Robust dissimilarity comparisons with ordinal outcomes

Francesco Andreoli[a,b,*] and Claudio Zoli[a]

[a] *Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy*
[b] *Luxembourg Institute of Socio-Economic Research (LISER), 11 Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg*

February 2022
Preliminary and incomplete, do not cite or quote

Abstract

The analysis of many economic phenomena requires partitioning societies into groups, gathering individuals sharing the same circumstances of birth like gender, birthplace, cohort, ability or parental background, and studying the extent at which these groups are distributed with different intensities across relevant ordered outcomes, such as income, health or cognitive score levels. When the groups are similarly distributed, their members could be seen as having equal chances to achieve any of the attainable outcomes. Otherwise, a form of dissimilarity prevails. We frame dissimilarity comparisons of multi-group distributions defined over ordinal outcomes by showing the equivalence between axioms underpinning information criteria, basic transformations of the data regarded to as unambiguously preserving or reducing dissimilarity and a new empirical test based on sequential dominance conditions. Mainstream approaches related to intergenerational mobility, equality of opportunity, discrimination and the analysis of distance between distributions are shown to be embedded within the dissimilarity model that we characterize. An application to Swedish data highlight the usefulness of the criteria to identify the intergenerational distributional consequences of a large education reform which took place in the 1960s.

[*]Contact details: `francesco.andreoli@univr.it` (F. Andreoli, corresponding author) and `claudio.zoli@univr.it` (C. Zoli). We are indebted to Valentino Dardanoni, Vincenzo Denicoló, Louis Eeckhaudt, Martin Foster, Alessio Fusco, Arnaud Lefranc, Brice Magdalou, Patrick Moyes, Eugenio Peluso, Gaston Yalonetzky and an anonymous referee for their useful comments. A preliminary version of this paper was presented at the 2011 ECINEQ meeting, LAGV 2012, 2012 SSCW meeting, 2012 EEA-ESEM meeting, and seminars at IT2012 Winter School, THEMA, Verona, Palermo, CORE, Bocconi University, Leicester University, UABarcelona, CERDI/FERDI. This paper forms part of the research projects *The Measurement of Ordinal and Multidimensional Inequalities* (Contract No. ANR-16-CE41-0005-01 - ORDINEQ) of the French National Agency for Research, the research project MOBILIFE (grant RBVR17KFHX) and PRE-OPP (grant RBVR19FSFA) of the University of Verona Basic Research scheme and the NORFACE/DIAL project *IMCHILD: The impact of childhood circumstances on individual outcomes over the life-course* (grant INTER/NORFACE/16/11333934/IMCHILD) of the Luxembourg National Research Fund (FNR), whose financial support is gratefully acknowledged.

Keywords: Dissimilarity, mobility, equality of opportunity, sequential dominance.

JEL Codes: D63, J71, J62, D30.

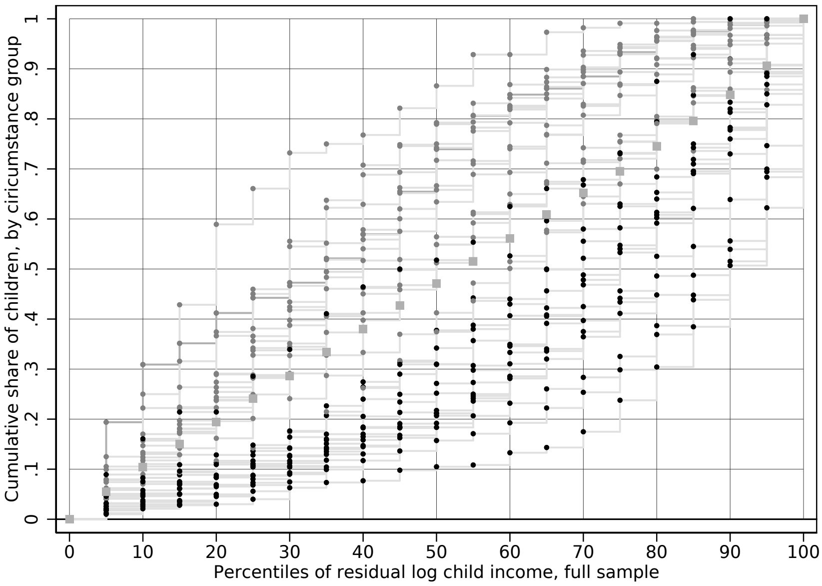
1 Introduction

Many economic phenomena are concerned with situations in which individuals are partitioned into different social groups on the basis of characteristics they share in common. In this paper, we are interested in comparing conditional bivariate distribution, in which one attribute is ordinal (the outcome) and the other is categorical (the group label). Each distribution specifies the proportions of each group that attain each of a finite number of ordered outcomes.

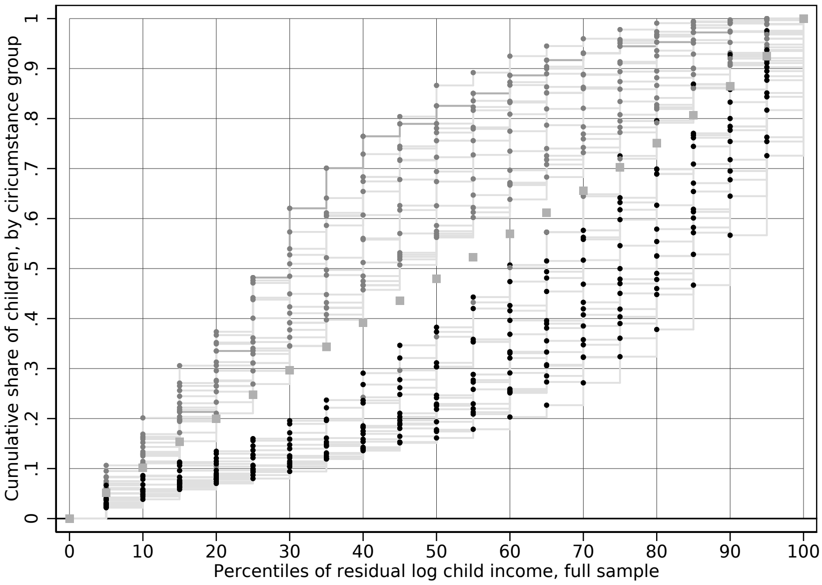
For instance, *discrimination* analysis is concerned with the way in which groups, defined along the lines of gender or race, attain different wage levels or occupations ordered by social prestige with different intensities. *Health or educational inequality* arises when the chances of attaining each one of the ordered outcomes (such as self-assessed health or standardized test scores) differ across individuals with different characteristics (such as by location, socio-economic status, demographics). Transition matrices, specifying the probability of achieving a given percentile in the child's income distribution conditionally on the percentile of departure in the parents' income distribution, are used to asess *intergenerational mobility.* A transition matrix displays low mobility when the distributions are highly dissimilar. More broadly, in analyzing *unfair inequality*, the focus is not much on the inequality in outcomes realized by people with similar traits that fall beyond individual responsibility, but rather on differences across groups defined by different traits.

Figure 1 depicts an empirical example of unfair inequality. Each panel reports the distribution of 32 groups, defined on the basis of gender, ability, place of birth and parental education, across population income vingitiles estimated from a representative sample of the Swedish population born 1948 and 1953.[1] Black and gray circles in each graph identify cumulative distributions conditional on group belonging, distinguishing those treated by a large compulsory education reform (the treatment group) and those who were not (the

[1]The data, from Meghir and Palme (2005), are described in Section 5.



(a) Control group, matrix **C**



(b) Treatment group, matrix **T**

Figure 1: Cumulative frequencies across income percentiles, by circumstances group and treatment status.

Note: Authors' computations based on Meghir and Palme (2005) data. Groups formed by interacting information on parents education, gender, ability and location. Gray dots are for groups gathering women. Gray lines allow to recollect each groups distribution. The gray squares correspond to averages of groups distributions by realization.

control group). There are 20 outcomes and 32 groups in both examples. Figure 1 portraits a comprehensive perspective about unfair inequality in both treatment and control groups, with some groups attaining worst income realizations with larger intensity than others. Some groups even suffer a disadvantage compared to others across all percentile, arguably a strong violation of fairness. This is the case, for instance, for the groups gathering women (black dots) as opposed to men (gray dots). In both panels of the figure, a form of dissimilarity prevails.

This paper investigates a criterion for ranking cumulative bivariate distributions by the extent of dissimilarity they display. The criterion can be used, for instance, to compare distributions in the treatment and control groups of Figure 1, and to conclude about whether dissimilarity in distributions (capturing unfair inequalities) has been reduced by effect of the intervention. We provide an axiomatic characterization of a robust dissimilarity criterion for ranking multivariate distributions of groups across ordered (but not necessarily cardinal) classes of realizations. The criterion is robust in the sense that it gathers agreement on the ranking of configurations for a variety of different perspectives about dissimilarity. The dissimilarity criterion induces a partial order of configurations.

There is widespread agreement in the literature about what constitutes perfect intergenerational mobility or lack of discrimination. These are situations in which the groups are *similarly distributed* across the attainable outcomes. The relevant notion of similarity dates back to the work of Gini (1914), where it is argued that two (or more) groups are similarly distributed whenever "the overall populations of the two groups take the same values with the same frequency."[2] In this case, groups are equally represented in each class, albeit groups may take on different outcomes with different intensity. Conversely, the case of maximal dissimilarity occurs when the groups membership can be identified from the knowledge of the class of outcomes. This is always the case when groups are ordered by

[2]See Gini (1914, p. 189) and XXX Andreoli and Zoli for a discussion about similarity.

stochastic dominance and their distributions are not overlapping, i.e. the highest realization achieved by any of the groups is always smaller than the lowest realization achieved by any other group which is better off.

We resort to an axiomatic model to rationalize comparisons of the configurations in-between the similarity and maximal dissimilarity cases. Consider again Figure 1. Visually, overall dissimilarity comes down to the extent of heterogeneity in groups cumulative frequencies at any ordered outcome. Reducing dissimilarity would require to reduce heterogeneity. If the distribution were similar, all dots in the graphs would overlay the average distribution (gray squares). We argue that dissimilarity always decreases whenever improvements of the situation for a proportion of the group experiencing low realizations with higher frequency is counterbalanced by a deterioration of the situation for an equal proportion of the group experiencing high realizations with higher frequency. This sequence of displacements is compounded into an *exchange* transformation.

The main result of the paper is a dissimilarity analog of the fundamental theorems in the measurement of inequality and risk (see Hardy, Littlewood and Polya 1934, Marshall, Olkin and Arnold 2011, Gajdos and Weymark 2012, Andreoli and Zoli 2020), The main theorem shows the equivalence between (i) the unanimity in ranking configurations for all dissimilarity orderings consistent with the effects of exchange transformations as well of operations that unambiguously preserve dissimilarity, (ii) a representation of the partial order through dissimilarity indices, (iii) the majorization condition which relates dissimilarity to the extent of inequality in groups cumulative frequencies across all outcomes and (iv) an implementation criterion which generalizes the orthant orders for bivariate distributions.[3] The theorem complements XXX Andreoli Zoli, which offers a characterization for robust dissimilarity comparisons when classes are categorical and non-ordered. Our main result

[3]Landmark contributions introducing orthant orders are Epstein and Tanny (1980), Tchen (1980), Dardanoni (1993) and (see also Ch. 6.G, Shaked and Shanthikumar 2006). Claim (iv) extends the orthant order criterion to bivariate distributions where groups are not ordered according to an exogenous label and their distributions cannot be ordered by stochastic dominance.

is also shown to rationalize and to generalize to the multi-group setting a variety of sparse and apparently unrelated results on the measurement of discrimination[4], unfair inequality[5] (often assuming that groups distributions are ordered by stochastic dominance), intergenerational mobility[6] (assuming that groups labels are exogenously ordered) and distance between distributions[7] (limited to the case with two groups).

The rest of the paper is organized as follows. Section 2 provides the setting, while axioms and the main results are in Section 3. The criteria related to dissimilarity are discussed in Section 4. Section 5 outlines an application of the dissimilarity criterion to evaluate the implications of the Swedish education reform on unfair inequality. Section 6 concludes. Proofs are collected in Appendix A, whereas Appendix B provides an algorithm to implement the empirical dissimilarity criterion.

2 Dissimilarity comparisons preserving ordinal information

2.1 Definition

We outline a dissimilarity criterion for population cumulative distributions (cdf) of a cardinal or ordinal attribute. Let $X_i \sim F_i$ indicate a random variable distributed as F_i, which is the population cdf for group i defined over the domain $x \in \mathbb{R}$. We consider the case in which the dissimilarity between distributions F_1, F_2 has to be compared with dissimilarity

[4]For a formal treatment based on discrimination curves, see Gastwirth (1975), Dagum (1980), Butler and McDonald (1987), Jenkins (1994), Le Breton, Michelangeli and Peluso (2012) and Fusco and Silber (2013). Lefranc, Pistolesi and Trannoy (2009) and Roemer (2012) refer more explicitly to equality of opportunity.

[5]See Lefranc et al. (2009) and Roemer (2012), who refer more explicitly to equality of opportunity. Andreoli, Havnes and Lefranc (2019), **?** and **?** resort on distance-based criteria to evaluate dissimilarity between distributions of a cardinal attribute and use this criterion to evaluate the opportunity equalizing effect of public policies.

[6]Dardanoni (1993) and Van de gaer, Schokkaert and Martinez (2001) discuss the robust ordering of (intergenerational) mobility matrices where both groups and achievements (respectively, the distribution of departure and that of destination) are ordered.

[7]See Shorrocks (1982) and Ebert (1984).

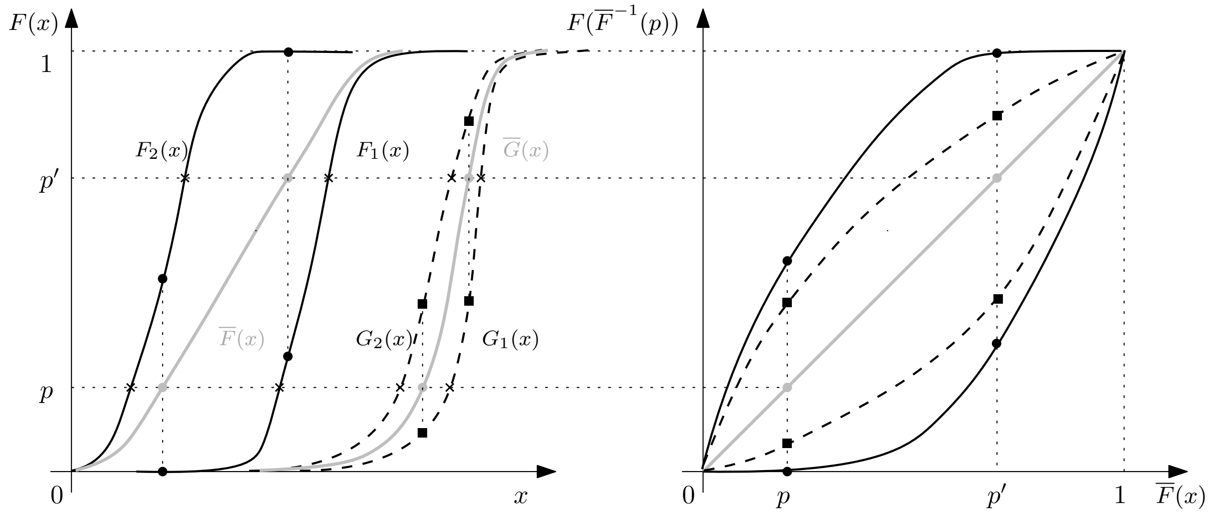


Figure 2: The dissimilarity criterion, based on cdf (F_1, F_2) and (G_1, G_2) (left panel) and their representations preserving ordinal information (right panel).

among distributions G_1, G_2. These comparisons should preserve exclusively the ordinal information underlying these distributions, which can be related, for instance, to fundamental heterogeneity unobservable to the researcher (Athey and Imbens 2006). We use random variables $U_i \sim \tilde{F}_i$ to represent the extent of unobserved heterogeneity.

Figure 2 outlines the evaluation problem concerning pairs of cdfs F_1, F_2 and G_1, G_2, pictured on the left-hand side of the figure. In the example, we care about robust assessments of dissimilarity in $(\tilde{F}_1, \tilde{F}_2)$. An intuitive criterion is to focus on gaps between distributions: we conclude that $(\tilde{G}_1, \tilde{G}_2)$ displays at most as much dissimilarity as $(\tilde{F}_1, \tilde{F}_2)$ whenever

$$|\tilde{G}_1(u) - \tilde{G}_2(u)| \leq |\tilde{F}_1(u) - \tilde{F}_2(u)|, \ \forall u \in [0, 1]. \quad (1)$$

However, we only observe (F_1, F_2) and (G_1, G_2). Each pair of cdfs displays dissimilarity, but the two pairs are not readily comparable: they do not even share the same domain of realizations. In order to compare the two distributions, we link each pair to a reference distribution $\overline{F}$ and $\overline{G}$, plotted in gray on the graph. The two distributions lie in-between F_1, F_2 and G_1, G_2 but $\overline{G}(x) \neq \overline{F}(x)$ at any x.

Consider an onto function $\overline{F}:\mathbb{R}\mapsto[0,1]$ mapping $\overline{F}(x)=\overline{F}(F_1(x),F_2(x))$. We require $\overline{F}$ to satisfy some desirable properties, such as consistency, symmetry (which introduces a form of anonymity with respect to groups weights) and monotonicity (ruling out distributions that are not identified over the entire domain of realizations).[8] A meaningful representation of $\overline{F}$ consistent with these desirable properties is $\overline{F}(x)=\frac{1}{2}\sum_{i=1}^{2}F_i(x)$ for any $x\in\mathbb{R}$, which is denoted the *average groups distribution* in the population. This function is continuous and invertible, such that for any fractional rank $p\in[0,1]$, $x=\overline{F}^{-1}(p)$. If we further assume that observable realizations and unobservable heterogeneity are related by the mapping $x=h(u)$, with h a monotonic increasing transformation h, then we can recover the unobservable $\tilde{F}_i$ as follows:

$$\begin{aligned}\tilde{F}_i(u) &:= \Pr[U_i\leq u] \;=\; \Pr[U_i\leq h^{-1}(x)]\\ &= \Pr[h(U_i)\leq x] \;=\; \Pr[X_i\leq \overline{F}^{-1}(p)]\\ &= F_i(\overline{F}^{-1}(p))\end{aligned}$$

In Figure 2, we project quantiles of $\overline{F}^{-1}(p)$ and, at any of these quantiles, we identify levels $(F_1(\overline{F}^{-1}(p)),F_2(\overline{F}^{-1}(p)))$ corresponding to all fractional rank $p\in[0,1]$. Levels corresponding to distributions F_1,F_2 and G_1,G_2 are identified respectively by black circles and boxes symbols. The dissimilarity criterion (1) is identified by the condition:

$$|G_1(\overline{G}^{-1}(p))-G_2(\overline{G}^{-1}(p))|\leq|F_1(\overline{F}^{-1}(p))-F_2(\overline{G}^{-1}(p))|,\ \forall p\in[0,1].$$

The gaps in cumulative groups frequencies are identified by the vertical distance between each pair of distributions at given level p. Gaps in each pair of distributions are evaluated at quantiles of the reference distributions $\overline{F}$ and $\mathbf{G}$. The right-hand side of Figure 2 provides

[8] Consistency requires that $\overline{F}(F(x),\ldots,F(x))=F(x)$. Symmetry requires instead that $\overline{F}(F_{\pi 1}(x),\ldots,F_{\pi d}(x))=F(x)$ for any permutation πi of the label of the groups. Monotonicity requires that $\overline{F}(F_1(x),\ldots,F_d(x))<\overline{F}(F_1(x)+\epsilon_1,\ldots,F_d(x)+\epsilon_d)$, $\forall\epsilon_i\in[0,1]$ small enough and $\exists i$: $\epsilon_i>0$.

a convenient and equivalent representation of the dissimilarity criterion, which consists in analyzing that the dispersion of distributions $(G_1(\overline{G}^{-1}(p)), G_2(\overline{G}^{-1}(p)))$ is smaller than that between distributions $(F_1(\overline{F}^{-1}(p)), F_2(\overline{F}^{-1}(p)))$ (measured on the vertical axis) at any proportion p (on the horizontal axis).

The latter representation is useful for a variety of reasons. First, with two distributions, the dissimilarity criterion can be readily compared with gap curve dominance conditions, introduced by Andreoli et al. (2019) in the context of robust unfair inequality analysis. The gap curve dominance criterion is based on quantile gaps at any p. In the left-hand side panel of figure 2, relevant quantiles in each pair of distributions are identified by crosses. Evaluations in gap curve dominance are subject to the cardinalization of the scale, whereas any monotonic increasing transformation of distributions F_1, F_2 and G_1, G_2 does not modify the right-hand side of the figure.

