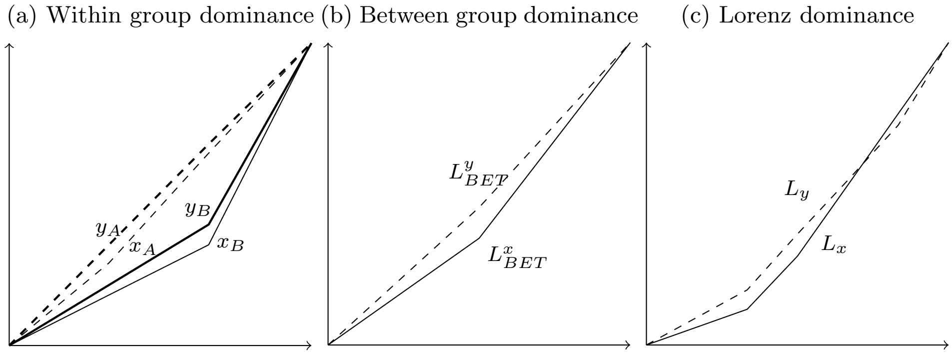
Observe that, by assumption, $q_A^x \left(L_x^A(r_A) - L_y^A(r_A)\right) \geq 0$ and $\epsilon \geq 0$. It is then sufficient to notice that, by definition, $L_y^A \leq 1$. The desired result.

□

In words, between group Lorenz dominance and within group dominance are sufficient to induce dominance in terms of concentration curves. However, $x \succ_{BET} y$ and $x \succ_{WIT} y$, may not be sufficient for $L_x > L_y$. To see this, consider the following example.

Let $x^A = (4, 4, 3)$ and $x^B = (4, 1, 1)$ and $y^A = (2, 2, 2)$ and $y^B = (1, 3, 1)$. In this case we have $\mu_A > \mu_B$ for both distributions, $L_y^A > L_x^A$, $L_y^B > L_x^B$ and $L_{BET}^y > L_{BET}^x$. Nevertheless, we do not have $L_y > L_x$ (see Figure 3). The cause of this result is the overlap between groups, captured by the area between $C_{\tilde{x}}$ and L_x.

Figure 3: Within and between partial orders do not imply Lorenz dominance



In absence of overlap, in contrast, between and within Lorenz dominance do suffice to guarantee full Lorenz dominance.

Proposition 3. *Let $x, y \in \mathbb{R}_+$ be two distribution for the population $A \cup B$, with $\mu_A \geq \mu_B$. If $\min x_A > \max x_B$ and $\min y_A > \max y_B$, then $x \succ_{BET} y$ and $x \succ_{WIT} y$ imply $L_x > L_y$.*

Proof. This result directly follows from the previous proposition and the fact that when there is no overlap, the curves of $C_{\tilde{x}}$ and L_x coincide. □

The reader may notice the link between $\succ_{BET}$, $\succ_{WIT}$ and, respectively, G^{Add}_{BET}, G^{Path}_{WIT}. In this sense, Propositions 2 and 3 offer normative justification for considering eq. (10) a suitable decomposition of the Gini coefficient.

5 Conclusion

After reviewing and comparing two alternative procedures for decomposing the Lorenz curve into between and within groups inequality, we have highlighted the presence of substantial differences between them. The standard additive decomposition process, defines within group inequality as the weighted sum of the Gini indices in each group, and the between group inequality as the inequality in the smoothed distribution. We showed that this approach offers a clean measure of the between group inequality which is less sensitive to changes in the distribution that do not alter the relative difference between the groups' averages. The path independent decomposition redefines the concept of within group inequality looking at a standardized distribution where groups have the same average income and do not overlap. This offers an unbiased measure of the within group inequality which remains insensitive to changes in the between group component.

We have underlined the strengths and limitations of the two approaches and proposed another decomposition of the Gini index which combines the definition of between group inequality stemming from the additive decomposition, with the measure of within group inequality at the base of the path independent decomposition.

Our discussion on the decomposition of the partial Lorenz ordering showed the difficulty of defining sufficient conditions for two Lorenz curves not to intersect. Much of this complexity comes from the residual term in the Gini decomposition. In this sense, our results show that the concentration curve of the lexicographic parade constitutes a useful too to decompose the Lorenz ordering while netting out the effect of the group's overlap.

Despite the link with the residual term of the Gini decomposition (Aronson & Lambert, 1994), the literature lacks a normative assessment of the equity role played by the group's overlap (or transvariation). We believe that the Gini index and his decomposition has not been completely understood yet, and the widespread use of this measure strongly motivates further research on its property and interpretation.

References

Aronson, J Richard, & Lambert, Peter J. 1994. Decomposing the Gini coefficient to reveal the vertical, horizontal, and reranking effects of income taxation. *National Tax Journal*, **47**(2), 273–294.

Bhattacharya, Nath, & Mahalanobis, B. 1967. Regional disparities in household consumption in India. *Journal of the American Statistical Association*, **62**(317), 143–161.

Blackorby, Charles, Donaldson, David, & Auersperg, Maria. 1981. A new procedure for the measurement of inequality within and among population subgroups. *Canadian Journal of Economics*, 665–685.

Checchi, Daniele, & Peragine, Vito. 2010. Inequality of opportunity in Italy. *The Journal of Economic Inequality*, **8**(4), 429–450.