Second, the criterion in (1) can be generalized to the multi-groups setting with d distributions. We conclude that $(\tilde{G}_1, \ldots, \tilde{G}_d)$ displays less dissimilarity than $(\tilde{F}_1, \ldots, \tilde{F}_d)$ whenever the heterogeneity across distributions $(G_1(\overline{G}^{-1}(p)), \ldots, G_d(\overline{G}^{-1}(p))$ is smaller than that of distributions $(F_1(\overline{F}^{-1}(p)), \ldots, F_d(\overline{F}^{-1}(p))$ for any $p \in [0, 1]$, where $\overline{F}(x) = \sum_{i=1}^d F_i(x)$ and $\overline{G}(x) = \sum_{i=1}^d F_i(x)$ whereas in general $\overline{F}(x) \neq \overline{G}(x)$ for at least some x. Visually, each of the d distribution can be represented in the space on the right-hand side panel of figure 2, and dissimilarity assessed by looking at compositional differences at any proportion p. In this paper, we characterize on normative grounds both the choice of the average group distribution and the adoption of a robust criterion for evaluating heterogeneity in distributions that is related to *uniform majorization.*

Third, the dissimilarity criterion for population distributions F_i can be readily implemented by knowledge of the empirical counterparts $\hat{F}_i$ estimated from random samples of the underlying populations. Owing to the linearity of the cdf estimators and of $\overline{F}$, the law of large numbers allows to conclude that $\hat{F}_i(\hat{\overline{F}}^{-1}(p))$ converges in probability to $F_i(\overline{F}^{-1}(p))$

for any p and for every i.

Lastly, the dissimilarity criterion can be extended to distributions defined over discrete, ordered outcomes, cardinal or not (for instance education or standardized test scores achievements). In this situation, the graph of the distributions $F_1, \ldots, F_d$ would be only right continuous, since the underlying distribution forms probability masses over a finite number of realizations. Under an arbitrary assumption about the continuity of the average population distribution $\overline{F}$ across realizations, the graphs of $F_i(\overline{F}^{-1}(p))$ is well defined across all the domain of p. Hence, the graph on the right-hand side panel of figure 2 still provides a valid representation of the test.

In the following section, we provide an empirical setting for analyzing dissimilarity on the basis of discrete distributions of outcomes that are ordered. Whether these outcomes are cardinal or not is irrelevant for performing dissimilarity assessments, insofar ordinal information is preserved.

2.2 Notation

We compare *distribution matrices* of size $d \times n$, depicting sets of distributions (indexed by rows) of $d \geq 1$ *groups* across $n \geq 2$ disjoint non-ordered *categories* (indexed by columns). We develop dissimilarity comparisons of distribution matrices with a fixed number d of groups and a variable number of classes. These matrices are collected in the set

$$\mathcal{M}_d := \left\{ \mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_{n_A}) : \ \mathbf{a}_j \in [0,1]^d, \sum_{j=1}^{n_A} a_{ij} = 1 \ \forall i, \ for \ n_A \geq 2 \right\},$$

where a_{ij} is the proportion of group i observed in class j. The column vector $\mathbf{a}_j$ collects the proportions of all groups in class j. The distribution matrices in $\mathcal{M}_d$ are hence *row stochastic*, meaning that matrix $\mathbf{A} \in \mathcal{M}_d$ represents a collection of d elements of the unit simplex Δ^{n_A}. We interpret the rows of $\mathbf{A}$ as distributions of groups frequencies.

The *cumulative distribution matrix* $\overrightarrow{\mathbf{A}} \in \mathbb{R}_+^{d,n_A}$ is constructed by sequentially cumulating the elements of the classes of $\mathbf{A}$, so that $\overrightarrow{\mathbf{a}}_k := \sum_{j=1}^k \mathbf{a}_j$. Figure 1 reports information on the empirical cdf of 32 groups defined over 20 classes, which can be organized into 32×20 matrices $\overrightarrow{\mathbf{C}}$ (for the control group) and $\overrightarrow{\mathbf{T}}$ (for the treatment group).We use j and k to denote classes of a distribution matrix.

The following numerical example, which we recurrently use in the paper, represents the distribution matrix $\mathbf{A} \in \mathcal{M}_2$ and its cumulation $\overrightarrow{\mathbf{A}}$:

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.1 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0 & 0.5 \end{pmatrix} \text{ and } \overrightarrow{\mathbf{A}} = \begin{pmatrix} 0.4 & 0.5 & 0.8 & 1 \\ 0.1 & 0.5 & 0.5 & 1 \end{pmatrix}. \tag{2}$$

Here, for instance, $a_{13} = 0.3$ indicates that the frequency of group one in class three is 30%, while $\overrightarrow{a}_{13} = 0.8$ indicates that the cumulative frequency of group one achieving realizations smaller or equal than those in class three is 80%.

Furthermore, let group h dominates group ℓ whenever $\overrightarrow{a}_{hj} \leq \overrightarrow{a}_{\ell j}$ for all classes $j = 1, ..., n$, with a strict inequality ($<$) holding for at least a class.[9] That is, ℓ is over-represented at the bottom of the realizations domain compared to h. This makes ℓ the disadvantaged group.

We follow the convention of using boldface letters to indicate column vectors, so that $\mathbf{i}_j$ is a column vector corresponding to column j of an identity matrix $\mathbf{I}_n$ of size n, $\mathbf{1}_n = \sum_j \mathbf{i}_j$ is the column vector with n entries all equal to 1 and similarly $\mathbf{0}_n := (0, \ldots, 0)'$, where the superscript denotes transposition. We denote $\mathcal{P}_n$ the set of $n \times n$ permutation matrices.

2.3 The empirical dissimilarity criterion

Let $p \in [0, 1]$ indicate a proportion of the *average of the cumulative distributions across groups* associated to a distribution matrix $\mathbf{A} \in \mathcal{M}_d$, with $p_j = \frac{1}{d}\mathbf{1}_d^t \cdot \overrightarrow{\mathbf{a}}_j$ for $j = 1, \ldots, n$. Let $\overrightarrow{a}_i(p)$ for each group i be the onto function on $[0, 1]$ such that $\overrightarrow{a}_i(p_j) = \overrightarrow{a}_{ij}$ for any class

[9] A strong version of stochastic dominance requires instead that $\overrightarrow{a}_{hj} < \overrightarrow{a}_{\ell j}$ holds for all classes $j = 1, \ldots, n-1$.

j of the distribution matrix, with $\overrightarrow{a}_i(0) = 0$ and $\overrightarrow{a}_i(p_n) = 1$. All in-between proportions $p \in [p_{j-1}, p_j]$, $j = 1, \ldots, n$, are a solution to the functional equation $p = \frac{1}{d}\sum_i \overrightarrow{a}_i(p)$ for every group i. When $p = p_j$ for any j, the solution to the equation is $\overrightarrow{a}_{ij}$. When $p \in [p_j, p_{j+1}]$, the solution is obtained by interpolating linearly outcomes $\overrightarrow{a}_{ij}$ and $\overrightarrow{a}_{ij+1}$ with a parameter common to all groups.[10]

Plotting the solutions to the equation across levels $p \in [0, 1]$ gives a piecewise linear graph on $[0, 1]$. The underlying functional representation of this graph is $\overrightarrow{a}_i(p)$ for every group i. In compact notation we write $\overrightarrow{\mathbf{a}}(p) = (\overrightarrow{a}_1(p), \ldots, \overrightarrow{a}_d(p))^t$. In Figure 3, we report in gray the cumulative distribution functions collected in vector $\overrightarrow{\mathbf{c}}(p)$ for the control municipalities and $\overrightarrow{\mathbf{t}}(p)$ for the treatment municipalities. The dissimilarity criterion we apply in this empirical setting to evaluate whether unfair inequality has been reduced by effect of the Swedish education reform consists in assessing whether $\overrightarrow{\mathbf{t}}(p) \preccurlyeq^U \overrightarrow{\mathbf{c}}(p)$ $p \in [0, 1]$.

On more general grounds, the dissimilarity empirical criterion is informative about dissimilarity in the population. If $\mathbf{A}$ is estimated from a random sample from $F_1, \ldots, F_d$ then the law of large numbers leads to conclude that:

$$\overrightarrow{\mathbf{a}}(p_j) \longrightarrow^p \left(F_1(\overline{F}^{-1}(p_j)), \ldots, F_d(\overline{F}^{-1}(p_j))\right), \; j = 1, \ldots, n.$$

The convergence takes results from linearity of the operators involved. Convergence extends to all $p \in [0, 1]$ when XXXX

We characterize axiomatically the dissimilarity ordering.

[10] To identify the interpolating parameter, let define $\overrightarrow{\mathbf{a}}(p) = (\overrightarrow{a}_1(p), \ldots, \overrightarrow{a}_d(p))^t$ such that $\overrightarrow{\mathbf{a}}(p_j) = \overrightarrow{\mathbf{a}}_j$. For $p \in [p_{j-1}, p_j]$ the functions satisfy $\overrightarrow{\mathbf{a}}(p) = \overrightarrow{\mathbf{a}}_{j-1} + \lambda \mathbf{a}_j$ with $\lambda \in [0, 1]$ implicitly defined by $p = p_{j-1} + \lambda \frac{1}{d}\overline{a}_j = (1 - \lambda)p_{j-1} + \lambda p_j$. The scalar $\lambda = (p - p_{j-1})/(p_j - p_{j-1})$ is the interpolating parameter that we use.

3 Characterization

3.1 Dissimilarity orders

The cases of perfect similarity and maximal dissimilarity can be formalized in matrix notation. A *perfect similarity* matrix $\mathbf{S}$ represents a situation in which the distributions of all groups coincide across classes and can be represented by the same row vector $\mathbf{s}' \in \Delta^n$. A *maximal dissimilarity* matrix $\mathbf{D}$ represents instead situations where each class is occupied at most by one group *and* each group occupies separate classes. In compact notation:

$$\mathbf{S} := \begin{pmatrix} \mathbf{s}' \\ \vdots \\ \mathbf{s}' \end{pmatrix} \quad \text{and} \quad \mathbf{D} := \begin{pmatrix} \mathbf{d}_1' & \dots & \mathbf{0}_{n_d}' \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{n_1}' & \dots & \mathbf{d}_d' \end{pmatrix}. \tag{3}$$

In the first case, $\mathbf{S}$, all groups are equally represented with the same intensity in each class. Conversely, in the second case, $\mathbf{D}$, it is possible to forecast the group occupying each class. This is the case because the distributions of the groups $\mathbf{d}_1' \in \Delta^{n_1}, \dots, \mathbf{d}_d' \in \Delta^{n_d}$ do not overlap across classes. [11] In this situation, not only groups are ordered by stochastic dominance (a condition that describes violations of equality of opportunity, as in Roemer (1998), Lefranc et al. (2009) and Andreoli et al. (2019)), but the highest realization achieved by any of the groups is always smaller than the lowest realization achieved by any other group which is better off.

This paper investigates the possibility of ordering distribution matrices according to the dissimilarity they display. A *dissimilarity ordering* is a complete and transitive binary relation $\preccurlyeq$ on the set $\mathcal{M}_d$ with symmetric part $\sim$, that ranks $\mathbf{B} \preccurlyeq \mathbf{A}$ whenever $\mathbf{B}$ is *at most as dissimilar as* $\mathbf{A}$.[12] Given $\mathbf{A} \in \mathcal{M}_d$, any dissimilarity ordering should rank $\mathbf{S} \preccurlyeq \mathbf{A} \preccurlyeq \mathbf{D}$

[11] This condition represents the case in which group identity and realizations display the highest degree of *connectivity*, a condition regarded to in Gini (1914) and subsequent literature (see Bertino, Drago, Landenna, Leti and Marasini 1987) as the maximal dissimilarity scenario.

[12] For any $\mathbf{A},\ \mathbf{B},\ \mathbf{C} \in \mathcal{M}_d$ the relation $\preccurlyeq$ is *transitive* if $\mathbf{C} \preccurlyeq \mathbf{B}$ and $\mathbf{B} \preccurlyeq \mathbf{A}$ then $\mathbf{C} \preccurlyeq \mathbf{A}$ and *complete*

for any perfect similarity matrix $\mathbf{S}$ and for any maximal dissimilarity matrix $\mathbf{D}$. There are infinitely many matrices that can be represented as $\mathbf{S}$ and $\mathbf{D}$ in (3). They are all regarded as equivalent representations of perfect similarity or of maximal dissimilarity, the focus being on differences across group distributions and not on the degree of heterogeneity in the distribution of each group across realizations. The condition $d \leq n$ is, nevertheless, necessary for $\mathbf{D}$ to exist. If $\mathbf{A}$ is such that $d > n$, then it can display some dissimilarity (as in the examples we use in the introduction), but not maximal dissimilarity.

3.2 Axioms

We outline axioms that establish the behavior of any relevant dissimilarity ordering when data are transformed in ways that unambiguously preserve or reduce dissimilarity. We study first the exchange transformations within a restricted domain of matrices where *all groups are ordered* according to *stochastic dominance*. Unless the distributions of groups h and ℓ coincide, if h dominates ℓ then there must exist a class k where $\overrightarrow{a}_{hk} < \overrightarrow{a}_{\ell k}$ such that $a_{\ell k} > 0$. When this is the case, dissimilarity can be reduced through an exchange operation, consisting in an upward movement of a *small enough* proportion $\varepsilon > 0$ of group ℓ, over-represented at the bottom of the realizations domain, from class k to any other class $k' > k$ associated with better realizations. This change is counterbalanced by a downward movement of an equal proportion ε of group h from class k' to k. By "small enough" we mean that, after the exchange, the dominance relations between *all* groups (and, notably, between h and ℓ) are preserved. This bears two consequences. First, the amount ε exchanged should, at most, compensate the disadvantage of ℓ in every class, but it should never swap the ranking of ℓ and h. Second, the transfer should not induce a re-ranking of ℓ and h with respect to the other groups. These conditions are made explicit in the definition of axiom E.

if either $\mathbf{A} \preccurlyeq \mathbf{B}$ or $\mathbf{B} \preccurlyeq \mathbf{A}$ or both, in which case $\mathbf{B} \sim \mathbf{A}$.

Axiom *E (Exchange)* *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B = n$ *where group* h *dominates group* ℓ *and* $k' > k$, *if* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by an exchange transformation such that (i)* $b_{hk} = a_{hk} + \varepsilon$ *and* $b_{hk'} = a_{hk'} - \varepsilon$, *(ii)* $b_{\ell k} = a_{\ell k} - \varepsilon$ *and* $b_{\ell k'} = a_{\ell k'} + \varepsilon$, *(iii)* $b_{ij} = a_{ij}$ *in all other cases, (iv)* $\varepsilon > 0$ *so that if* $\overrightarrow{a}_{ij} \leq \overrightarrow{a}_{i'j}$ *then* $\overrightarrow{b}_{ij} \leq \overrightarrow{b}_{i'j}$ *for all groups* $i \neq i'$ *and for all classes* j, *then* $\mathbf{B} \preccurlyeq \mathbf{A}$.

Only a subset of matrices in $\mathcal{M}_d$ can be compared (i.e., transformed one into the other) through exchange operations. We identify this class through the property of *ordinal comparability.*

Definition 1 (Ordinal comparability) *The matrices* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *are ordinal comparable if (i)* $\mathbf{e}_d^t \cdot \mathbf{A} = \mathbf{e}_d^t \cdot \mathbf{B}$, *(ii) all groups are ordered according to stochastic dominance in* $\mathbf{A}$ *and* $\mathbf{B}$, *and (iii) the order of the groups is the same in* $\mathbf{A}$ *and* $\mathbf{B}$.

Condition *(i)* above has two implications. First, ordinal comparable matrices must have the same size, that is $n_A = n_B = n$. Second, the *average distribution across groups*, denoted for matrix $\mathbf{A}$ by the n-dimensional row vector $\frac{1}{d}\mathbf{e}_d^t \cdot \mathbf{A}$, should coincide for matrices $\mathbf{A}$ and $\mathbf{B}$, that is $\frac{1}{d}\mathbf{e}_d^t \mathbf{a}_j = \frac{1}{d}\mathbf{e}_d^t \mathbf{b}_j$ for any class j.

Every exchange operation preserves the number of the classes of a distribution matrix along with its average distribution across groups, and does not produce re-ranking of the distributions of the groups, as required by conditions *(ii)* and *(iii)* in the definition of ordinal comparability.

Consider first the possibility of relaxing condition *(i)* in the definition of ordinal comparability (Definition 1). The axioms *IEC* and *SC* introduce operations that reshape the number and size of the classes of a distribution matrix without affecting dissimilarity.

Axiom *IEC (Independence from Empty Classes)* *For any* $\mathbf{A},\ \mathbf{B},\ \mathbf{C},\ \mathbf{D} \in \mathcal{M}_d$ *and* $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$, *if* $\mathbf{B} = (\mathbf{A}_1, \mathbf{0}_d, \mathbf{A}_2)$, $\mathbf{C} = (\mathbf{0}_d, \mathbf{A})$, $\mathbf{D} = (\mathbf{A}, \mathbf{0}_d)$ *then* $\mathbf{B} \sim \mathbf{C} \sim \mathbf{D} \sim \mathbf{A}$.

The *IEC* axiom places the emphasis on dissimilarity originated from non-empty columns of a distribution matrix. If $\mathbf{A}$ and $\mathbf{B}$ differ only because of $|n_A - n_B|$ empty classes in one of the two matrices, then the dissimilarity in $\mathbf{A}$ should be regarded to as an equivalent representation of that in $\mathbf{B}$. Adding or eliminating an empty class changes the number of classes without affecting the average distribution across groups.

The second transformation increases the number of classes by *splitting proportionally* (the groups densities in) a class into two new classes. This transformation requires to replicate one column of a distribution matrix and then to scale the entries of the original and of the replicated columns by the splitting coefficients $\beta \in (0,1)$ and $1-\beta$, respectively. This operation guarantees that the resulting distribution matrix is row stochastic and that the degree of proportionality of the groups frequencies in the new columns coincides with that in the original column. In the schooling segregation example, splitting a school would require to randomly allocate its students population into two smaller institutes, so that ethnic proportions in the two new institutes are not altered.

Axiom *SC* *(Independence from Split of Classes)* *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *with* $n_B = n_A + 1$*, if* $\exists\, j$ *such that* $\mathbf{b}_j = \beta \mathbf{a}_j$ *and* $\mathbf{b}_{j+1} = (1-\beta)\mathbf{a}_j$ *with* $\beta \in (0,1)$*, while* $\mathbf{b}_k = \mathbf{a}_k$ $\forall k < j$ *and* $\mathbf{b}_{k+1} = \mathbf{a}_k$ $\forall k > j$*, then* $\mathbf{B} \sim \mathbf{A}$.

The *SC* axiom highlights that dissimilarity arises from the disproportionality of the groups composition in some classes. A split transformation increases the number of classes and modifies the shape of a distribution matrix, but it does not alter the proportionality of the groups. For this reason, it is regarded to as dissimilarity preserving. A finite sequence of split of classes and insertion/elimination of empty classes can then be used to extend the degree of comparability in terms of axiom E to matrices that satisfy conditions *(ii)* and *(iii)* in Definition 1, but not condition *(i)*. An example clarifies this point.

Consider the distribution matrix $\mathbf{A}'$, obtained from $\mathbf{A}$ by merging, element by element, the classes two and three as follows:

$$\mathbf{A}' = \begin{pmatrix} 0.4 & 0 & 0.4 & 0.2 \\ 0.1 & 0 & 0.4 & 0.5 \end{pmatrix} \text{ and } \overrightarrow{\mathbf{A}}' = \begin{pmatrix} 0.4 & 0.4 & 0.8 & 1 \\ 0.1 & 0.1 & 0.5 & 1 \end{pmatrix}. \tag{4}$$

In both matrices $\mathbf{A}$ in (2) and $\mathbf{A}'$ the groups are ranked according to stochastic dominance, but the average distributions across groups differ. Hence, condition *(i)* in Definition 1 does not hold and the two matrices cannot be compared exclusively on the basis of exchange operations. Matrix $\mathbf{A}'$, however, can be transformed by splitting its third class according to proportions 5/8 and 3/8, and then eliminating its second class, which is empty. These operations reshape matrix $\mathbf{A}'$ size and the corresponding average distribution across groups, so that the new matrix is ordinal comparable to $\mathbf{A}$, which can be now obtained by exchanging a proportion $\varepsilon = 0.15$ of the two groups as follows:

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.25 - \varepsilon & 0.15 + \varepsilon & 0.2 \\ 0.1 & 0.25 + \varepsilon & 0.15 - \varepsilon & 0.5 \end{pmatrix}. \tag{5}$$

All dissimilarity orderings consistent with axioms *IEC*, *SC* and *E* agree that $\mathbf{A} \preccurlyeq \mathbf{A}'$, even though $\mathbf{A}$ and $\mathbf{A}'$ are not ordinal comparable.[13]

If the focus is on the departure from similarity and not on what group dominates the others, then any permutation of the distributions of the groups should not affect dissimilarity. The axiom *IPG* would shift the focus from the label of the groups to their distributions, thus extending comparability to matrices where all groups are ordered by stochastic dominance, but their labels do not coincide across matrices.

Axiom *IPG* (Independence from Permutations of Groups) *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$, *if* $\mathbf{B} = \mathbf{\Pi}_d \cdot \mathbf{A}$ *for a permutation matrix* $\mathbf{\Pi}_d \in \mathcal{P}_d$ *then* $\mathbf{B} \sim \mathbf{A}$.

[13] The example illustrates that dissimilarity evaluations based on the axiom E could be incompatible with the implications of axiom *IPC*. As argued in Section 3.2, in fact, when the classes of $\mathbf{A}$ and of $\mathbf{A}'$ can be permuted one should rather conclude in favor of the opposite ranking $\mathbf{A}' \preccurlyeq \mathbf{A}$, which holds for all orderings consistent with axioms *IPC*, *IEC*, *SC* and *MC*.

One additional implication of the axiom is that the cumulative distributions of the groups should be treated symmetrically in dissimilarity assessments, meaning that both positive and negative gaps between the distributions contribute equally to measured dissimilarity.

The implications of *IPG* can be strengthened by assuming that if the cumulative distributions of two or more groups coincide in a class, then any permutation of the name of these groups from that class onward should lead to a new distribution matrix exhibiting the same degree of dissimilarity. This concept is formalized through the *Interchange of Groups (I)* axiom.

Axiom *I* *(Interchange of Groups)* *For any* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B = n$, *if* $\exists \mathbf{\Pi}_{h,\ell} \in \mathcal{P}_d$ *permuting only groups* h *and* ℓ *whenever* $\overrightarrow{a}_{hk} = \overrightarrow{a}_{\ell k}$, *such that* $\mathbf{B} = (\mathbf{a}_1, ..., \mathbf{a}_k, \mathbf{\Pi}_{h,\ell} \cdot \mathbf{a}_{k+1}, ..., \mathbf{\Pi}_{h,\ell} \cdot \mathbf{a}_{n_A})$, *then* $\mathbf{B} \sim \mathbf{A}$.

According to axiom *I*, if the cumulative distributions of at least two groups coincide in class k of $\mathbf{A}$, then an interchange of these groups from class $k+1$ onward reproduces the effect of a permutation of the labels of these groups. When the distributions of the groups can be ordered according to *strong* stochastic dominance, similarity is clearly violated but axiom *I* cannot apply. When, instead, the distributions of the groups can be ordered by a weaker form of stochastic dominance (or when there is no dominance relation at all) the gaps in cumulative group distributions compensate (or reverse) in at least one class, thus indicating a less clear violation of similarity. In these situations, axiom *I* postulates that cases of weak dominance and non-dominance that could arise after the application of interchange transformations are sources of indifference in terms of dissimilarity.

Consider matrix $\mathbf{A}''$ below, obtained from $\mathbf{A}$ in (2) by interchanging the labels of the groups from class three onward, so that

$$\mathbf{A}'' = \begin{pmatrix} 0.4 & 0.1 & 0 & 0.5 \\ 0.1 & 0.4 & 0.3 & 0.2 \end{pmatrix} \quad \text{with} \quad \overrightarrow{\mathbf{A}}'' = \begin{pmatrix} 0.4 & 0.5 & 0.5 & 1 \\ 0.1 & 0.5 & 0.8 & 1 \end{pmatrix}. \tag{6}$$

The gaps in the cumulative distributions of the groups in $\overrightarrow{\mathbf{A}}''$ and $\overrightarrow{\mathbf{A}}$ are of the same magnitude but of different sign. According to axiom *I*, the two matrices should be regarded as equally dissimilar. The axiom emphasizes dissimilarity that arises from the extent, and not from the sign, of these gaps. It also allows to compare the gaps of the cumulative distributions of the groups in matrix $\overrightarrow{\mathbf{A}}'$ in (4), which always have the same sign, with those in matrix $\overrightarrow{\mathbf{A}}''$, whose sign changes across classes.

Axioms *IEC*, *SC*, *IPG* have been introduced in XXX Andreoli zoli and are maintained in our setting. These axioms validity rests on the fact that outcomes realizations are not cardinal. XXX Andreoli introduced an axiom setting *independence with respect to permutations of the classes*. This axiom is irrelevant when outcomes are ordinal. Finally, XXX Andreoli study the merge of classes operation, which consists in merging, distribution by distribution, proportions of the groups across subsequent classes to generate a new class of larger size. The axioms *MC* posits that any merge of classes operation cannot increase dissimilarity. Axiom *MC* may lead to counterintuitive evaluations when classes are ordered: in fact, a merge operation could level differences in groups frequencies in correspondence of some realizations that could compensate for disadvantages by some of these groups at lower realizations.[14] Exchange operations, instead, are robust to these critics. Axioms *E* and *I* are new and exploit information obtained by cumulating groups frequencies across ordered realizations.

[14] As an example, consider merging classes two and three of $\mathbf{A}$ might have a counterintuitive impact on dissimilarity. This transformation moves 10% of the population of the disadvantaged group 1, along with 40% of the population of group 2, from class two to three, thus leading to the new configuration $\mathbf{A}'$ in (4). Though the transformation improves the situation of both groups, the cumulative distributions in $\overrightarrow{\mathbf{A}}'$ move apart and the gap between the two distributions does not "compensate" anymore in class two.

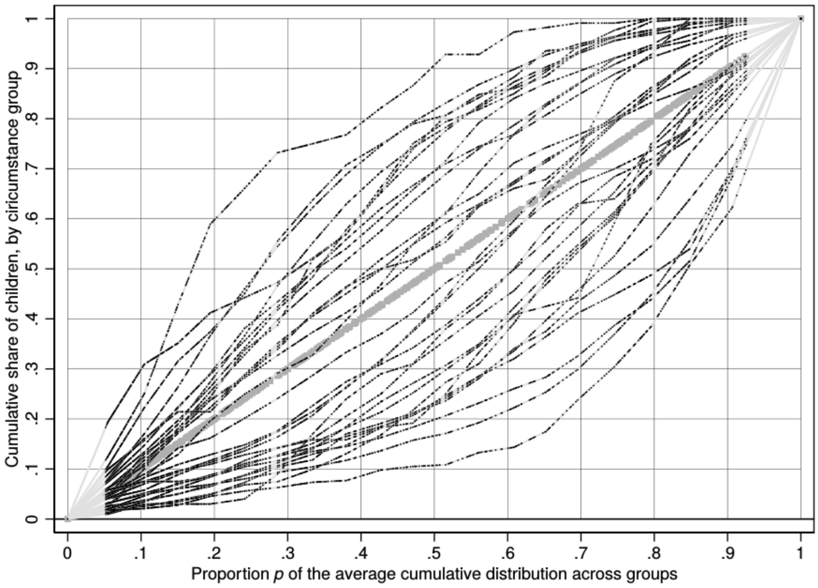
3.3 Dissimilarity indices and majorization conditions

Additional notation. Let $p \in [0,1]$ indicate a proportion of the *average of the cumulative distributions across groups* associated to a distribution matrix $\mathbf{A} \in \mathcal{M}_d$. The proportion associated to cumulative distributions in class j is p_j, with $p_j = \frac{1}{d}\mathbf{e}_d^t \cdot \overrightarrow{\mathbf{a}}_j$. Any proportion $p \in [p_{j-1}, p_j]$ (with $p_0 = 0$) can be obtained by splitting class j of $\mathbf{A}$ into classes j' and $j'+1$ so that, after the split, $p_j = p_{j'+1}$ and $p = p_{j'}$. Every such split operation carries also consequences for the cumulative distribution of each group. Denote $\overrightarrow{a}_i(p)$ the onto function on $[0,1]$ which gives the cumulative proportion of group i that would be observed in class j' if class j were split into classes j' and $j'+1$ as above. There is a function $\overrightarrow{a}_i(p)$ for each group i. The function is such that $\overrightarrow{a}_i(p_j) = \overrightarrow{a}_{ij}$ for any class j of the distribution matrix, with $\overrightarrow{a}_i(0) = 0$ and $\overrightarrow{a}_i(p_n) = 1$. All in-between proportions $p \in [p_{j-1}, p_j]$, $j = 1, \ldots, n$, are a solution to the functional equation $p = \frac{1}{d}\sum_i \overrightarrow{a}_i(p)$ with $\overrightarrow{a}_i(p)$ continuous and piecewise linear on $[0,1]$ for every group i.[15] For ease of exposition, distribution functions are organized in a vector $\overrightarrow{\mathbf{a}}(p) = (\overrightarrow{a}_1(p), \ldots, \overrightarrow{a}_d(p))^t$ at proportion p, with $\overrightarrow{a}_i(p)$ defined as above.

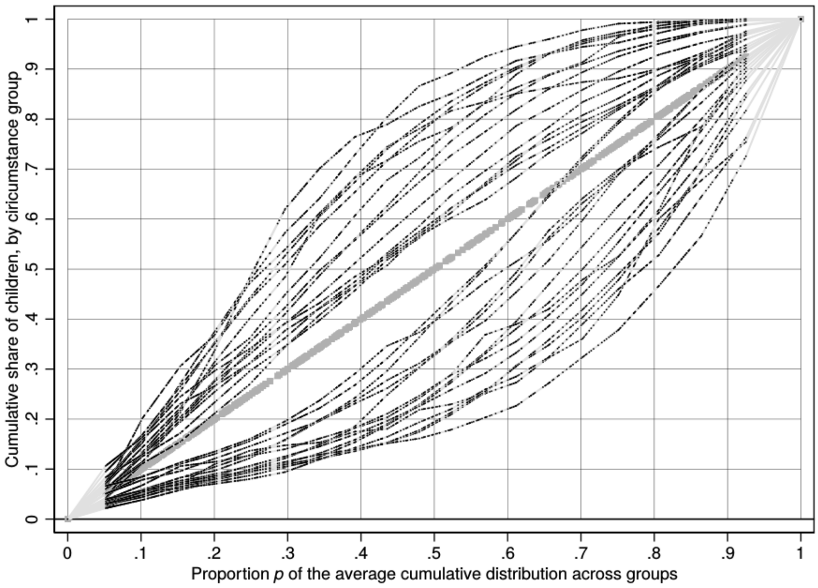
There are two key advantages from representing for each group i its cumulative distribution by the function $\overrightarrow{a}_i(.)$. The first advantage is that any dissimilarity ordering satisfying *SC* and *IEC* considers a distribution matrix $\mathbf{A}$ and the associated functions $\overrightarrow{a}_i(.)$ for all $i = 1, \ldots, d$ as equivalent representations of the data.[16] The second advantage is that every distribution matrix admits a representation through the cumulative distribution functions of the groups. We argue that any two distribution matrices that are not ordinal comparable can always be compared by focusing on the dispersion of the distribution functions

[15] More formally, the vector of functions $\overrightarrow{\mathbf{a}}(p) = (\overrightarrow{a}_1(p), \ldots, \overrightarrow{a}_d(p))^t$ is such that $\overrightarrow{\mathbf{a}}(p_j) = \overrightarrow{\mathbf{a}}_j$ for every class $j = 1, \ldots, n_A$. For $p \in [p_{j-1}, p_j]$ the function also satisfies $\overrightarrow{\mathbf{a}}(p) = \overrightarrow{\mathbf{a}}_{j-1} + \lambda \mathbf{a}_j$ with $\lambda \in [0,1]$ implicitly defined by $p = p_{j-1} + \lambda \frac{1}{d}\overline{a}_j = (1-\lambda)p_{j-1} + \lambda p_j$. The scalar $\lambda = (p - p_{j-1})/(p_j - p_{j-1})$ can be interpreted as the split parameter β in the definition of the SC axiom. The function $\mathbf{a}_i(p)$ is then continuous in p and piecewise linear for every group i.

[16] As argued in the previous footnote $\overrightarrow{\mathbf{a}}(p)$ is obtained by applying split transformations to the distribution matrix $\mathbf{A} \in \mathcal{M}_d$. For any p, any sequence of these transformations will lead to the same $\overrightarrow{\mathbf{a}}(p)$. Moreover, adding or eliminating empty classes does not affect $\overrightarrow{\mathbf{a}}(p)$.



(a) Control group



(b) Treatment group

Figure 3: Cumulative groups distributions, by treatment status.

Note: Authors' computations based on Meghir and Palme (2005) data. Groups formed by interacting information on parents education, gender, ability and location. Piecewise linear gray curves are the graphs of the cumulative group distributions $\mathbf{c}(p)$ (control) and $\mathbf{t}(p)$ (treatment). Black dots identify elements of the ordinal comparable matrices $\overrightarrow{\mathbf{C}}^*$ and $\overrightarrow{\mathbf{T}}^*$.

$\overrightarrow{a}_1(p), \ldots, \overrightarrow{a}_d(p)$ at every proportion $p \in [0, 1]$.

Consider again the example discusses in the introduction. Original data can be organized in distributions matrices $\mathbf{T}$ (for the treatment group) and $\mathbf{C}$ (for the comparison group) is size 32×20. The corresponding distribution functions $\overrightarrow{\mathbf{t}}(p)$ and $\overrightarrow{\mathbf{c}}(p)$ are represented through piecewise gray lines in Figure 3. These figures are obtained by reporting the intercepts of the groups distributions in Figure 1 at any proportion $p \in [0, 1]$ of the average population distribution, also represented on the same figure (by gray square symbols). The dissimilarity criteria we analyze allow to conclude about the policy effect by confronting matrices $\mathbf{T}$ and $\mathbf{C}$ on the basis of their representations $\overrightarrow{\mathbf{t}}(p)$ and $\overrightarrow{\mathbf{c}}(p)$.

Majorization A first concern of this paper is to establish a majorization condition which represent the notion of robust dissimilarity evaluations outlined above. Consider $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ with cumulative groups distributions $\overrightarrow{\mathbf{a}}(p)$ and $\overrightarrow{\mathbf{b}}(p)$ respectively. The criterion of dissimilarity reduction that we analyze consists in evaluating the extent of dispersion in cumulative groups distributions in $\overrightarrow{\mathbf{a}}(p)$ and $\overrightarrow{\mathbf{b}}(p)$ at proportions $p \in [0, 1]$ of the average population distribution. $\mathbf{B}$ is as most as dissimilar as $\mathbf{A}$ if the proportions of the groups adding up to the bottom $p100\%$ of the average of the cumulative distributions across groups in $\mathbf{A}$ (i.e $\overrightarrow{\mathbf{a}}(p)$) are unambiguously more dispersed than the corresponding proportions in $\mathbf{B}$ (i.e. $\overrightarrow{\mathbf{b}}(p)$), for any $p \in [0, 1]$. It is standard in the literature on economic inequality to represent the relation "unambiguously more dispersed than" by resorting on uniform majorization conditions for $d-$dimensional vectors (Kolm 1969, Cowell 2000, Marshall et al. 2011, Andreoli and Zoli 2020), denoted $\preccurlyeq^U$.[17] The robust dissimilarity criterion is a rather demanding condition: it requires to verify that uniform dominance holds at any percentile p of the average population distribution, that is $\overrightarrow{\mathbf{b}}^t(p) \preccurlyeq^U \overrightarrow{\mathbf{a}}^t(p)$ for all $p \in [0, 1]$.

[17]For any pair of vectors $\mathbf{a}^t, \mathbf{b}^t \in \mathcal{M}_1$ with $\mathbf{a}^t \cdot \mathbf{e}_d = \mathbf{b}^t \cdot \mathbf{e}_d$, $\mathbf{b}$ majorizes $\mathbf{a}$, denoted $\mathbf{b} \preccurlyeq^U \mathbf{a}$, if and only if $\mathbf{b} \in conv\{\Pi_n \cdot \mathbf{a} : |\Pi_n \in \mathcal{P}_n\}$. Following Marshall et al. (2011), theorems 1.A.3 and 2.B.2, this condition is necessary and sufficient to guarantee that any element of $\mathbf{b}$ can be obtained through a sequence of fundamental inequality reducing (Pigou-Dalton) transfers applied to elements of $\mathbf{a}$.

Indices. A second concern of the paper is about measurement of dissimilarity. We investigate the possibility of measuring dissimilarity by a linear rank-dependent evaluation function $D_w(\mathbf{A})$, which is an average, taken over all proportions p, of the inequality displayed by vectors $\overrightarrow{\mathbf{a}}(p)$. Inequality is measured itself as a weighted average, were realizations $\overrightarrow{a}_i(p)$ are weighted by the function $w_i(p)$. This function is non-decreasing in i at any p and it is assumed to be bounded and continuous in p almost everywhere. The set of all weighting functions satisfying these properties is denoted $\mathcal{W}$. For any $w \in \mathcal{W}$ let

$$D_w(\mathbf{A}) := \int_0^1 \sum_{i=1}^{d} w_i(p) \overrightarrow{a}_{(i)}(p) dp, \tag{7}$$

where $\overrightarrow{a}_{(i)}(p)$ is the i-th smaller element of vector $\overrightarrow{\mathbf{a}}(p)$. The index can be interpreted as the average degree of dispersion of the cumulative distributions of the groups. The shape of the weighting function $w_i(p)$ allows to address the extent of sensitivity of the index to heterogeneity in groups composition at any proportion p. All weighting functions are restricted so that $\sum_i w_i(p) = 0$ for all p, which guarantee to focus on distributional concerns. Under these constraints, any exchange transformation of the data reduces the dispersions in the groups cumulative distributions, hence dissimilarity.[18] Furthermore, the index can be normalized to 0 when perfect similarity is reached.

Yaari (1987) and Weymark (1981) have outlined a particular parametric class of weighting functions belonging to $\mathcal{W}$, denoted the single-parameter *S-Gini* weights, which generalize the Gini inequality index.[19] XXX AAberghe et al XXX have derived different generalizations of the rank-dependent measures of inequality which also include the Gini index a special case. All these weighting schemes can be used to construct parametric measures of dissimilarity. A robust dissimilarity evaluation requires agreement on $D_w(\mathbf{B}) \leq D_w(\mathbf{A})$

[18] Every exchange transformation originates a rank-preserving progressive transfer (Fields and Fei 1978) of *cumulated* frequencies of these groups in at least one class. Every such transfer implies a reduction in the heterogeneity of groups cumulative distributions.

[19] Following notation in Maccheroni, Muliere and Zoli (2005), the discrete counterpart of the S-Gini weights is obtained by setting $w_i(p) = \frac{1}{p}\left(1 - ((1 - \frac{i-1}{d})^k - (1 - \frac{i}{d})^k)\right)$ for k a positive integer. When $k = 2$, the weights coincide with those of the Gini inequality index.

for *all* weighting schemes $w \in \mathcal{W}$.

3.4 Main result and discussion

We characterize the dissimilarity partial orders induced by the intersection of the dissimilarity orderings (Donaldson and Weymark 1998) satisfying desirable properties. The following theorem shows equivalences between the dissimilarity partial order characterized by the dissimilarity axioms, the dissimilarity dominance criterion based on uniform majorization, and the representation based on rank dependent dissimilarity measures. The three conditions establish robust and equivalent representation of the dissimilarity partial order. Nonetheless, comparisons based on these criteria cannot be empirically tested. The theorem offers an equivalent implementable condition which is empirically tractable.

Theorem 1 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *the following statements are equivalent:*

(i) $\mathbf{B} \ \preccurlyeq\ \mathbf{A}$ *for every ordering* $\preccurlyeq$ *satisfying axioms E, SC, IEC, IPG and I.*

(ii) $D_w(\mathbf{B}) \ \leq\ D_w(\mathbf{A})$ *for all* $w \in \mathcal{W}$*.*

(iii) $\overrightarrow{\mathbf{b}}^t(p) \preccurlyeq^D \overrightarrow{\mathbf{a}}^t(p)$ *for all* $p \in [0,1]$*.*

(iv) There exist $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ *ordinal comparable that are obtained from* $\mathbf{A}$ *and* $\mathbf{B}$ *respectively through elimination of empty classes, split of classes, interchanges and permutation of groups operations, such that* $\sum_{i=1}^h \overrightarrow{b}^*_{(i)j} \ \geq\ \sum_{i=1}^h \overrightarrow{a}^*_{(i)j}$ *for all* $h = 1, \ldots, d$*,* $j = 1, \ldots, n^*$*.*

Exchange transformations are related to correlation-decreasing transfers, (see Epstein and Tanny 1980, Tchen 1980, Atkinson and Bourguignon 1982). E combines distance in cdfs and correlation reduction as two equivalent perspectives. But when combined with I, the focus shift from correlation to distance.