Costa, Michele. 2008. Gini index decomposition for the case of two subgroups. *Communications in Statistics—Simulation and Computation®*, **37**(4), 631–644.

Dagum, Camilo. 1998. A new approach to the decomposition of the Gini income inequality ratio. *Pages 47–63 of: Income inequality, poverty, and economic welfare.* Springer.

Duncan, Otis Dudley, & Duncan, Beverly. 1955. A methodological analysis of segregation indexes. *American sociological review*, **20**(2), 210–217.

Fedorov, Leonid. 2002. Regional inequality and regional polarization in Russia, 1990–99. *World development*, **30**(3), 443–456.

Foster, James E, & Shneyerov, Artyom A. 2000. Path independent inequality measures. *Journal of Economic Theory*, **91**(2), 199–222.

Gastwirth, Joseph L. 1975. Statistical measures of earnings differentials. *The American Statistician*, **29**(1), 32–35.

Gini, Corrado. 1916. *Il concetto di" transvariazione" e le sue prime applicazioni.* Athenaeum.

Handcock, Mark S, & Morris, Martina. 2006. *Relative distribution methods in the social sciences.* Springer Science & Business Media.

Kestenbaum, Bert. 1979. Coefficients of between-group inequality: a review. *ORES Working Paper*, **No.**(8).

Lambert, Peter J, & Aronson, J Richard. 1993. Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal*, **103**(420), 1221–1227.

Lambert, Peter J, & Decoster, André. 2005. The Gini coefficient reveals more. *Available at SSRN 809004.*

Lambert, Peter J, & Subramanian, Subbu. 2014. Disparities in socio-economic outcomes: some positive propositions and their normative implications. *Social Choice and Welfare*, **43**(3), 565–576.

Le Breton, Michel, Michelangeli, Alessandra, & Peluso, Eugenio. 2012. A stochastic dominance approach to the measurement of discrimination. *Journal of Economic Theory*, **147**(4), 1342–1350.

Mussard, Stéphane, Alperin, María Noel Pi, Seyte, Françoise, & Terraza, Michel. 2005. Extensions of Dagum's Gini decomposition. *In: International Conference in Memory of Two Eminent Socia Scientists: C. Gini and MO Lorenz. Their impact in the XX-th century development of probability, statistics and economics.*

Mussini, Mauro. 2013. A matrix approach to the Gini index decomposition by subgroup and by income source. *Applied Economics*, **45**(17), 2457–2468.

Shorrocks, Anthony F. 1980. The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society*, 613–625.

Wolff, Edward N. 1976. Occupational earnings behavior and the inequality of earnings by sex and race in the United States. *Review of Income and Wealth*, **22**(2), 151–166.

Zhang, Xiaobo, & Kanbur, Ravi. 2001. What difference do polarisation measures make? An application to China. *Journal of development studies*, **37**(3), 85–98.

Appendices

A The many group case in brief

Let the groups be $A_1, A_2, ..., A_m$ be of sizes $n_{1,2}, ..., n_m$ and let $p_i = \frac{n_i}{n}$ and $q_i = \frac{n_i \mu_i}{n\mu}$ where n is the total population and μ the average income, so that $q_i = p_i \frac{\mu_i}{\mu}$. Suppose the means are ordered as $\mu_1 \geq \mu_2 \geq .... \geq \mu_m$. In these terms we have

$$G_{BET}^{Add} = \sum_i \sum_{(i,j):i<j} |p_i q_j - p_j q_i| \tag{17}$$

$$G_{WIT}^{Add} = \sum_i p_i q_i G_i \tag{18}$$

$$G_{WIT}^{Path} = \sum_i p_i^2 G_i \tag{19}$$

$$G_{BET}^{Path} = \sum_i \sum_{(i,j):i<j} |p_i q_j - p_j q_i| + \sum_i p_i (q_i - p_i) G_i \tag{20}$$

Consider now orderings of income distributions in $\cup_{i=1}^m A_i$. Let $\sum_{j<i} p_j \leq r \leq \sum_{j \leq i} p_j$, so that the income unit at rank r in the lexicographic parade belongs to A_i. Then $t_i = \frac{r - \sum_{j<i} p_j}{p_i} \in (0, 1)$ is hte rank of this income unit in its group's income parade.

Certain results in the main text can be generalized to this situation. Namely, for the concentration curves we have

$$C_{\tilde{x}}(r) - L_{BET}(r) = q_i \left(L_x^i(t_i) - t_i \right) \tag{21}$$

$$C_{\tilde{x}_{sta}}(r) = \sum_{j<i} p_j + p_i L_x^i(t_i) \tag{22}$$

Taking another income distribution y over the same population, with the same ordering of means, we have these equivalents of 14 and 15:

$$C_{\tilde{x}}(r) - C_{\tilde{y}}(r) = \left[L^x_{BET}(r) - L^y_{BET}(r) \right] + q_i^x \left(L_x^i(t_i) - t_i \right) - q_i^y \left(L_y^i(t_i) - t_i \right) \tag{23}$$

$$C_{\tilde{x}_{sta}}(r) - C_{\tilde{y}_{sta}}(r) = p_i \left[L_x^i(t_i) - L_y^i(t_i) \right] \tag{24}$$

with the same implications for orderings as in Propositions 1, 2 and 3.