The criterion displays similarities with *orthant tests* investigated in the stochastic orders literature (see Ch. 6.G in Shaked and Shanthikumar 2006). Tchen (1980) has proposed an orthant test to analyze concordance in matrices with fixed margins, where the order of the groups is fixed exogenously but group distributions are not necessarily ordered by stochastic dominance (the underlying welfare order has been developed in Atkinson and Bourguignon 1982). Differently from Tchen's result, claim *(ii)* holds when groups are endogenously ordered by stochastic dominance relations. Hence, we face more constraints than Tchen in showing that sequential majorization can be decomposed into a series of exchange transformations that preserve the order of the groups, which is a reasonable and normatively appealing feature in dissimilarity analysis (but not necessarily in other situations). We develop on this observation in the proof of Lemma 2 in Appendix A.1.

Claim *(iv)* can also be related to the orthant order introduced by Meyer and Strulovici (2013) to assess supermodularity in matrices with different class margins. They decompose the dominance condition implied by the orthant test into operations that are different (and weaker) than the exchange transformations, but that are meaningful to characterize supermodular stochastic orderings of interdependence between the rows of a distribution matrix.

Axioms *SC*, *IEC* and *I* extend the validity of the dissimilarity model to the class $\mathcal{M}_d$ of distribution matrices. The main result of this section shows that for any pair of distribution matrices $\mathbf{A}$ and $\mathbf{B}$ that are not ordinal comparable, there always exists a pair of matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ that are ordinal comparable and such that the groups in $\mathbf{A}^*$ and $\mathbf{B}^*$ display the same cumulative distribution functions observed in $\mathbf{A}$ and $\mathbf{B}$, respectively. Then, $\mathbf{A}^*$ and $\mathbf{B}^*$ can be obtained from $\mathbf{A}$ and $\mathbf{B}$ through operations of splits of classes, insertion/elimination of empty classes and groups interchanges. Every dissimilarity ordering consistent with axioms *IEC*, *SC* and *I* concludes that $\mathbf{A}^* \sim \mathbf{A}$ and $\mathbf{B}^* \sim \mathbf{B}$.

There are of many equivalent pairs of distribution matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ that can be obtained in this way. For any such pair, $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ for every dissimilarity ordering consistent with axiom *E*. By transitivity, $\mathbf{B} \preccurlyeq \mathbf{A}$ must hold for the subset of orderings satisfying axioms *IEC*, *SC*, *I and* axiom *E*. Claim *iv)* shows that the dissimilarity partial orders are implemented by testing sequentially *Lorenz dominance*[20] on each class $j = 1, \ldots, n^*$ of ordinal equivalent matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ that are equivalent, in terms of dissimilarity, to $\mathbf{A}$ and $\mathbf{B}$. In Appendix B we discuss an algorithm that allows to retrieve ordinal comparable matrices making from any pair of data matrices.

4 Related orders

4.1 Mobility and equality of opportunity

An increasingly popular notion of inequality, alternative to inequality of outcomes, is that of *inequality of opportunity* (Roemer 2012). According to this theory, outcomes are generated by individual effort (gathering all dimensions upon which people have full control and responsibility), by circumstances (such as the background of origin), and by the interaction of these two. Inequality of opportunity criteria account for the implications of the unequal distribution of circumstances on the distribution of some relevant outcome. In the context of income opportunities, some authors[21] have suggested to use as benchmark the counterfactual fair income distribution (representing the income distribution that would have occurred if the implications of the circumstances on income were eliminated). Inequality of opportunity stems from the dissimilarity in the distribution of the actual income shares across the entire population and the distribution of the counterfactual (fair) income shares in the same population. Theorem 1 hints on the possibility of using Zonotopes inclusion to

[20] Recall that for any pair of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d_+$, $\mathbf{b}$ Lorenz dominates $\mathbf{a}$ if and only if $\sum_{i=1}^h b_{(i)} \geq \sum_{i=1}^h a_{(i)}$ $\forall h = 1, \ldots, d$, with equality holding for $h = d$ (see Marshall et al. 2011).

[21] For a review, see Ramos and Van de gaer (2016).

test inequality of opportunity for income, and provides a consistent measurement framework. Other contributions have stressed the importance of inequality of opportunity in dimensions other than income, such as education, health or wellbeing. In these cases, inequality of opportunity arises when groups, defined by circumstances, are dissimilarly distributed across attainable ordinal realizations. Theorem 1 provides the foundations for its measurement.

Dardanoni (1993) investigates the social welfare ordering induced by sequential majorization in the context of income mobility. Dardanoni explores situations in which groups and classes respectively denote the income levels in the distribution of departure and in that of destination, and hence coincide across distribution matrices. The social welfare functions proposed by Dardanoni could be related to the indices $D_{\mathbf{w}}$.[22] Claim *(??)* replicates the mobility orthant order in Dardanoni (1993) when income levels in the underlying mobility matrix are chosen to represent quantiles of the distributions of departure and of destination. With this specification, in fact, groups and classes are equally and uniformly weighted. Theorem **??** then extends Dardanoni's result by showing that agreement in social welfare evaluations of mobility can always be traced down to the existence of a sequence of elementary exchange transformations. Each exchange transformation reduces dissimilarity across the rows of the mobility matrix by improving the mobility prospects (i.e., shifting probability mass towards higher income quantiles of the distribution of destination) for those individuals starting at the bottom quantiles of the distribution of departure and by deteriorating the mobility prospects of those at the top quantiles in the distribution of departure. Theorem **??** applies as well to mobility assessments in which the distribution of departure and that of destination differ. This occurs, for instance, when the distribution

[22] In Dardanoni (1993) the weights w_{ij} coincide with the product of normative weights, associated with the income levels of departure, and utility evaluations, attached to the income levels in the distribution of destination. The social welfare evaluations are weighted by the equilibrium (ergodic) income distribution probabilities, while in our case the groups (indicating information on the distribution of departure) always receive uniform weights, and the classes (indicating information on the distribution of destination) are weighted according to the actual average distribution across groups.

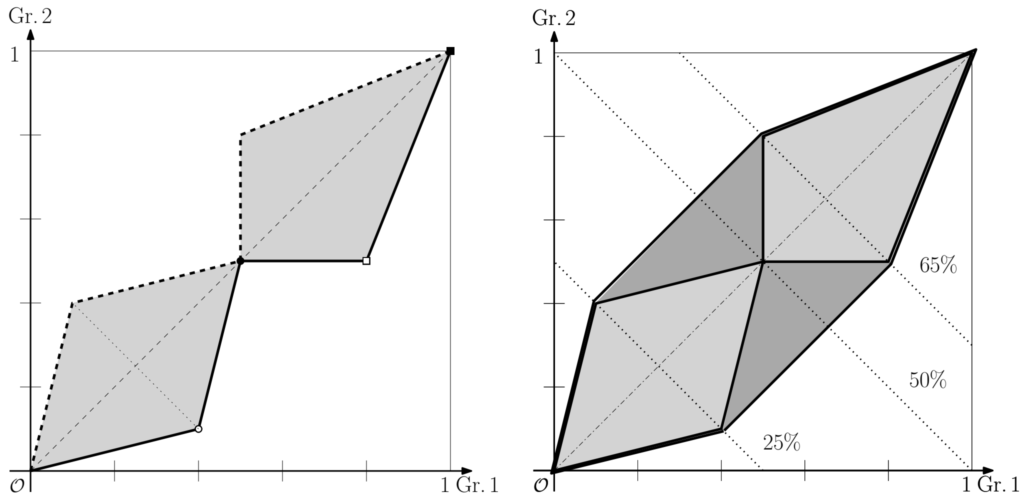
of departure is given in terms of income deciles and that of destination in terms of income centiles. In this situation, the index $D_{\mathbf{w}}$ provides an appropriate metric for social welfare. In more general cases, claim *(iv)* and the criterion in Dardanoni (1993) do not match.

If groups identify percentiles of the parental income distribution, while classes correspond to percentiles of the children income distribution, then any distribution matrix represents an intergenerational *mobility matrix*, where both groups and classes are ordered. If the mobility matrix is *monotone*[23], the dissimilarity model presented in Section 4 provides implementable criteria for assessing changes in mobility across percentiles in the distribution of destination, provided that a perfectly mobile society could be described as one where the distribution of children incomes is independent from that of the parents incomes (as in Shorrocks 1978, Stiglitz 2012, Kanbur and Stiglitz 2016). For any pair of monotone mobility matrices with fixed margins both for rows and columns, the dissimilarity criterion in Theorem 1 coincides with the test of the orthants (Tchen 1980, Dardanoni 1993).[24] If margins differ, the dissimilarity order extends mobility comparisons, and can be tested through Path Polytopes inclusion. This is an important aspect for empirical research, where in many cases the parental income distribution can only be observed with a degree of precision that is smaller (for instance, in deciles) than that of the distribution of children income (for instance, in percentiles). These cases could not be compared within the mobility framework, although they could be compared in terms of the dissimilarity criteria in Theorem 1. If the mobility matrices are non-monotone, the dissimilarity criterion imposes stronger conditions than the traditional mobility test.[25]

[23]In a monotone matrix, the group in row $i+1$ stochastic dominates the group in row i, for any i. If group distributions are suitably rearranged, every pair of ordinal comparable distribution matrices can be interpreted as monotone matrices with given marginals.

[24]After proportions of the groups have been cumulated first by row and then by column, the test of the orthants requires to verify that the entries of the resulting matrix expressing higher mobility are nowhere smaller than the entries of the resulting matrix expressing lower mobility. Given that the matrices are monotone with fixed margins, they satisfy ordinal comparability and the test coincides with the Lorenz dominance criterion in statement *(iv)* of Theorem **??**.

[25]Empirical evidence suggests, however, that monotonicity of mobility matrices is unlikely to be rejected by the data (Dardanoni, Fiorini and Forcina 2012).



(a) Path Polytope of $\mathbf{A}$ in (2)

(b) Path Polytopes inclusion

Figure 4: Path Polytopes of matrices $\mathbf{A}$ (light grey area) and $\mathbf{A}'$ (dark grey area), with matrix $\mathbf{B}$ Monotone Path (dashed line in panel (b)).

4.2 The geometry of dissimilarity

We introduce a geometric criterion for dissimilarity analysis which is based on the *Path Polytope* representation of distribution matrices. The Path Polytope $PP(\mathbf{A}) \subseteq [0,1]^d$ of a matrix $\mathbf{A} \in \mathcal{M}_d$ is the convex hull of the permutations of a *Monotone Path* $MP(\mathbf{A}) \subseteq [0,1]^d$ with respect to the d-dimensional hypercube diagonal. The Monotone Path $MP^*(\mathbf{A})$ is a graphical piecewise-linear arrangement of n_A segments starting from the origin of the positive orthant, and sequentially connecting the points with coordinates given by the columns of $\overrightarrow{\mathbf{A}}$, up to the point with coordinates $\mathbf{e}_d$ (see Ziegler 1995). Formally:

$$MP(\mathbf{A}) \ := \ \left\{ \overrightarrow{\mathbf{a}}(p) : p \in [0,1] \right\}.$$

The order of the vertices of the Monotone Path, denoted $\mathbf{v}_j \in MP^*(\mathbf{A})$ with $j = 0, 1, \ldots, n_A$, coincides with the one of the classes of $\mathbf{A}$, so that $\mathbf{v}_j = \overrightarrow{\mathbf{a}}_j$, $\mathbf{v}_0 = \mathbf{0}_d$ and $\mathbf{v}_{n_A} = \mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{e}_d$. In compact notation,

$$MP(\mathbf{A}) := \left\{ \mathbf{z} \ : \ \mathbf{z} \in conv\left\{ \mathbf{\Pi}_d \cdot \mathbf{p} \right\}, \ \mathbf{\Pi}_d \in \mathcal{P}_d, \ \mathbf{p} \in MP(\mathbf{A}) \right\},$$

where the *conv* operator denotes the convex hull of all permutations of each point $\mathbf{p}$ along the Monotone Path. As the order of $\mathbf{A}$'s classes is given, $MP^*(\mathbf{A})$ is unique and by construction identifies a unique Path Polytope. A distribution matrix displaying some dissimilarity originates a Path Polytope that lies in $PP(\mathbf{D})$ and shares the same reference diagonal $PP(\mathbf{S})$, where $\mathbf{D}$ and $\mathbf{S}$ are, respectively, maximal dissimilarity and the perfect similarity matrices.

Consider the Monotone Path of $\mathbf{A}$ in (2). It is represented by a thick solid line in Figure 4(a). This line connects four points marked with different symbols, their coordinates being given by the values of the cumulative frequencies of the groups in $\overrightarrow{\mathbf{A}}$ classes. In comparisons involving two groups distributions, the Monotone Path identifies for each proportion of group one the corresponding proportion of group two achieving similar or worse realizations. For every point along the Monotone Path, there exists a corresponding point obtained as its permutation. For instance, the point with coordinates $(0.4, 0.1)^t$, marked by a white dot in the figure, defines a symmetric point with coordinates $(0.1, 0.4)^t$. The Path Polytope $PP(\mathbf{A})$ is identified by the grey area in the figure, delimited by the Monotone Path and by its symmetric counterpart (around the diagonal).

Any Path Polytope could represent more than one distribution matrix. In fact, transformations such as splitting classes, insertion/elimination of empty classes and interchanges, if applied to matrix $\mathbf{A}$ do not affect the graph of $PP(\mathbf{A})$.[26] Nonetheless, exchange transformations have consequences for the shape of the Path Polytope. The next corollary shows that the ordering of distribution matrices induced by the inclusion of Path Polytopes provides an equivalent representation of the dissimilarity partial order characterized in Theorem 1.

Corollary 1 *For any pair of matrices* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, $\overrightarrow{\mathbf{b}}^t(p) \preccurlyeq^D \overrightarrow{\mathbf{a}}^t(p)$ *for all* $p \in [0, 1]$ *if*

[26]Every point $\mathbf{p} \in MP(\mathbf{A})$ comprised between the Monotone Path vertices $\mathbf{v}_{j-1}$ and $\mathbf{v}_j$ can be written as $\mathbf{p} = \overrightarrow{\mathbf{a}}_{j-1} + \lambda \mathbf{a}_j$ with $\lambda \in [0, 1]$, which defines a split operation of class j. Furthermore, if class j is empty, then $\mathbf{v}_{j-1} = \mathbf{v}_j$.

and only if $PP(\mathbf{B}) \subseteq PP(\mathbf{A})$.

The Path Polytope inclusion criterion provides a graphical account of the dissimilarity dominance condition. Consider, for instance, the matrices $\mathbf{A}$ in (2) and $\mathbf{A}'$ in (4). As shown in Figure 4(b), $Z^*(\mathbf{A}) \subseteq Z^*(\mathbf{A}')$. The two distribution matrices are not ordinal comparable and, according to claim *(iv)*, sequential majorization has to be verified at every proportion p of the respective average distribution across groups. Every proportion p identifies an isopopulation (hyper)plane intersecting the Path Polytopes. The inclusion test is verified if the intersection of the Path Polytope of the more dissimilar distribution and any isopopulation hyperplane includes[27] the area originated from the intersection of the Path Polytope of the less dissimilar distribution and the same isopopulation hyperplane. For instance, gaps among the cumulative distributions of groups 1 and 2 in $\mathbf{A}$ perfectly compensate at proportion $p100\% = 50\%$, while a gap in the cumulative distributions of the groups persists in $\mathbf{A}'$. Theorem 1 allows to conclude that $\mathbf{A}'$ displays unambiguously more dissimilarity than $\mathbf{A}$, despite the former being obtained from the latter by a merge transformation. This example further highlights that dissimilarity evaluations of distribution matrices with ordered classes may be inconsistent with, and deserves a different treatment from, dissimilarity evaluations involving categorical outcomes. This distinction is mirrored in the potential inconsistencies in rankings of distribution matrices produced by Path Polytope and Zonotope tests.

Furthermore, the Path Polytope inclusion partial order is the natural multi-group extension of the dominance tests based on non-intersecting interdistributional Lorenz curves (at order zero) proposed by Butler and McDonald (1987) and often employed in discrimination analysis. This is discussed in the following section.

[27]That is, the extremes points of the intersection are more dispersed in the sense of Lorenz. In fact, a cross-section of the Path Polytope originates the so-called “Kolm triangles” representations of the Lorenz curve, as in Kolm (1969).

4.3 Discrimination

Consider now the situation in which groups are distributed over ordered outcomes. The Monotone Path generated by the distributions of two groups coincides with the graph of the *concentration curve* of these distributions (Mahalanobis 1960, Butler and McDonald 1987). When the distribution of one group stochastic dominates the distribution of the other group, the Monotone Path always lies below the unit square diagonal. Its graph delimits a *discrimination curve* (Le Breton et al. 2012). For instance, the groups in matrix $\mathbf{A}$ in (2) are distributed so that group 2 stochastic dominates group 1. Their discrimination curve coincides with the lower boundary of the Path Polytope $Z^*(\mathbf{A})$ in Figure 4(a).

For any pair $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$ in which group 2 stochastic dominates group 1 in both matrices, $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ always indicates that the discrimination curve of $\mathbf{B}$ lies nowhere below the discrimination curve of $\mathbf{A}$. This dominance criterion is related to robust evaluations of statistical discrimination (Gastwirth 1975) and affluence (Dagum 1980). Theorem 1 delivers three contributions on this literature. First, the theorem shows that every robust discrimination assessment based on non-intersecting discrimination curves is supported by the existence of a finite sequence of dissimilarity preserving transformations and of exchange transformations. Hence, discrimination analysis bears similarities with other orderings, but it is inconsistent with the prevailing notions of segregation and inequality (based on merge operations). Second, the theorem points out that the Path Polytopes inclusion order extends the discrimination curve analysis to the multi-group setting.[28] Third, the theorem reveals a simple interpretation of multi-group discrimination: it coincides with the extent of the dispersion in the composition of groups making up the bottom $p100\%$ of the average of the cumulative distributions across groups.

[28]The properties of the associated class of indicators that can be used to assess multi-group discrimination are also provided.

4.4 Distance between distributions

The dissimilarity model for ordered classes can be related to the measurement of the distance between two or more distributions. This approach could be informative when distribution matrices depict distributions of the groups across the realizations of an observable outcome (for instance wages or test score achievements) that is monotonically and increasingly, but not necessarily linearly, related to a latent variable of interest (such as skills) (Athey and Imbens 2006, Bonhomme and Sauder 2011). Theorem 1 would help identifying the effects of an hypothetical educational policy, whose objective could be to reduce the dissimilarity in skills of children experiencing different family backgrounds, by focusing on the dissimilarity in observed wages and test scores achievements of the children treated by the policy. The dissimilarity model allows to single out the policy effect without recurring on ad hoc parametric restrictions on the relation between observable outcomes and the unobservable policy objective.

We maintain in this section that the classes of a distribution matrix delimit a partition of a continuous domain of realizations $\mathcal{X} \subseteq \mathbb{R}$, such that each class k of the matrix $\mathbf{A}$ is connected with an interval of realizations and adjacent classes always indicate adjacent intervals. This is the case, for instance, if $\mathbf{A}$'s classes identify groups frequencies at given wage intervals or test scores achievements. We also assume that these frequencies are uniformly distributed within each interval. Under these conditions, a distribution matrix is an equivalent representation of the histograms of the distributions of the groups.

Let denote group i cumulative distribution by $F_i(x)$. It gives the cumulative proportion of group i members achieving a realization that is smaller than or equal to $x \in \mathcal{X}$. The graph of $F_i(x)$ is piecewise linear, continuous and non-decreasing over $\mathcal{X}$. The kinks in the graph correspond to the upper bounds of the intervals of the partition of $\mathcal{X}$ delimited by $\mathbf{A}$'s classes. Denoting by x_j this realization, then $F_i(x_j) = \overrightarrow{a}_{ij}$.

In a multi-group setting, there are d distributions, denoted $F_1, \ldots, F_d$. The average of the distributions across all groups evaluated at realization x is $\overline{F}(x) = \frac{1}{d}\sum_{i=1}^{d} F_i(x)$. The dissimilarity of $F_1, \ldots, F_d$ can be measured by the vertical distance (according to some metric) between these cumulative distributions evaluated at $\overline{F}^{-1}(p) \in \mathcal{X}$, which is the p-th quantile of $\overline{F}(.)$.

When $d = 2$, a natural metric for the distance is $\left|F_1(\overline{F}^{-1}(p)) - F_2(\overline{F}^{-1}(p))\right|$. An aggregate distance indicator $D(F_1, F_2)$ could then be the average of these differences:

$$D(F_1, F_2) \quad := \quad \int_0^1 \left|F_1(\overline{F}^{-1}(p)) - F_2(\overline{F}^{-1}(p))\right| dp, \tag{8}$$

which is a measure belonging to the class identified in claim *(iii)* of Theorem 1.[29] After a change in variable, the index rewrites:

$$D(F_1, F_2) \quad := \quad \int_{\mathcal{X}} |F_1(x) - F_2(x)|\, d\overline{F}(x). \tag{9}$$

The distance $D(F_1, F_2)$ can be readily compared to measures of distance between distributions of a continuous variable that have been developed in the literature (Shorrocks 1982, Ebert 1984), the most intuitive one being the Manhattan distance $D_M(F_1, F_2) := \int_{\mathcal{X}} |F_1(x) - F_2(x)|\, dx$. Compared to these distances, $D(F_1, F_2)$ is invariant to monotone transformations of the scale of $\mathcal{X}$. This has implications for the policy evaluation example above. In fact, according to $D(F_1, F_2)$, the distance between group-specific skills distributions coincides with the distance between group-specific distributions of wages and of test score achievements. The distance $D(F_1, F_2)$ thus allows to use observable information on wages and test scores to identify the policy effects without assuming linearities of skills effects in the wage and test score functions. On the contrary, alternative distance metrics that aggregate gaps in cumulative distribution functions of the groups based on observable outcomes could mechanically magnify actual skills gaps across groups (if, for

[29] Suppose that distribution F_2 stochastic dominates distribution F_1. In this case, it is sufficient to set $w_2(p) = -1$ and $w_1(p) = 1\ \forall p$ in (7) to obtain $D(F_1, F_2)$.

instance, wages and test-scores are more volatile than skills), thus leading to biased policy evaluations.

The distance $D(F_1, F_2)$ has also other desirable properties. Its upper bound is $D(F_1, F_2) = 1$ when F_1 and F_2 do not overlap (i.e., the Path Polytope coincides with the unit square), a situation identifying maximal dissimilarity. The minimum is $D(F_1, F_2) = 0$ and is achieved whenever $F_1 = F_2$. These bounds follow from the next equivalence:

Corollary 2 *$D(F_1, F_2)$ is proportional to the area of the Path Polytope $PP(\mathbf{A})$.*

For any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$, the condition $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ is therefore sufficient (but not necessary) for $D(F_1^B, F_2^B) \leq D(F_1^A, F_2^A)$.

Extensions of the distance index to $d \geq 3$ can be obtained from (9) by assuming that the distance is measured by an average of the inequality within each class weighted by the overall population distribution across these classes. Let $I_d : [0,1]^d \to [0,1]$ be an onto function representing an inequality indicator consistent with the Lorenz criterion and such that $I_2(F_1, F_2) = |F_1 - F_2|$. The multi-group distance index is $D(F_1, \ldots, F_d) := \int_{\mathcal{X}} I_d(F_1(x), \ldots, F_d(x)) d\overline{F}(x)$. Compared to the family of indicators discussed in statement *(iii)* of Theorem **??**, $D(F_1, \ldots, F_d)$ can be used to rank *any* pair of distribution matrices irrespectively from their number and size of classes, since it is invariant to splits and insertion/elimination of empty classes.

5 Application: Education reforms and the intergenerational transmission of advantage

5.1 Data and estimation

We investigate the implications a large scale education reform on unfair inequality in long-term outcomes. We focus on the Swedish education reform, which increased compulsory

	mean	sd	min	max	count
Outcome variable					
Annual income (ln)	7.439	0.549	0.029	10.68	203,155
Controls and treatment definition					
Income year	90.42	3.433	85	96	203,155
Cohort 1953	0.464	0.499	0	1	203,155
Always treated municipality	0.213	0.410	0	1	203,155
Always control municipality	0.167	0.373	0	1	203,155
Treatment == 1	0.561	0.496	0	1	203,155
Circumstances					
Female	0.486	0.500	0	1	203,155
Father education:					
Primary	0.822	0.383	0	1	203,155
Vocational	0.078	0.268	0	1	203,155
Secondary	0.064	0.245	0	1	203,155
Higher	0.036	0.187	0	1	203,155
High ability	0.511	0.500	0	1	203,155
Urban	0.168	0.374	0	1	203,155

Table 1: Table of descriptives for the using sample

education duration, abolished streaming after grade six and introduced a uniform national curriculum. The reform was gradually introduced across a selected (albeit non-randomized) group of Swedish municipalities in 1949 until 1962. Afterwards, the reform was gradually extended to the universe of municipalities. Meghir and Palme (2005) provides an exhaustive description of the reform, the identification strategy and the data.[30] In this section, we investigate the implications of the reform for the distribution of unfair inequality in income acquisition among treated and non-treated children.

The sample we use, from Meghir and Palme (2005), covers about 10% of the Swedish population born in 1948 and 1953.[31] The sample gathers longitudinal data for about 18,000 boys and girls born in 1948 and 1953, for which income during adulthood is observed in 1985 through 1996, gathering a total of 203,155 observations. Summary statistics are available in Table 1.

[30]Literature has focused on the average effects of the reform on earnings (Meghir and Palme 2005, **?**), education (Holmlund 2008), mortality (Lager and Torssander 2012) and health (Meghir, Palme and Simeonova 2018).

[31]Anonymized data are accessible online.

The cohort 1948 roughly corresponds to the group of students which were already completing compulsory education in the early stages of the implementation of the reform in 1962, and thus were experiencing the old compulsory education system in their municipality of residence. The cohort 1953 gathers instead students who were entering secondary school when the reform was already in place in most of Sweden.

Our exercise consists in comparing dissimilarity in the income opportunities of those that born in the post-implementation period in municipalities that are exposed to the reform (the treatment group) with the dissimilarity displayed by the remaining units. The outcome of interest is given by vingitiles of the sample income distribution residuals, obtained after regressing log yearly income on a income year indicators, cohort, municipality and county fixed effects and municipality and cohort trends.[32] We separate observations into treatment and control groups. 56.1% of the sample units are in the treatment group. For each of them, we outline the distributions across the income vingitiles of 32 mutually exclusive groups, gathering observations with similar gender (two categories), father education (four categories), ability score collected in school (two categories) and being born in Stockholm, Gotheborg or Malmo. Data for both treated and control groups can be organized into 32×20 matrices, denoted respectively $\mathbf{T}$ and $\mathbf{C}$ ($\mathbf{T}, \mathbf{C} \in \mathcal{M}_{32}$). The black and gray dots in Figure 1 represent elements of the matrices $\overrightarrow{\mathbf{T}}$ (panel b) and $\overrightarrow{\mathbf{C}}$ (panel a).

5.2 Results

We use matrices $\mathbf{T}$ and $\mathbf{C}$ to test about the impact of the large-scale Swedish compulsory school reform on unfair inequality, originating from the intergenerational implications of circumstances determined at birth (father's education, gender, place of residence and abilities) on income opportunities (captured by the position each units occupies on the sample

[32]Following Meghir and Palme (2005) and Holmlund (2008), identification rests on the quasi-random assignment of the reform across municipalities. By netting out fixed effects and trends in log income we make pre and post reform cohorts income profiles comparable. Differences in incomes across cohorts of those living in municipalities that switch treatment status over the period considered (about 65% of the sample) identifies the effect of interest.

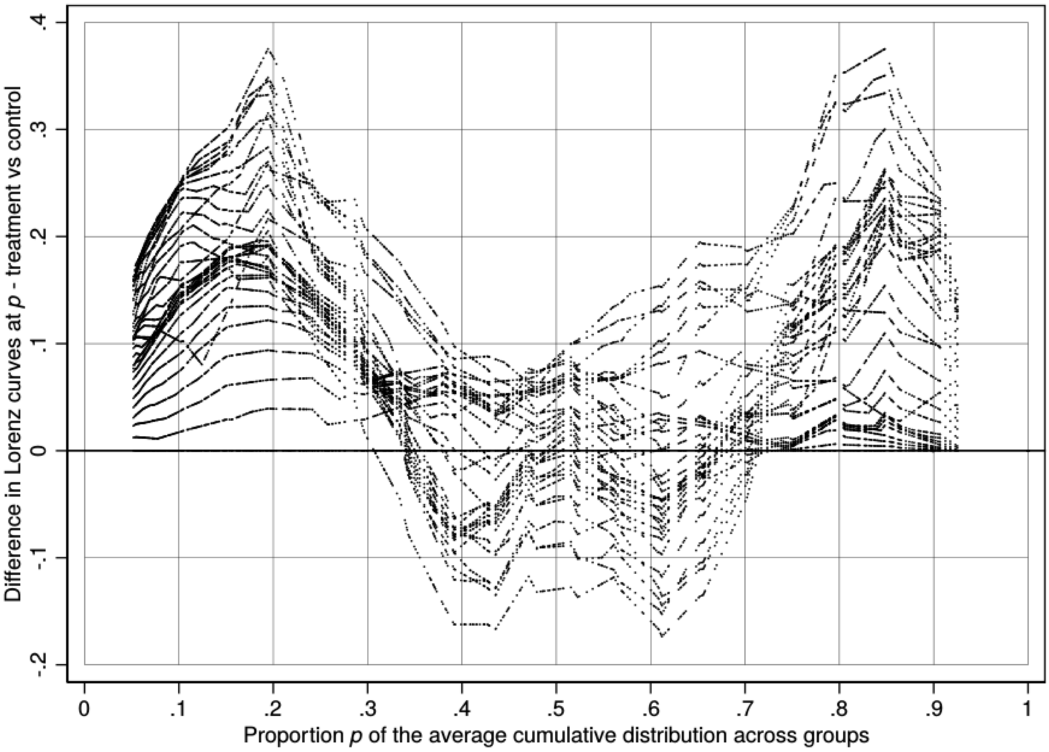


Figure 5: Dominance test for $\mathbf{T} \preccurlyeq \mathbf{C}$.

Note: Authors' computations based on Meghir and Palme (2005) data. Average population distributions for matrices $\mathbf{T}^*$ and $\mathbf{C}^*$ are reported on the horizontal axis. Claim iv) in Theorem 1 is implemented: the figure reports differences in Lorenz curves of cumulative groups distribution at given intercepts.

distribution of income). We resort on the notion of dissimilarity between distributions (circumstances) across ordered classes of realizations (income vingitiles in the sample) as a way to empirically elicit the extent of unfair inequality in the treatment and control group. We test the relation $\mathbf{T} \preccurlyeq \mathbf{C}$ to gather evidence about the unfair inequality-reducing effect of the reform.

We use the empirical criterion outlined in claim iv) of Theorem 1 to test $\mathbf{T} \preccurlyeq \mathbf{C}$. As shown in Figure 1, $\mathbf{T}$ and $\mathbf{C}$ are not ordinal comparable. We use dissimilarity preserving operations to obtain the 32×493 matrices $\mathbf{T}^* = (\mathbf{t}_1, \ldots, \mathbf{t}_{493})$ and $\mathbf{C}^* = (\mathbf{c}_1, \ldots, \mathbf{c}_{493})$, $\mathbf{T}^*, \mathbf{C}^* \in \mathcal{M}_{32}$, that are ordinal comparable. The elements of matrices $\overrightarrow{\mathbf{C}}^*$ and $\overrightarrow{\mathbf{T}}^*$ are represented by black dots in panels a) and b) respectively in Figure 3. We use these data to implement the criterion iv). For each class $j = 1, \ldots, 493$ (corresponding to a fraction p_j

of the average population distribution) and $h = 1, \ldots, 32$ of these matrices we compute the quantity $\sum_{i=1}^{h} \overrightarrow{t}^{*}_{(i)j} - \sum_{i=1}^{h} \overrightarrow{c}^{*}_{(i)j}$ and report its coordinate on the vertical axis of Figure 5. This quantity is the difference between Lorenz curves of vectors $\overrightarrow{\mathbf{t}}^{*}_{j}$ and $\overrightarrow{\mathbf{c}}^{*}_{j}$ and is positive in the large majority of the comparisons, except for central percentiles of the sample income distribution. This graph provides a formal test for the conclusions one can draw comparing panels a) and b) of figures 1 or 3, that is, that the distance between groups' distributions is smaller in the treatment group compared to the control group, albeit distributions appear more polarized along the lines of gender. As a consequence, the dominance criterion is not satisfied and we cannot conclude in favor of $\mathbf{T} \preccurlyeq \mathbf{C}$. This does not exclude, however, that consensus over the changes in dissimilarity can be reached for some interesting parametric families of dissimilarity indices.

Figure 6 reports the levels of the S-Gini family of inequality indices of groups cumulative frequencies, measured at any share of the average groups distributions, for the treatment (solid lines) and control groups (dashed lines). The parameter k expresses increasing inequality aversion. The graph provides compelling evidence that dissimilarity in groups distribution is smaller in the treatment groups compared to the comparison group. After aggregating these assessments into the dissimilarity index D_k (a version of the index in 7 where the weighting function $w \in \mathcal{W}$ is parametrized by the S-Gini weights), we find robust evidence that the schooling reform has reduced dispersion in earnings profiles: the differences $D_k(\mathbf{T}) - D_k(\mathbf{C})$ are always positive for reasonable selections of the inequality aversion parameter.[33]

[33]Differences $D_k(\mathbf{T}) - D_k(\mathbf{C})$ take values $0.215 - 0.241$ for $k = 2$, $0.317 - 0.35$ for $k = 3$, $0.373 - 0.413$ for $k = 4$ and $0.407 - 0.454$ for $k = 5$.

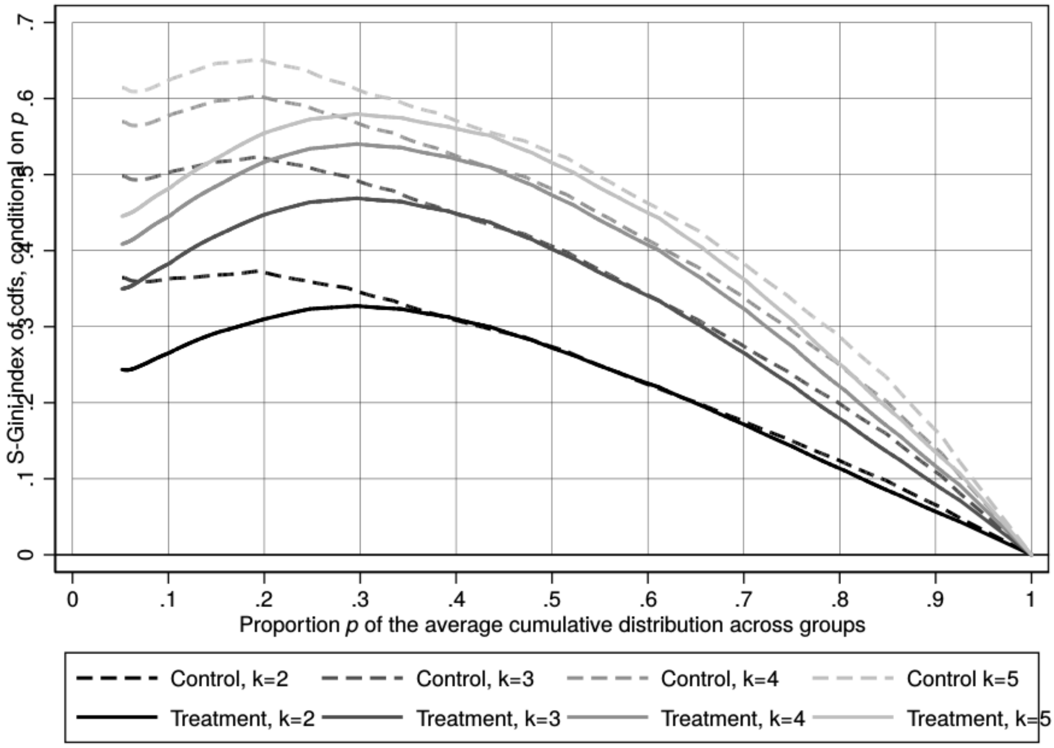


Figure 6: Comparing distributions by mean of S-Gini indices at selected proportions of the average groups' cdfs.

Note: Authors' computations based on Meghir and Palme (2005) data. Average population distributions for matrices $\mathbf{T}^*$ and $\mathbf{C}^*$ are reported on the horizontal axis. The figure reports the estimator for $\sum_{i=1}^{32} w_i t^*_{(i)}(p)$ and $\sum_{i=1}^{32} w_i t^*_{(i)}(p)$ for selected $p \in [0,1]$, where w_i is the S-Gini weighting function (parametrized by $k = 1, \ldots, 5$.

6 Concluding remarks

A large and sparse literature on segregation, discrimination, mobility, inequality and distance measurement has proposed criteria for ranking multi-groups distributions according to the dissimilarity they exhibit. The theorems presented in the paper establish the foundation of the dissimilarity model, which provides an organized and integrated measurement framework for the aforementioned phenomena. For empirical purposes, the interesting result is that the existence of dissimilarity preserving and/or reducing transformations mapping one configuration into another can be tested upon Zonotopes and Path Polytopes inclusion. The two geometric tests can be seen as multidimensional generalizations of the segregation, discrimination and concentration curves, and can lead to relevant applications in policy evaluation analysis.

For instance, a policymaker interested in reducing ethnic segregation of students across schools located in a given school district, might propose a portfolio of policy measures, none of which has to do with more “elementary” transformations such as splitting, merging, permuting schools or adding empty schools. Nonetheless, these “elementary” transformations might still be targeted as obviously segregation-preserving/reducing. If the “complex” policy measures reshape the students distribution across schools in a way that is consistent with the existence of sequences of more “elementary” transformations, then the policymaker can safely conclude that his de-segregation objective has been achieved. The policymaker can conclude that such sequence *exists* upon verification of the Zonotopes inclusion empirical test, based on the available data. Routines are made available to facilitate this task.

The same procedure applies to the analysis of dissimilarity with ordered classes. In these cases, the policymaker can use the Path Polytope inclusion criterion to assess problems of ethnic-based discrimination in different dimensions, such as earnings, accessibility to health, educational achievements or standardized test scores. This test is useful in situation where comparability issues are at stake. It might occur, for instance, when assessing whether ethnic-based labor market discrimination is stronger in terms of workers’ earnings or in terms of their determinants, such as the education and the skills of the workers (measured though standardized test scores). If the Path Polytope inclusion criterion is not rejected, then the policymaker can conclude on the existence of finite sequence of dissimilarity preserving and dissimilarity reducing exchange operations mapping one distribution matrix into another. Routines implementing the Path Polytope inclusion test are also made available.

There might be cases where Zonotopes or Path Polytopes inclusion is rejected by the data. Making use of dissimilarity indices in the classes discussed in the two theorems, it is possible to produce conclusive evaluations of the changes in dissimilarity that are consistent with the implications of the "elementary" transformations. Evaluations based on one or few dissimilarity indicators, however, are not robust and can always be challenged on the perspective offered by alternative measures. The complete characterization of the dissimilarity indicators presented in the paper is left for future research.

A Proofs

A.1 Preliminary results

We develop a rank-preserving version of Tchen's (1980) algorithm to show that the sequential majorization in claim *(??)* of Theorem 1 is always supported by the existence of a finite sequence of exchange transformations mapping the distribution matrix $\mathbf{A}$ into the less dissimilar one $\mathbf{B}$. The algorithm applies to ordinal comparable matrices, a subset of the matrices with fixed marginals analyzed in Tchen (1980). As a consequence, Tchen's algorithm is not appropriate in our setting because it does not guarantee that the rank of the groups is preserved at every step of the algorithm. Therefore conditions *(ii)* and *(iii)* of Definition 1 might be violated by Tchen's algorithm, while this is not the case for consistency by the result in Theorem 1.

Additional notation. We focus here on ordinal comparable matrices, where the order of the groups coincides with the one of the rows, so that group i dominates group $i-1$, for any i. That is, for $\mathbf{A} \in \mathcal{M}_d$, $\overrightarrow{a}_{ij} \leq \overrightarrow{a}_{i-1\,j}$, $\forall i, j$. Moreover, let (x, y) identify the cell corresponding to row x and column y of a distribution matrix, with $x \in \{1, \ldots, d\}$ and $y \in \{1, \ldots, n\}$. The lexicographic order on $\{1, \ldots, d\} \times \{1, \ldots, n\}$ that we consider is denoted by $(x, y) < (x', y')$ if $y < y'$ or if $y = y'$ and $x > x'$. We also use $i \in [x, x']$ to denote $i \in \{x, \ldots, x' | x < \ldots < x'\}$. Furthermore, the *doubly cumulative distribution matrix* of $\mathbf{A}$ is denoted by $\overrightarrow{\overrightarrow{\mathbf{A}}}$, with $\overrightarrow{\overrightarrow{a}}_{ij} = \sum_{x \geq i} \overrightarrow{a}_{xj}$. Using this compact notation, the Lorenz dominance criterion rewrites $\overrightarrow{\overrightarrow{\mathbf{B}}} \geq \overrightarrow{\overrightarrow{\mathbf{A}}}$.

Strategy of the proof. The algorithm is built in two steps that are illustrated respectively in Lemma 1 and Lemma 2. The *first step* of the algorithm delimits the building blocks of the analysis by developing a rank-preserving version of Tchen's algorithm (see Theorem 1 in Tchen 1980), from where the notation is taken. Given two ordinal comparable matrices

$\mathbf{H}$, $\mathbf{H}' \in \mathcal{M}_d$ with two elements h_{ij} and h'_{ij} satisfying $h_{ij} < h'_{ij}$, and such that $\overset{\rightrightarrows}{\mathbf{H}} \leq \overset{\rightrightarrows}{\mathbf{H}'}$ and $\overset{\rightrightarrows}{h}_{xy} = \overset{\rightrightarrows}{h'}_{xy}$ for all $(x, y) < (i, j)$, Lemma 1 will identify the sequence of transfers of groups population masses that, when applied to $\mathbf{H}$, leads to matrix $\mathbf{H}'$ by leveling the difference $h'_{ij} - h_{ij}$ in cell (i, j). This result is achieved through a finite sequence of M steps. Each step identifies a matrix $\mathbf{K}^m$ with $m \in \{1, \ldots, M\}$, where $\overset{\rightrightarrows}{\mathbf{H}} \leq \overset{\rightrightarrows}{\mathbf{K}}^m \leq \overset{\rightrightarrows}{\mathbf{K}}^{m+1} \leq \overset{\rightrightarrows}{\mathbf{H}'}$ with element k^m_{ij} such that $h_{ij} < k^m_{ij} \leq h'_{ij}$. The Lemma 1 guarantees that every matrix $\mathbf{K}^m$ is transformed into $\mathbf{K}^{m+1}$ through a finite sequence S of transfers of equal magnitude that delimits a chain of exchange transformations. The construction of the algorithm guarantees that the rank of the groups is always preserved throughout the sequence reducing the quantity $h'_{ij} - h_{ij}$.

Given two ordinal comparable distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ such that $\overset{\rightrightarrows}{\mathbf{A}} \leq \overset{\rightrightarrows}{\mathbf{B}}$, the *second step* of the algorithm develops the sequences of transfers of groups masses transforming $\mathbf{A}$ into $\mathbf{B}$ in a way that preserves, at each step of the sequence, the ranking of the groups. The first sequence, indexed by $q \in \{1, \ldots, Q\}$, identifies the cells of $\mathbf{A}$ that have to be transformed into the corresponding cells of $\mathbf{B}$. The sequence starts in $q = 1$ with cell $(d, 1)$ and moves according to the lexicographic order, from any cell (i, j) to $(i-1, j)$ if $i > 2$ or to $(d, j+1)$ if $i = 2$, and so on.[34] At each step q of the sequence the gap $b_{ij} - a_{ij}$ in (i, j) has to be eliminated before moving to step $q + 1$. In order to preserve the rank of the groups in class j, however, groups $i-1, i-2, ...$ should remain dominated by i when shifting from $\mathbf{A}^q$ to $\mathbf{A}^{q+1}$. The transformations that guarantee this no-reranking condition should sequentially transfer mass to groups $i-1, i-2, ...$ before affecting group i in class j. This subsequence is indexed by $p \in \{1, \ldots, P\}$. The two sequences together induce transfers that are bounded and guarantee:

$$\overset{\rightrightarrows}{\mathbf{A}} \leq \ldots \leq \overset{\rightrightarrows}{\mathbf{A}}^q = \overset{\rightrightarrows}{\mathbf{A}}^{q,1} \leq \ldots \leq \overset{\rightrightarrows}{\mathbf{A}}^{q,p} \leq \overset{\rightrightarrows}{\mathbf{A}}^{q.p+1} \leq \ldots \leq \overset{\rightrightarrows}{\mathbf{A}}^{q,P} = \overset{\rightrightarrows}{\mathbf{A}}^{q+1} \leq \ldots \leq \overset{\rightrightarrows}{\mathbf{B}}.$$

[34] This is so because, by ordinal comparability, a_{1j} and b_{1j} are determined by the remaining $d-1$ elements of $\mathbf{a}_j$ and $\mathbf{b}_j$.

By construction, P is finite. In fact, the matrices $\mathbf{A}^{q,p}$ and $\mathbf{A}^{q,p+1}$ can be considered as $\mathbf{H}$ and $\mathbf{H}'$ in the first step of the algorithm. Thus, $\mathbf{A}^{q,p+1}$ is obtained from $\mathbf{A}^{q,p}$ exclusively through a finite sequence of exchange operations. Extending this reasoning, also $\mathbf{B}$ is obtained from $\mathbf{A}$ exclusively through a finite sequence of exchange operations, which will prove Lemma 2.

First step of the algorithm. For any pair $\mathbf{H}, \mathbf{H}' \in \mathcal{M}_d$ of ordinal comparable matrices with $h_{ij} < h'_{ij}$, $\overrightarrow{\overrightarrow{\mathbf{H}}} \leq \overrightarrow{\overrightarrow{\mathbf{H}'}}$ and $\overrightarrow{\overrightarrow{h}}_{xy} = \overrightarrow{\overrightarrow{h'}}_{xy}$ for all $(x, y) < (i, j)$, consider the sequence of matrices $\mathbf{K}^m$ with $m \in \{1, \ldots, M\}$ where $\mathbf{K}^1 = \mathbf{H}$. Let $\mathbf{K}$ and $\mathbf{K}'$ denote two consecutive matrices in this sequence. Lemma 1.1 in Tchen (1980) identifies the operations mapping $\mathbf{K}$ into $\mathbf{K}' \in \mathcal{M}_d$ that preserve the monotonicity of $\mathbf{K}$ (i.e., that guarantee that $\overrightarrow{k}'_{ij} \leq \overrightarrow{k}'_{ij+1}, \forall i, j$). These transformations can be represented by a subsequence of matrices $\mathbf{K}^s$ with $s \in \{1, \ldots, S\}$ leading to $\mathbf{K}'$ from $\mathbf{K}$. We present a version of this subsequence that is also rank-preserving (i.e., that guarantees that $\overrightarrow{k}'_{ij} \geq \overrightarrow{k}'_{i+1\,j}, \forall i, j$).

We first show that the subsequence of matrices $\mathbf{K}^s$ exists, is finite and is related to exchange operations. For a given cell (i, j), set a row i^* such that $i^* < i$ and $k_{i^*j} > 0$, and consider $\mathbf{K}$ satisfying the following conditions:

$$k_{ij} < h'_{ij} \text{ and } \overrightarrow{\overrightarrow{k}}_{xy} = \overrightarrow{\overrightarrow{h'}}_{xy} \quad \text{for all} \quad (x, y) < (i, j), \tag{10}$$

$$\delta = \min \left\{ \overrightarrow{k}_{i-1\,j} - \overrightarrow{k}_{ij}, \overrightarrow{k}_{i^*j} - \overrightarrow{k}_{i^*+1\,j}, \frac{1}{2}(\overrightarrow{k}_{i^*j} - \overrightarrow{k}_{ij}) \right\} \; > \; 0. \tag{11}$$

Condition (10) is as in Tchen (1980), while condition (11) is new. It secures that there is enough mass that can be moved from cell (i^*, j) and added to (i, j) so that the rank of the groups is preserved. Given $\mathbf{K}$, define the sequence $S(\mathbf{K}, \mathbf{H}'|i^*) := (x_s, y_s)_{s \in \{1, \ldots, S\}}$ by setting

$$
\begin{aligned}
x_1 &= i \\
y_1 &= \min\{c|c \geq j+1,\ k_{ic} > 0\} \\
x_s &= \max\{r|i^* < r < i,\ k_{rc} > 0 \ \text{ for some } j < c < y_{s-1}\} \\
y_s &= \min\{c|c \geq j+1,\ k_{x_s c} > 0\}
\end{aligned}
$$

if $s < S$, while $(x_S, y_S) = (i^*, j)$. This sequence is nonempty with $x_S = i^* < x_{S-1} < \ldots < x_1 = i$ and $y_1 > y_2 > \ldots > y_S = j$, and leads to $\mathbf{K}'$.

Define $\mathbf{K}^1 = \mathbf{K}$ and $\mathbf{K}^s$ as the distribution matrix obtained from $\mathbf{K}^{s-1}$ where at most a mass $\Delta > 0$ is subtracted from (i, y_{s-1}) and (x_s, y_s) and added to (x_s, y_{s-1}) and (i, y_s). The mass Δ that can be moved should coincide with the smallest quantity between (i) $h'_{ij} - k_{ij}$ (the quantity that should be compensated), (ii) the frequency of group x_s in class y_s (this guarantees the monotonicity), (iii) the gap between the cumulative distributions of group i and group $i-1$, and (iv) the gap between group x_s and group $x_s + 1$. These two latter conditions guarantee that the rank of the groups is preserved by the transfer. When $x_s = i - 1$, at most half of the gap $\overrightarrow{k}_{x_s j} - \overrightarrow{k}_{ij}$ can be transferred. By construction of the sequence, at every step s $k_{x_s y} = 0$ $\forall x_s$, $\forall y \in [y_s, y_{s-1} - 1]$. Thus, conditions (iii) and (iv) are always satisfied when (11) holds. Altogether these conditions give:

$$
\Delta \ := \ \min\left\{h'_{ij} - k_{ij},\ \min_{S(\mathbf{K},\mathbf{H}'|i^*)}\{k^s_{x_s,y_s}\},\ \delta\right\}. \tag{12}
$$

Lemma 1 *Let* $\mathbf{K}$ *satisfy conditions (10) and (11), there exists* $\mathbf{K}' \in \mathcal{M}_d$ *obtained from* $\mathbf{K}$ *through a sequence of exchanges, such that* $\overrightarrow{\overrightarrow{\mathbf{K}'}} \leq \overrightarrow{\overrightarrow{\mathbf{H}'}}$ *and* $k'_{ij} = k_{ij} + \Delta$*, with* $\Delta > 0$ *as in (12).*

Proof Consider $S(\mathbf{K}, \mathbf{H}'|i^*)$ defined as above and let $\mathbf{K}^1 = \mathbf{K}$. For $s = 1$ a mass Δ is subtracted from (i, y_1) and (x_2, y_2) and added to (i, y_2) and (x_2, y_1), thereby representing an exchange transformation. In fact, by definition (12) this quantity must be lower than k_{iy_1} and $k_{x_2y_2}$, which guarantees that $\overrightarrow{k}_{iy_1} - \Delta \geq 0$ and $\overrightarrow{k}_{x_2y_2} - \Delta \geq \overrightarrow{k}_{x_2+1\, y_2}$. This operation leads to $\mathbf{K}^2$. Then, a mass Δ is subtracted from (i, y_2) (x_3, y_3) and added to (i, y_3) and (x_3, y_2) giving $\mathbf{K}^3$. By (12), also this operation is supported by an exchange transformation. The last step of this sequence involves moving mass from (i^*, j) to (i, j) where $k_{i^*j} > 0$ by definition. Recall that $i \geq 2$, hence i^* always exists. To show that $\overrightarrow{\overrightarrow{\mathbf{K}}}^s \leq \overrightarrow{\overrightarrow{\mathbf{H}}}'$ for any s, assume by recurrence that $\overrightarrow{\overrightarrow{\mathbf{K}}}^{s-1} \leq \overrightarrow{\overrightarrow{\mathbf{H}}}'$. For $(x, y) < (i, y_s)$, $k^s_{xy} = k^{s-1}_{xy}$. By definition, Δ is such that the order of groups i and $i-1$ is preserved, hence $k^s_{iy_s} > k^{s-1}_{iy_s}$ and $\overrightarrow{\overrightarrow{k}}^s_{iy_s} > \overrightarrow{\overrightarrow{k}}^{s-1}_{iy_s}$ while $\overrightarrow{\overrightarrow{k}}^s_{iy_s} < \overrightarrow{\overrightarrow{k}}^s_{i-1\, y_s}$. Moreover, $k^s_{xy} = k^{s-1}_{xy}$ for $x \in [x_s + 1, i - 1]$ and $y \in [y_s, y_{s-1}]$, hence $\overrightarrow{\overrightarrow{k}}^s_{xy} > \overrightarrow{\overrightarrow{k}}^{s+1}_{xy}$. Finally, $k^s_{x_sy_s} < k^{s-1}_{x_sy_s}$ and $\overrightarrow{\overrightarrow{k}}^s_{x_sy_s} = \overrightarrow{\overrightarrow{k}}^{s-1}_{x_sy_s}$, as well as $\overrightarrow{\overrightarrow{k}}^s_{xy_{s-1}} = \overrightarrow{\overrightarrow{k}}^{s-1}_{xy_{s-1}}$ for $x \in [x_s, i]$. Combining these conditions, the required result is obtained. *Q.E.D.*

Under (10) and (11), the iteration of the sequence $S(\mathbf{K}, \mathbf{H}'|i^*)$ in Lemma 1 might lead to three alternative outcomes. (i) The iteration might identify a transfer $\Delta = h'_{ij} - k_{ij}$ such that $k'_{ij} = h'_{ij}$, in which case $\mathbf{K}' = \mathbf{K}^M = \mathbf{H}'$ and the sequence is completed. Alternatively $\Delta < h'_{ij} - k_{ij}$, then $\mathbf{K}' \neq \mathbf{H}'$ and Lemma 1 must be reiterated. (ii) In this case, if $\delta > h'_{ij} - k_{ij}$ the rank-preserving constraints are not binding, so that $\Delta = k_{x_s\, y_s}$, where $(x_s, y_s) \in S(\mathbf{K}, \mathbf{H}'^*)$. If the condition holds starting from $\mathbf{K} = \mathbf{K}^1 = \mathbf{H}$, then it should also hold in all the following steps, since it indicates that there is enough mass in cell (i^*, j) to level the difference $h'_{ij} - h_{ij}$ and preserve the groups rankings. Lemma 1 introduces the sequence $S(\mathbf{K}^1, \mathbf{H}'^*)$ leading to $\mathbf{K}^2$. A second iteration of the lemma would give the sequence $S(\mathbf{K}^2, \mathbf{H}'^*)$ leading to $\mathbf{K}^3$, and so on. Generally, repeated iterations of the lemma lead to a sequence of distribution matrices $\mathbf{K}^m$, $m \in \{1, \ldots, M\}$ where $h'_{ij} - k^{m+1}_{ij} < h'_{ij} - k^m_{ij}$. Each of these matrices is supported by a sequence $S(\mathbf{K}^m, \mathbf{H}'^*)$ so that if $\Delta = k^m_{x_sy_s}$ for some $(x_s, y_s) \in S(\mathbf{K}^m, \mathbf{H}'^*)$, then $S(\mathbf{K}^{m+1}, \mathbf{H}'^*)$ must contain all the points of $S(\mathbf{K}^m, \mathbf{H}'^*)$ except from (x_s, y_s). Hence the former develops on a larger set of cells than the latter.

The sequence finally converges to $k_{ij}^M = h'_{ij}$ given that $S(\mathbf{K}^m, \mathbf{H}'^*)$ is a strictly increasing sequence on a finite range, indicating that it is always possible to move from $\mathbf{K}$ to $\mathbf{H}'$ in a finite number M of steps.

Finally, (iii) if instead $\delta < h'_{ij} - k_{ij}$ the iteration of Lemma 1 does not guarantee that $\mathbf{H}'$ is reached, because the rank-preserving constraint becomes binding at some point. This can be avoided by suitably redefining i^*. Next result in Lemma 2, presented in the second step of the main algorithm, will show how to iteratively construct matrices $\mathbf{H}$ and $\mathbf{H}'$ where either situation (i) or (ii) can occur.

Second step of the algorithm. The goal of the second step is to develop a sequence of rank-preserving transfers of groups masses mapping $\mathbf{A}$ into $\mathbf{B}$ whenever $\overset{\rightrightarrows}{\mathbf{B}} \geq \overset{\rightrightarrows}{\mathbf{A}}$. Every transfer of mass is constructed in such a way that Lemma 1 always applies. Thus, each transfer breaks down into a finite number of exchange transformations.

Lemma 2 *For* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *satisfying ordinal comparability, (i)* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *through a finite sequence of exchange transformations if and only if (ii)* $\overset{\rightrightarrows}{\mathbf{B}} \geq \overset{\rightrightarrows}{\mathbf{A}}$.

Proof $(i) \Rightarrow (ii)$. Suppose that $\mathbf{B}$ is obtained from $\mathbf{A}$ by an exchange transformation involving classes k and $k' > k$. Then there exists $\varepsilon > 0$ such that $\overrightarrow{b}_{hj} = \overrightarrow{a}_{hj} + \varepsilon$ and $\overrightarrow{b}_{\ell j} = \overrightarrow{a}_{\ell j} - \varepsilon$ with $\overrightarrow{b}_{ij} = \overrightarrow{a}_{ij}$ for all groups $i \neq h, \ell$ and for all classes j such that $k \leq j < k'$, while $\overrightarrow{\mathbf{b}}_j = \overrightarrow{\mathbf{a}}_j$ for all other classes. Consider first $k' = k + 1$. If $h = \ell + 1$ then $\varepsilon \leq \frac{1}{2}(\overrightarrow{a}_{\ell k} - \overrightarrow{a}_{\ell+1\ k})$. If $h > \ell + 1$ then $\varepsilon \leq \min\{(\overrightarrow{a}_{\ell k} - \overrightarrow{a}_{\ell+1\ k}), (\overrightarrow{a}_{h-1\ k} - \overrightarrow{a}_{hk})\}$. These conditions define a *rank-preserving progressive transfer* (RPPT) applied in the space of cumulative groups frequencies. If $k' > k + 1$, the exchange transformation originates a sequence of RPPT ε_j across classes $k \leq j < k'$. Setting $\varepsilon = \min_j\{\varepsilon_j\}$ guarantees that $\overrightarrow{\mathbf{b}}_j$ is obtained from $\overrightarrow{\mathbf{a}}_j$ though a RPPT, $\forall j = k, \ldots, k' - 1$. Every RPPT induces Lorenz dominance (Fields and Fei 1978), hence (ii) holds.

$(ii) \Rightarrow (i)$. Let $\overset{\rightrightarrows}{\mathbf{B}} \geq \overset{\rightrightarrows}{\mathbf{A}}$. For a given (i, j) consider a matrix $\mathbf{A}^q \in \mathcal{M}_d$ that is ordinal comparable to $\mathbf{A}$, with $q \in \{1, \ldots, Q\}$ where $\mathbf{A}^1 = \mathbf{A}$ and $\overrightarrow{\overrightarrow{\mathbf{A}}}^q \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$ such that $\overrightarrow{\overrightarrow{a}}^q_{xy} = \overrightarrow{\overrightarrow{b}}_{xy}$ for all $(x, y) < (i, j)$ and $a^q_{ij} < b_{ij}$. The sequence indexed by q identifies cells of $\mathbf{A}$. We now develop a sequence of transformations that guarantees to obtain $\mathbf{A}^{q+1} \in \mathcal{M}_d$ from $\mathbf{A}^q$ satisfying $\overrightarrow{\overrightarrow{\mathbf{A}}} \leq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q+1} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$, $\overrightarrow{\overrightarrow{a}}^{q+1}_{xy} = \overrightarrow{\overrightarrow{b}}_{xy}$ for all $(x, y) < (i, j)$ and $a^{q+1}_{ij} = b_{ij}$. There are two distinct cases where different sequences of transformations apply.

Case (a). For any class j, denote $i^* = \max\{r | r < i,\ a^q_{rj} > 0,\ \overrightarrow{a}^q_{rj} > \overrightarrow{a}^q_{ij}\}$, which defines an interval $[i^* + 1, i]$. Consider the case where $\overrightarrow{a}^q_{xj} = \overrightarrow{a}^q_{ij}$ for all $x \in [i^* + 1, i]$. To avoid re-rankings of the groups in $[i^* + 1, i]$, consider adding recursively mass to groups in class j starting from the group in position $i^* + 1$ and sequentially moving to the group in position i. The whole procedure defines a subsequence $p \in \{1, \ldots, P\}$ of transformations of $\mathbf{A}^q$, denoted $\mathbf{A}^{q,p}$ with $\mathbf{A}^{q,1} = \mathbf{A}^q$, where $\mathbf{A}^{q,2}$ is obtained only by letting $\overrightarrow{a}^{q,2}_{i^*+1\,j} = \overrightarrow{a}^{q,1}_{i^*+1\,j} + \Delta_{ij}(i^*)$ and $\overrightarrow{a}^{q,2}_{i^*j} = \overrightarrow{a}^{q,1}_{i^*j} - \Delta_{ij}(i^*)$, then $\mathbf{A}^{q,3}$ is obtained only by letting $\overrightarrow{a}^{q,3}_{i^*+2\,j} = \overrightarrow{a}^{q,2}_{i^*+1\,j}$ and $\overrightarrow{a}^{q,3}_{i^*j} = \overrightarrow{a}^{q,1}_{i^*j} - 2\Delta_{ij}(i^*)$, and for a general p the matrix $\mathbf{A}^{q,p}$ is obtained only by letting $\overrightarrow{a}^{q,p}_{i^*+p-1\,j} = \overrightarrow{a}^{q,p-1}_{i^*+p-2\,j}$ and $\overrightarrow{a}^{q,p}_{i^*j} = \overrightarrow{a}^{q,1}_{i^*j} - (p-1)\Delta_{ij}(i^*)$ until p reaches $i - i^* + 1$, where

$$\Delta_{ij}(i^*) \;=\; \min\left\{\overrightarrow{b}_{ij} - \overrightarrow{a}^q_{ij},\ \frac{1}{i - i^* + 1}\left(\overrightarrow{a}^q_{i^*j} - \overrightarrow{a}^q_{ij}\right)\right\}. \tag{13}$$

The sequence then has reached cell (i, j), giving by construction $\overrightarrow{\overrightarrow{\mathbf{A}}}^{q,1} \leq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p-1} \leq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$. If $\overrightarrow{a}^{q,p}_{ij} = \overrightarrow{b}_{ij}$, the sequence is completed and $p = P$. Otherwise $\overrightarrow{a}^{q,1}_{i^*j} - (i - i^*)\Delta_{ij}(i^*) = \overrightarrow{a}^{q,p}_{i^*+1\,j} = \ldots = \overrightarrow{a}^{q,p}_{ij} < \overrightarrow{b}_{ij}$. In this case then reset $i^{*\prime} < i^*$ and reiterate the sequence of transfers of mass $\Delta_{ij}(i^{*\prime})$. The index of the sequence moves further to $p+1$ where $\mathbf{A}^{q,p+1}$ is obtained only by letting $\overrightarrow{a}^{q,p+1}_{i^{*\prime}+1\,j} = \overrightarrow{a}^{q,p}_{i^{*\prime}+1\,j} + \Delta_{ij}(i^{*\prime})$ and $\overrightarrow{a}^{q,p+1}_{i^{*\prime}j} = \overrightarrow{a}^{q,p}_{i^{*\prime}j} - \Delta_{ij}(i^{*\prime})$ which gives $\overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p} \leq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p+1}$, and so on. By construction, this sequence develops on a finite number P of steps leading to $\overrightarrow{a}^{q,P}_{ij} = \overrightarrow{b}_{ij}$.

Case (b). Alternatively, there exist (at least one) groups in the interval $[i^* + 1, i]$ that have no mass in class j, but their cumulative distributions differ from the one of group

i. Define $\widetilde{i} := \max\{r | r \in [i^*+1, i], \overrightarrow{a}^q_{rj} > \overrightarrow{a}^q_{ij}, a^q_{rj} = 0\}$. The group occupying position $\widetilde{i}$ delimits the interval $[\widetilde{i}+1, i]$ with $\widetilde{i}+1 \leq i$. To avoid re-rankings, consider adding recursively mass in class j to the groups in $[\widetilde{i}+1, i]$, starting from the group occupying position $\widetilde{i}+1$ and sequentially moving to the group in position i. In a finite number of iterations, these transfers can either compensate the gap $\overrightarrow{b}_{ij} - \overrightarrow{a}^q_{ij}$, thus leading to $\mathbf{A}^{q+1}$, or increase groups masses in class j until the cumulative distributions of the groups in $[\widetilde{i}+1, i]$ end up coinciding with the one of group $\widetilde{i}$. The whole procedure defines a subsequence $p \in \{1, \ldots, P\}$ of transformations of $\mathbf{A}^q$, denoted $\mathbf{A}^{q,p}$ with $\mathbf{A}^{q,1} = \mathbf{A}^q$, where $\mathbf{A}^{q,2}$ is obtained only by letting $\overrightarrow{a}^{q,2}_{\widetilde{i}+1\,j} = \overrightarrow{a}^{q,1}_{\widetilde{i}+1\,j} + \Delta_{ij}(i^*, \widetilde{i})$ and $\overrightarrow{a}^{q,2}_{i^*j} = \overrightarrow{a}^{q,1}_{i^*j} - \Delta_{ij}(i^*, \widetilde{i})$, and for a generic step p the matrix $\mathbf{A}^{q,p}$ is obtained only by letting $\overrightarrow{a}^{q,p}_{\widetilde{i}+p-1\,j} = \overrightarrow{a}^{q,p-1}_{\widetilde{i}+p-2\,j}$ and $\overrightarrow{a}^{q,p}_{i^*j} = \overrightarrow{a}^{q,1}_{i^*j} - (p-1)\Delta_{ij}(i^*, \widetilde{i})$ until p reaches $i - \widetilde{i} + 1$, where

$$\Delta_{ij}(i^*, \widetilde{i}) \;=\; \min\left\{\overrightarrow{b}_{ij} - \overrightarrow{a}^q_{ij},\ \overrightarrow{a}^q_{\widetilde{i}j} - \overrightarrow{a}^q_{\widetilde{i}+1\,j},\ \frac{1}{i-\widetilde{i}}\left(\overrightarrow{a}^q_{i^*j} - \overrightarrow{a}^q_{i^*+1\,j}\right)\right\}. \tag{14}$$

The second and the third quantities in $\Delta_{ij}(i^*, \widetilde{i})$ define the rank-preserving constraints of groups i^* and $\widetilde{i}$. The sequence then has reached cell (i, j), giving by construction that $\overrightarrow{\overrightarrow{\mathbf{A}}}^{q,1} \leq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p-1} \leq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$. If $\overrightarrow{a}^{q,p}_{ij} = \overrightarrow{b}_{ij}$, the sequence is completed and $p = P$. Otherwise, at least one of the following constraints is binding:

$$\overrightarrow{a}^{q,p}_{\widetilde{i}j} = \overrightarrow{a}^{q,p}_{\widetilde{i}+1\,j} = \ldots = \overrightarrow{a}^{q,p}_{ij} < \overrightarrow{b}_{ij}, \tag{15}$$

$$\overrightarrow{a}^{q,p}_{i^*j} - (i-\widetilde{i})\Delta_{ij}(i^*, \widetilde{i}) = \overrightarrow{a}^{q,p}_{i^*+1\,j}. \tag{16}$$

If (15) holds but (16) does not hold, then the rank-preserving constraint for group $\widetilde{i}$ is binding. In this case, the algorithm proceeds by resetting $\widetilde{i}$ to $\widetilde{i}' \in [i^*, \widetilde{i}-1]$. The sequence updates to $p+1$ and generates a new matrix $\mathbf{A}^{q,p+1}$. If $\widetilde{i}' > i^*$, the sequence continues following the procedure outlined above, using transfers of mass $\Delta_{ij}(i^*, \widetilde{i}')$ defined in (14), to obtain $\mathbf{A}^{q,p+1}$ only by letting $\overrightarrow{a}^{q,p+1}_{\widetilde{i}'+1\,j} = \overrightarrow{a}^{q,p}_{\widetilde{i}'+1\,j} + \Delta_{ij}(i^*, \widetilde{i}')$ and $\overrightarrow{a}^{q,p+1}_{i^*j} = \overrightarrow{a}^{q,p}_{i^*j} - \Delta_{ij}(i^*, \widetilde{i}')$, and so on. Otherwise, if $\widetilde{i}' = i^*$ then the sequence proceeds as in *Case (a)* using the transfers of mass $\Delta_{ij}(\widetilde{i}')$ in (13) to obtain $\mathbf{A}^{q,p+1}$ only by letting $\overrightarrow{a}^{q,p+1}_{\widetilde{i}'+1\,j} = \overrightarrow{a}^{q,p}_{\widetilde{i}'+1\,j} + \Delta_{ij}(\widetilde{i}')$ and $\overrightarrow{a}^{q,p+1}_{\widetilde{i}'j} = \overrightarrow{a}^{q,p}_{\widetilde{i}'j} - \Delta_{ij}(\widetilde{i}')$. If, instead, (16) holds but (15) does not hold, i.e. $\overrightarrow{a}^{q,p}_{\widetilde{i}j} > \overrightarrow{a}^{q,p}_{\widetilde{i}+1\,j}$, then reset i^* to $i^{*\prime} < i^*$ and iterate again the sequence outlined above on the interval $[\widetilde{i}+1, i]$ while setting the feasible transfer to $\Delta_{ij}(i^{*\prime}, \widetilde{i})$. Finally, if both constraints are binding, both i^* and $\widetilde{i}$ must be reset and the algorithm is iterated. In all these situations, the order of transfers gives that $\overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p+1} \geq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p}$ by construction.

We now motivate that any given step of the algorithm leading from $\mathbf{A}^{q,p}$ to $\mathbf{A}^{q,p+1}$ can be decomposed into a finite sequence of exchange transformations, so that P must be finite as well. For any given $\mathbf{A}^{q,p}$ associated with cell (i, j), the step p identifies a cell (x, j) where $a^{q,p+1}_{xj} = a^{q,p}_{xj} + \Delta$, where Δ is defined either by (13) or by (14), depending on the prevailing case. Set $\mathbf{A}^{q,p} = \mathbf{H}$, denote with $\mathbf{H}'$ a matrix such that $\overrightarrow{\overrightarrow{h}}'_{zy} = \overrightarrow{\overrightarrow{a}}^{q,p}_{zy}$ for all $(z, y) < (x, j)$ and $h'_{xj} := a^{q,p+1}_{xj} > a^{q,p}_{xj}$. Thus $\mathbf{H}$ and $\mathbf{H}'$ satisfy condition (10). The two matrices also satisfy condition (11) as a consequence of the transfers identified in the three cases outlined above. Furthermore h'_{xj} is defined such that, given i^*, the rank-preserving constraint is never binding, i.e. $\delta > a^{q,p+1}_{xj} - a^{q,p}_{xj}$. The conditions in Lemma 1 apply, indicating that there exists a finite sequence $m \in \{1, \ldots, M\}$ with $\mathbf{K}^1 = \mathbf{H} = \mathbf{A}^{q,p}$ and with M finite, such that $\overrightarrow{\overrightarrow{k}}^{M}_{zy} = \overrightarrow{\overrightarrow{a}}^{q,p}_{zy}$ for all $(z, y) < (x, j)$ and $k^M_{xj} = a^{q,p+1}_{xj}$, thereby giving $\mathbf{K}^M = \mathbf{H}'$. It is now sufficient to set $\mathbf{A}^{q,p+1} = \mathbf{K}^M$ to be sure that $\mathbf{A}^{q,p+1}$ is obtained from $\mathbf{A}^{q,p}$ through a finite sequence of exchange transformations. So it is every step of the sequence $\{1, \ldots, P\}$, through which we conclude that P must be finite as well, and that $\mathbf{A}^{q+1} = \mathbf{A}^{q,P}$ with $a^{q+1}_{ij} = b_{ij}$ is obtained from $\mathbf{A}^q$ only through exchange transformations.

The proof of the lemma follows by iterating the algorithm outlined above, based on Lemma 1. First set $\mathbf{A}^1 = \mathbf{A}$ and $(i, j) = (d, 1)$ to obtain $\mathbf{A}^{1,P}$ where the sequence of transformations grants $\overrightarrow{\overrightarrow{\mathbf{A}}}^{1,P} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$ and $a^{1,P}_{d1} = b_{d1}$; then set $\mathbf{A}^2 = \mathbf{A}^{1,P}$ and $(i, j) = (d-1, 1)$ to obtain $\mathbf{A}^{2,P}$ with $\overrightarrow{\overrightarrow{\mathbf{A}}}^{2,P} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$, $a^{2,P}_{d1} = b_{d1}$ and $a^{2,P}_{d-1\,1} = b_{d-1\,1}$; and so on.

Q.E.D.

A.2 Proof of Theorem 1.

Proof We show that *(i)⇒(ii)⇒(iii)⇒(iv)⇒(v)⇒(i).*

(i) ⇒ (ii). All dissimilarity orderings consistent with split of classes, insertion/deletion of empty classes and interchange transformations agree that $\mathbf{A}^* \sim \mathbf{A}$ and $\mathbf{B}^* \sim \mathbf{B}$. Moreover, all dissimilarity orderings consistent with exchange transformations agree that $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$. The intersection of these orderings is thus characterized by axioms *E*, *SC*, *IEC*, *I,* and by transitivity leads to the relations $\mathbf{B} \sim \mathbf{B}^* \preccurlyeq \mathbf{A}^* \sim \mathbf{A}$ giving $\mathbf{B} \preccurlyeq \mathbf{A}$.

(ii) ⇒ (iii) ⇒ (iv). The proof is derived by analogy with Lemma **??**. There, the evaluations are taken by looking at separate classes, here they are taken at separate levels of $p \in [0,1]$.

(ii) ⇒ (iii). Note that by construction the index D_w is invariant with respect to split of classes, insertion/deletion of empty classes and interchange transformations and therefore satisfies axioms *SC*, *IEC* and *I.* Apply an exchange transformation of amount $\varepsilon > 0$ from group ℓ to h with $\ell > h$ involving adjacent classes j and $j+1$ of matrix $\mathbf{A}$. The change in D_w generated by this transformation is obtained as a weighted average of the associated changes in $\overrightarrow{a}_{(\ell)}(p)$ and $\overrightarrow{a}_{(h)}(p)$ weighted respectively by $w_\ell(p)$ and $w_h(p)$. Let $p_j := \frac{1}{d}\mathbf{e}_d^t \cdot \overrightarrow{\mathbf{a}}_j$ denote the proportion of population occupying the first j classes. By construction $\overrightarrow{a}_{(\ell)}(p)$ and $\overrightarrow{a}_{(h)}(p)$ are affected by the exchange transformation only for $p \in (p_{j-1}, p_{j+1})$. The population mass ε is transferred from group h to group ℓ uniformly in the interval (p_j, p_{j+1}) and in opposite direction, still uniformly, in the interval $(p_{j-1}, p_j]$. As a result the change in D_w is $\int_{p_{j-1}}^{p_j} [w_h(p) - w_\ell(p)]\, \varepsilon \frac{p - p_{j-1}}{p_j - p_{j-1}} dp + \int_{p_j}^{p_{j+1}} [w_h(p) - w_\ell(p)]\, \varepsilon \frac{p_{j+1} - p}{p_{j+1} - p_j} dp \leq 0$, given that $w_h(p) - w_\ell(p) \leq 0$ for all p by assumption. Thus the index is consistent also with axiom *E* and claim *(iii)* holds.

(iii) ⇒ (iv). Recall that condition *(iv)*, that is $\overrightarrow{\mathbf{a}}^t(p) \preccurlyeq^D \overrightarrow{\mathbf{b}}^t(p)$ for all $p \in [0,1]$, can

be rewritten as $\sum_{i=1}^{h} \overrightarrow{b}_{(i)}(p) \geq \sum_{i=1}^{h} \overrightarrow{a}_{(i)}(p)$ for all $h = 1, \ldots, d$ and for all $p \in [0,1]$ where by construction $\sum_{i=1}^{d} \overrightarrow{b}_{(i)}(p) = \sum_{i=1}^{d} \overrightarrow{a}_{(i)}(p)$. We show that if claim *(iv)* does not hold then also claim *(iii)* should not hold. Suppose that there exists a $q \in (0,1)$ and a group $h^* \in \{1, 2, \ldots, d-1\}$ such that the condition in claim *(iii)* is violated, that is $\sum_{i=1}^{h^*} \overrightarrow{b}_{(i)}(q) < \sum_{i=1}^{h^*} \overrightarrow{a}_{(i)}(q)$. Then by continuity of $\overrightarrow{b}_{(i)}(p)$ and of $\overrightarrow{a}_{(i)}(p)$ with respect to p it also holds that there exists an interval (q_L, q^H) such that $q \in (q_L, q^H)$ where $\sum_{i=1}^{h^*} \overrightarrow{b}_{(i)}(q) - \sum_{i=1}^{h^*} \overrightarrow{a}_{(i)}(q) < 0$ for all $q \in (q_L, q^H)$. Denote $\Delta_{(i)}(p) := \overrightarrow{b}_{(i)}(p) - \overrightarrow{a}_{(i)}(p)$, then the condition can be rewritten as $\sum_{i=1}^{h^*} \Delta_{(i)}(q) < 0$ for all $q \in (q_L, q^H)$.

Set $w_i(p) = 0$ for all $p \notin (q_L, q^H)$. It follows that $D_w(\mathbf{B}) - D_w(\mathbf{A}) = \int_{q_L}^{q^H} \sum_{i=1}^{d} w_i(p)\Delta_{(i)}(p)dp$. Let $w_i(p) = 1 - d/h^*$ for all $p \in (q_L, q^H)$ and $i = 1, 2, ..., h^*$ and $w_i(p) = 1$ for all $p \in (q_L, q^H)$ and $i = h^*+1, ..., d$, so that $\sum_{i=1}^{d} w_i(p) = 0$. Then $\sum_{i=1}^{d} w_i(p)\Delta_{(i)}(p) = \sum_{i=1}^{d} \Delta_{(i)}(p) - d/h^* \cdot \sum_{i=1}^{h^*} \Delta_{(i)}(p)$. Recalling that by construction $\sum_{i=1}^{d} \Delta_{(i)}(p) = 0$ it follows that

$$D_w(\mathbf{B}) - D_w(\mathbf{A}) = -d/h^* \cdot \int_{q_L}^{q^H} \sum_{i=1}^{h^*} \Delta_{(i)}(p)dp.$$

Given that $\sum_{i=1}^{h^*} \Delta_{(i)}(q) < 0$ for all $q \in (q_L, q^H)$, it follows that $D_w(\mathbf{B}) - D_w(\mathbf{A}) > 0$, thereby violating claim *(iii)*.

(iv) $\Rightarrow$ *(v)*. From Theorem A.2 in Marshall et al. (2011, p.30), statement *(iv)* implies that

$$conv\{\mathbf{\Pi}_d \cdot \overrightarrow{\mathbf{b}}(p) : \mathbf{\Pi}_d \in \mathcal{P}_d\} \subseteq conv\{\mathbf{\Pi}_d \cdot \overrightarrow{\mathbf{a}}(p) : \mathbf{\Pi}_d \in \mathcal{P}_d\} \tag{17}$$

for every $p \in [0,1]$, where the *conv* operator indicates the *convex hull*. Recall that, by construction, $\{\overrightarrow{\mathbf{a}}(p) : p \in [0,1]\} = MP^*(\mathbf{A})$ and similarly for $MP^*(\mathbf{B})$. Hence, the inclusion condition in (17) defined across all p's coincides with $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ given the definition of the Path Polytope $Z^*(\cdot)$.

(v) $\Rightarrow$ *(i)*. Starting from $\mathbf{A}$ and $\mathbf{B}$ and making use of the information embedded in $Z^*(\mathbf{A})$ and $Z^*(\mathbf{B})$ we will construct the associated $\mathbf{A}^*$ and $\mathbf{B}^*$ and prove claim *(i)*. In order

to construct matrix $\mathbf{A}^*$ we will consider a set of cross sections of the Path Polytope $Z^*(\mathbf{A})$.

Denote a cross-section of $Z^*(\mathbf{A})$, generated by an isopopulation hyperplane set at level $p = \frac{1}{d}\mathbf{e}_d \cdot \mathbf{p}$, by the set $conv\{\mathbf{\Pi}_d \cdot \mathbf{p} : \mathbf{\Pi}_d \in \mathcal{P}_d\}$, where $\mathbf{p}$ lies on the edge of $Z^*(\mathbf{A})$. We consider two sets of Path Polytope cross-sections associated with different proportions p of the average cumulative distributions across groups.

(i) The first set corresponds to all cross-sections delimited by $\mathbf{A}$'s classes. In this case the set of vectors $\mathbf{p}$ coincides with the vertices of $\mathbf{A}$ in $MP^*(\mathbf{A})$ that is $\overrightarrow{\mathbf{a}}_j$ for all j or their permutations such that $\exists \mathbf{\Pi}_d \in \mathcal{P}_d$ for which $\mathbf{\Pi}_d \cdot \mathbf{p} = \overrightarrow{\mathbf{a}}_j$. We define the first set as $\mathcal{S}_1 := \{p_j : p_j = \frac{1}{d}\mathbf{e}_d^t \cdot \overrightarrow{\mathbf{a}}_j, j = 1, \ldots, n_A\}$ with n_A denoting the number of classes of $\mathbf{A}$.

(ii) To identify the second set, we consider $MP^*(\mathbf{A})$ and we identify all vectors $\mathbf{p}_j \in MP^*(\mathbf{A})$ where it occurs a re-ranking of the proportions of the groups across $\mathbf{A}$'s classes. Define the indices $j = 1, \ldots, n_A^c$ associated with points $\mathbf{p}_j \in MP^*(\mathbf{A})$ that are *comonotonic*, i.e., such that for every group i and index j, the element p_{ij} of $\mathbf{p}_j$ is ordered with respect to any other element $p_{i'j}$ of $\mathbf{p}_j$, $i \neq i'$, in the same way as the element p_{ij+1} of $\mathbf{p}_{j+1}$ is ordered with respect to the element $p_{i'j+1}$ of $\mathbf{p}_{j+1}$, that is, such that $p_{ij} \geq p_{i'j} \rightarrow p_{ij+1} \geq p_{i'j+1}$ for all $i, i' \in \{1, 2, \ldots, d\}$.

To identify this set, start with $j = 1$ and set $\mathbf{p}_1 = \overrightarrow{\mathbf{a}}_1$, then for $j = 2$ derive

$$\mathbf{p}_2 := argmax_{\mathbf{p}}\{\mathbf{e}_d^t \cdot \mathbf{p} : \mathbf{p} \in MP^*(\mathbf{A}),\ \mathbf{e}_d^t \cdot \mathbf{p} > \mathbf{e}_d^t \cdot \mathbf{p}_1,\ \mathbf{p} \text{ is comonotonic to } \mathbf{p}_1\}.$$

If all group distributions are ordered in matrix $\mathbf{A}$ and there does not exist a pair of groups i, i' where (strict) re-ranking occurs, then $\mathbf{p}_2 = \overrightarrow{\mathbf{a}}_{n_A} = \mathbf{e}_d$ and $n_A^c = 2$. Else, if re-ranking occurs along $MP^*(\mathbf{A})$ starting from $\mathbf{p}_1$, then $\mathbf{p}_2$ denotes the vector in $MP^*(\mathbf{A})$ immediately preceding the vectors where (strict) re-ranking takes place and $\mathbf{p}_2$ is such that $1 > \mathbf{e}_d^t \cdot \mathbf{p}_2 > \mathbf{e}_d^t \cdot \mathbf{p}_1$. We can then reiterate the procedure to derive $\mathbf{p}_3$.

Recursively, step j of the algorithm would give

$$\mathbf{p}_j = argmax_{\mathbf{p}}\{\mathbf{e}_d^t \cdot \mathbf{p} : \mathbf{p} \in MP^*(\mathbf{A}),\ \mathbf{e}_d^t \cdot \mathbf{p} > \mathbf{e}_d^t \cdot \mathbf{p}_{j-1},\ \mathbf{p} \text{ is comonotonic to } \mathbf{p}_{j-1}\},$$

with the sequence ending after a finite number n_A^c of steps. The set of associated proportions of the average distribution across groups is denoted $\mathcal{S}_2 := \{p_j :\ p_j = \frac{1}{d}\mathbf{e}_d^t \cdot \mathbf{p}_j,\ \mathbf{p}_j \in MP^*(\mathbf{A}),\ j = 1, \ldots, n_A^c\}$.

Consider now the union of the sets derived in cases (i) and (ii), giving $\mathcal{S}_A = \mathcal{S}_1 \cup \mathcal{S}_2 := \{p_j : j = 1, \ldots, n_A^*\}$, where proportions are ordered such that $p_j < p_{j+1}$ $\forall j$ and $n_A \leq n_A^* \leq n_A + n_A^c$.

An analogous procedure identifies a set of proportions $\mathcal{S}_B := \{p_j : j = 1, \ldots, n_B^*\}$ from $Z^*(\mathbf{B})$. The union of the sets $\mathcal{S}_A$ and $\mathcal{S}_B$ is denoted $\mathcal{S} = \mathcal{S}_A \cup \mathcal{S}_B := \{p_j : j = 1, \ldots, n^*\}$ where proportions are ordered such that $p_j < p_{j+1}$ $\forall j$, with $\max\{n_A^*, n_B^*\} \leq n^* \leq n_A^* + n_B^*$, $p_1 = \frac{1}{d}\min\{\overline{a}_1, \overline{b}_1\}$ and $p_{n^*} = 1$.

We consider the sequence of indices $j = 1, \ldots, n^*$ and the associated p_j to derive the partitions of the n^* classes of the two ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ obtained from $\mathbf{A}$ and $\mathbf{B}$. These operations will not affect the shape of the Path Polytopes of $\mathbf{A}$ and of $\mathbf{B}$, respectively.

To see this, consider first matrix $\mathbf{A}$. Note that for every $p_j \in \mathcal{S}$ there exists a $\mathbf{p}_j^A \in MP^*(\mathbf{A})$ such that $p_j = \frac{1}{d}\mathbf{e}_d \cdot \mathbf{p}_j^A$. The vector $\mathbf{p}_j^A$ can be obtained from the vectors associated to the classes of $\mathbf{A}$ through *splits and elimination/insertion of empty classes*. For instance, if $\mathbf{p}_j^A$ is such that $\frac{1}{d}\mathbf{e}_d^t \cdot \overrightarrow{\mathbf{a}}_k < p_j < \frac{1}{d}\mathbf{e}_d^t \cdot \overrightarrow{\mathbf{a}}_{k+1}$ for a class k of $\mathbf{A}$, then it can be re-written as $\mathbf{p}_j^A = \overrightarrow{\mathbf{a}}_k + \lambda \mathbf{a}_{k+1}$ by construction of the Monotone Path upon which $\mathbf{p}_j^A$ lies. In this case, $\lambda \in (0, 1)$ can be interpreted as a split parameter.

The sequence of vectors $\mathbf{p}_1^A, \ldots, \mathbf{p}_{n^*}^A$ displays comonotonic elements. The condition of comonotonicity is defined for adjacent vectors in the sequence. However, the condition might not hold if it is applied to any pair of non-adjacent vectors. In fact, it is possible that

$\exists j$ such that $\mathbf{p}_j^A$ is not comonotonic to $\mathbf{p}_{j+k}^A$ for $k \geq 2$. Define, hence, a sequence of vectors $\mathbf{z}_1^A, \ldots, \mathbf{z}_{n^*}^A$ obtained by independently permuting the elements of each vector in $\mathbf{p}_1^A, \ldots, \mathbf{p}_{n^*}^A$ so that all these vectors become comonotonic to $\mathbf{p}_1^A$. We then obtain $\mathbf{z}_1^A = \mathbf{p}_1^A$, and $\mathbf{z}_2^A = \mathbf{p}_2^A$ by construction, and derive a set of permutation matrices $\mathbf{\Pi}_d^j \in \mathcal{P}_d$ so that $\mathbf{z}_3^A = \mathbf{\Pi}_d^3 \cdot \mathbf{p}_3^A$ is comonotonic with $\mathbf{z}_2^A$ (and also with $\mathbf{z}_1^A$), $\mathbf{z}_4^A = \mathbf{\Pi}_d^4 \cdot \mathbf{\Pi}_d^3 \cdot \mathbf{p}_4^A$ is comonotonic with $\mathbf{z}_3^A$ (and therefore also with $\mathbf{z}_2^A$ and $\mathbf{z}_1^A$), and so on, so that in general $\mathbf{z}_j^A = (\mathbf{\Pi}_d^j \cdot \mathbf{\Pi}_d^{j-1} \cdot \ldots \cdot \mathbf{\Pi}_d^3) \cdot \mathbf{p}_j^A$ is comonotonic with $\mathbf{z}_{j-1}^A, \ldots, \mathbf{z}_1^A$ for $j = 1, \ldots, n^*$. The vectors $\mathbf{z}_1^A, \ldots, \mathbf{z}_{n^*}^A$ identify either vertices or points along the edges of $Z^*(\mathbf{A})$.

Define the matrix $\overrightarrow{\mathbf{A}^*} := (\mathbf{z}_1^A, \ldots, \mathbf{z}_{n^*}^A)$, that by construction of the vectors $\mathbf{z}_j^A$ satisfies the properties of a cumulative distribution matrix. The underlying distribution matrix is denoted $\mathbf{A}^* := (\Delta\mathbf{z}_1^A, \ldots, \Delta\mathbf{z}_{n^*}^A)$ where $\Delta\mathbf{z}_j^A := \mathbf{z}_j^A - \mathbf{z}_{j-1}^A \geq \mathbf{0}_d$ with $\mathbf{z}_0^A := \mathbf{0}_d$ such that $\Delta\mathbf{z}_1^A = \mathbf{z}_1^A = \mathbf{p}_1^A$. The definition of $\mathbf{A}^*$ clarifies that the group permutations mapping points $\mathbf{p}_j^A$ into $\mathbf{z}_j^A$ can be *associated with a sequence of interchange of groups transformations* applied to $\mathbf{A}^*$. According to the definition of Axiom *I*, in fact, one can construct the sequence of the interchange of groups permutation matrices by considering the matrices $(\mathbf{\Pi}_d^j \cdot \mathbf{\Pi}_d^{j-1} \cdot \ldots \cdot \mathbf{\Pi}_d^3)$ for $j = 3, 4, \ldots, n^* - 1$ where each generic matrix $\mathbf{\Pi}_d^j$ that involve permutations of more than two groups could be decomposed itself into a sequence of matrices involving only permutations of two groups.

Define in a similar way the sequence $\mathbf{z}_1^B, \ldots, \mathbf{z}_{n^*}^B$ and the matrix $\mathbf{B}^* := (\Delta\mathbf{z}_1^B, \ldots, \Delta\mathbf{z}_{n^*}^B)$, where $\mathbf{z}_j^B$ is either a vertex or lies along the edges of $Z^*(\mathbf{B})$. By construction, the group distributions in matrix $\mathbf{A}^*$ and in matrix $\mathbf{B}^*$ are ordered by stochastic dominance, and the order of the groups coincides in both matrices up to an independent permutations of the rows of the matrices. Furthermore, $\frac{1}{d}\mathbf{e}_d^t \cdot \mathbf{A}^* = \frac{1}{d}\mathbf{e}_d^t \cdot \mathbf{B}^*$. Hence the following claim holds:

Claim A: $\mathbf{A}^*$ and $\mathbf{B}^*$ *are ordinal comparable matrices* obtained from $\mathbf{A}$ and $\mathbf{B}$ respectively through operations of split of classes, insertion/elimination of empty classes and interchange of groups transformations.

Moreover, these matrices are such that $Z^*(\mathbf{A}^*) = Z^*(\mathbf{A})$ and $Z^*(\mathbf{B}^*) = Z^*(\mathbf{B})$. Thus, $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ in claim *(v)* implies $Z^*(\mathbf{B}^*) \subseteq Z^*(\mathbf{A}^*)$, that is

$$conv\{\mathbf{\Pi}_d \cdot \overrightarrow{\mathbf{b}}^*_j : \mathbf{\Pi}_d \in \mathcal{P}_d\} \subseteq conv\{\mathbf{\Pi}_d \cdot \overrightarrow{\mathbf{a}}^*_j : \mathbf{\Pi}_d \in \mathcal{P}_d\},$$

for all $j = 1, \ldots, n^*$. According to Marshall et al. (2011) Theorems 1.A.2 and 2.B.2, this condition is equivalent to $\overrightarrow{\mathbf{b}}^{*t}_j$ Lorenz dominates $\overrightarrow{\mathbf{a}}^{*t}_j$ for all $j = 1, \ldots, n^*$. This latter condition is considered in Lemma **??**, as claim *(iv)*. According to the equivalence between claims *(iv)* and *(i)* in Lemma **??** it follows that $\mathbf{B}^*$ can be obtained from $\mathbf{A}^*$ through a finite sequence of exchange transformations, which combined with claim A presented above gives claim *(i)* and concludes the proof.

Q.E.D.

A.3 Proof of Corollary 1

Proof easy *Q.E.D.*

A.4 Proof of Corollary 2

Proof For $\mathbf{A} \in \mathcal{M}_2$, denote the boundaries of the Path Polytope $PP(\mathbf{A})$ by $\phi(p) = F_2(F_1^{-1}(p))$ and $\psi(p) = F_1(F_2^{-1}(p))$. The two boundaries delimit areas with respect to the diagonal expressing perfect similarity, that are denoted $A_\phi = \int_0^1 |p - \phi(p)|\, dp$ and $A_\psi = \int_0^1 |p - \psi(p)|\, dp$, respectively. The two Path Polytope boundaries are symmetric with respect to the diagonal, and by construction $\phi(p)$ and $\psi(p)$ satisfy $\phi \circ \psi(p) = p = \psi \circ \phi(p)$ at any p. Building on this, we express the two areas as $A_\phi = \int_{\mathcal{X}} |F_1(x) - F_2(x)| dF_1(x)$ by changing the variable of integration to $p = F_1(x)$ and $A_\psi = \int_{\mathcal{X}} |F_2(x) - F_1(x)| dF_2(x)$ by changing the variable of integration to $p = F_2(x)$. From this it is possible to write half of

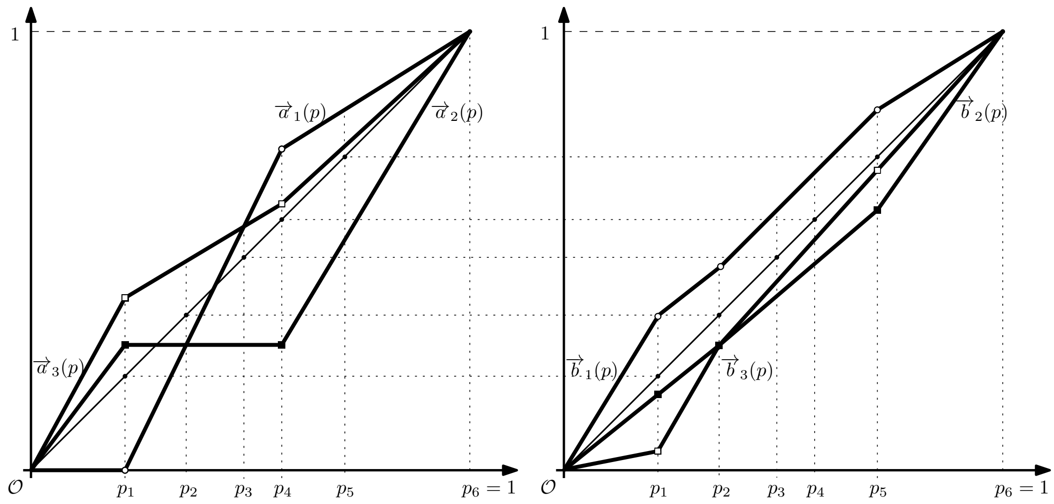


Figure 7: Cumulative distributions for three groups

the area of the Path Polytope (equal to $A_\phi + A_\psi$) as

$$\frac{1}{2}(A_\phi + A_\psi) = \int_{\mathcal{X}} |F_1(x) - F_2(x)| \, d\frac{1}{2}(F_1(x) + F_2(x)) = D(F_1, F_2).$$

Q.E.D.

B Implementation

This section describes a procedure to obtain ordinal comparable matrices, that employs dissimilarity preserving operations exclusively. The relation with cumulative distributions representations is also discussed. The cumulative distribution functions of matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_3$ are pictured in figure 7. Share of the average population distribution p are reported on the horizontal axis, whereas cdfs corresponding to any $p \in [0, 1]$ are on the vertical axis. Each distribution is identified by a specific marker, corresponding to groups proportions in different classes of the underlying distribution matrices. In the example, $n^A = 3$ and $n^B = 4$.

References

Andreoli, F., Havnes, T. and Lefranc, A. (2019). Robust inequality of opportunity comparisons: Theory and application to early childhood policy evaluation, *The Review of Economics and Statistics* **101**(2): 355–369.

Andreoli, F. and Zoli, C. (2020). From unidimensional to multidimensional inequality: A review., *METRON* **78**: 5–42.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models, *Econometrica* **74**(2): pp. 431–497.

Atkinson, A. B. and Bourguignon, F. (1982). The comparison of multi-dimensioned distributions of economic status, *The Review of Economic Studies* **49**(2): 183–201.

Bertino, S., Drago, E., Landenna, G., Leti, G. and Marasini, D. (1987). Dissimilarity and connection, *in* A. Naddeo (ed.), *Italian Contributions to the Methodology of Statistics*, Cleup Padova, chapter 3, pp. 83 – 109.

Bonhomme, S. and Sauder, U. (2011). Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling, *Review of Economics and Statistics* **93**(2): 479–494.

Butler, R. J. and McDonald, J. B. (1987). Interdistributional income inequality, *Journal of Business & Economic Statistics* **5**(1): pp. 13–18.

Cowell, F. (2000). Measurement of inequality, Vol. 1 of *Handbook of Income Distribution*, Elsevier, chapter 2, pp. 87 – 166.

Dagum, C. (1980). Inequality measures between income distributions with applications, *Econometrica* **48**(7): pp. 1791–1803.

Dardanoni, V. (1993). Measuring social mobility, *Journal of Economic Theory* **61**(2): 372 – 394.

Dardanoni, V., Fiorini, M. and Forcina, A. (2012). Stochastic monotonicity in intergenerational mobility tables, *Journal of Applied Econometrics* **27**(1): 85–107.

Donaldson, D. and Weymark, J. A. (1998). A quasiordering is the intersection of orderings, *Journal of Economic Theory* **78**(2): 382 – 387.

Ebert, U. (1984). Measures of distance between income distributions, *Journal of Economic Theory* **32**(2): 266 – 274.

Epstein, L. G. and Tanny, S. M. (1980). Increasing generalized correlation: A definition and some economic consequences, *The Canadian Journal of Economics* **13**(1): pp. 16–34.

Fields, G. S. and Fei, J. C. H. (1978). On inequality comparisons, *Econometrica* **46**(2): pp. 303–316.

Fusco, A. and Silber, J. (2013). On social polarization and ordinal variables: The case of self-assessed health, *The European Journal of Health Economics* pp. 1–11.

Gajdos, T. and Weymark, J. A. (2012). Introduction to inequality and risk, *Journal of Economic Theory* **147**(4): 1313 – 1330. Inequality and Risk.

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

Gini, C. (1914). Di una misura di dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche, *Atti del R. Istitituto Veneto di Scienze Lettere e Arti* **LXXIII**.

Hardy, G. H., Littlewood, J. E. and Polya, G. (1934). *Inequalities*, London: Cambridge University Press.

Holmlund, H. (2008). A researcher's guide to the Swedish compulsory school reform, *LSE Research Online Documents on Economics 19382*, London School of Economics and Political Science, LSE Library.

Jenkins, S. J. (1994). Earnings discrimination measurement: A distributional approach, *Journal of Econometrics* **61**: 81–102.

Kanbur, R. and Stiglitz, J. E. (2016). Dynastic inequality, mobility and equality of opportunity, *Journall of Economic Inequality* **14**(4): 419–434.

Kolm, S.-C. (1969). *The Optimal Production of Social Justice*, Public Economics, London: McMillan, pp. 145–200.

Lager, A. C. J. and Torssander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million swedes, *Proceedings of the National Academy of Sciences* **109**(22): 8461–8466.

Le Breton, M., Michelangeli, A. and Peluso, E. (2012). A stochastic dominance approach to the measurement of discrimination, *Journal of Economic Theory* **147**(4): 1342 – 1350.

Lefranc, A., Pistolesi, N. and Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France, *Journal of Public Economics* **93**(11-12): 1189 – 1207.

Maccheroni, F., Muliere, P. and Zoli, C. (2005). Inverse stochastic orders and generalized Gini functionals, *METRON* **63**(3): 529–559.

Mahalanobis, P. C. (1960). A method of fractile graphical analysis, *Econometrica* **28**(2): pp. 325–351.

Marshall, A. W., Olkin, I. and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*, Springer.

Meghir, C. and Palme, M. (2005). Educational reform, ability, and family background, *American Economic Review* **95**(1): 414–424.

Meghir, C., Palme, M. and Simeonova, E. (2018). Education and mortality: Evidence from a social experiment, *American Economic Journal: Applied Economics* **10**(2): 234–56.

Meyer, M. and Strulovici, B. H. (2013). The supermodular stochastic ordering, *Technical report*, CEPR Discussion Paper No. DP9486.

Ramos, X. and Van de gaer, D. (2016). Approaches to inequality of opportunity: Principles, measures and evidence, *Journal of Economic Surveys* **30**(5): 855–883.

Roemer, J. (1998). *Equality of Opportunity*, Harvard University Press, Cambridge.

Roemer, J. E. (2012). On several approaches to equality of opportunity, *Economics and Philosophy* **28**: 165–200.

Shaked, M. and Shanthikumar, G. (2006). *Stochastic Orders*, Springer, New York.

Shorrocks, A. F. (1978). The measurement of mobility, *Econometrica* **46**(5): 1013–1024.

Shorrocks, A. F. (1982). On the distance between income distributions, *Econometrica* **50**(5): pp. 1337–1339.

Stiglitz, J. (2012). *The Price of Inequality*, Norton.

Tchen, A. H. (1980). Inequalities for distributions with given marginals, *The Annals of Probability* **8**(4): 814–827.

Van de gaer, D., Schokkaert, E. and Martinez, M. (2001). Three meanings of intergenerational mobility, *Economica* **68**(272): 519–37.

Weymark, J. A. (1981). Generalized Gini inequality indices, *Mathematical Social Sciences* **1**(4): 409 – 430.

Yaari, M. E. (1987). The dual theory of choice under risk, *Econometrica* **55**(1): pp. 95–115.

Ziegler, G. (1995). *Lectures on Polytopes*, number 152 in *Graduate Texts in Mathematics*, Springer-Verlag, New York.