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*

January 15, 2023

Abstract

The analysis of many economic phenomena requires partitioning societies into mutually exclusive groups of individuals sharing similar characteristics and studying the extent at which these groups are distributed with different intensities across ordered realizations of a relevant outcome, 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. This paper introduces a novel empirical robust criterion for dissimilarity which is based on sequential dominance comparisons and is capable of ordering multi-group distributions defined over ordinal outcomes. The criterion is characterized in terms of existence of a finite sequence of basic transformations of the data regarded to as unambiguously preserving or reducing dissimilarity. An application to Sweden highlights the usefulness of the criterion to identify the intergenerational distributional consequences of a large education reform which took place in the 1960s.

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

JEL Codes: D63, J71, J62, D30.

*Contact details: `francesco.andreoli@univr.it` (F. Andreoli) and `claudio.zoli@univr.it` (C. Zoli, corresponding author). We are indebted to Valentino Dardanoni, Vincenzo Denicoló, Louis Eeckhaudt, Martin Foster, Alessio Fusco, Nicolas Gravel, Arnaud Lefranc, Brice Magdalou, Patrick Moyes, Eugenio Peluso, Philippe Van Kerm and Gaston Yalonetzky for their useful comments. We are also grateful to participants to ECINEQ, LAGV, SSCW meeting, EEA-ESEM conferences and seminars at Canazei Winter School, THEMA, Verona, Palermo, CORE, Bocconi, Leicester, Universitat Autonoma Barcelona, 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 PREOPP (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.

1 Introduction

1.1 The problem

Many economic phenomena are concerned with the way a society is split into mutually exclusive groups, defined for instance along the lines of gender, age, ethnicity, parental background, and the way such groups are distributed across ordered realizations of an underlying outcome of interest.

For instance, *discrimination* occurs when groups attain different wage levels with different intensities (see for instance Gastwirth 1975, Dagum 1980, Jenkins 1994). *Unfair inequality* arises instead when the chances of attaining a certain outcome (such as self-assessed health, education, human capital, skills or income) differ across individuals with different characteristics (for a review, see Roemer and Trannoy 2016). 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 often employed to assess *intergenerational mobility* (see for instance Dardanoni 1993, Jäntti and Jenkins 2015). A transition matrix displays low mobility when the probability of achieving any of the percentiles in the children income distribution depends on the income of the parents. Conversely, *origin independence* is achieved when child distributions coincide across parental background groups.

There is widespread agreement in the literature about what constitutes lack of discrimination, or a fair distribution of resources, or origin independence. 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, p. 189), 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."When this is the case, groups are equally represented at each realization, albeit they may take on different outcomes with different intensity. Conversely, the case of maximal dissimilarity occurs when the groups membership can be inferred from the knowledge of the realization. This is always the case when the groups distributions are not overlapping, i.e. the highest realization achieved by any of the groups is smaller than the lowest realization achieved by any other group which dominates it in the sense of first order stochastic dominance. Arguably, all these cases identify situations of maximal discrimination (as in Le Breton, Michelangeli and Peluso 2012), immobility (Dardanoni 1993, Van de gaer, Schokkaert and Martinez 2001) or social envy (Roemer 1998, Fleurbaey 2008).

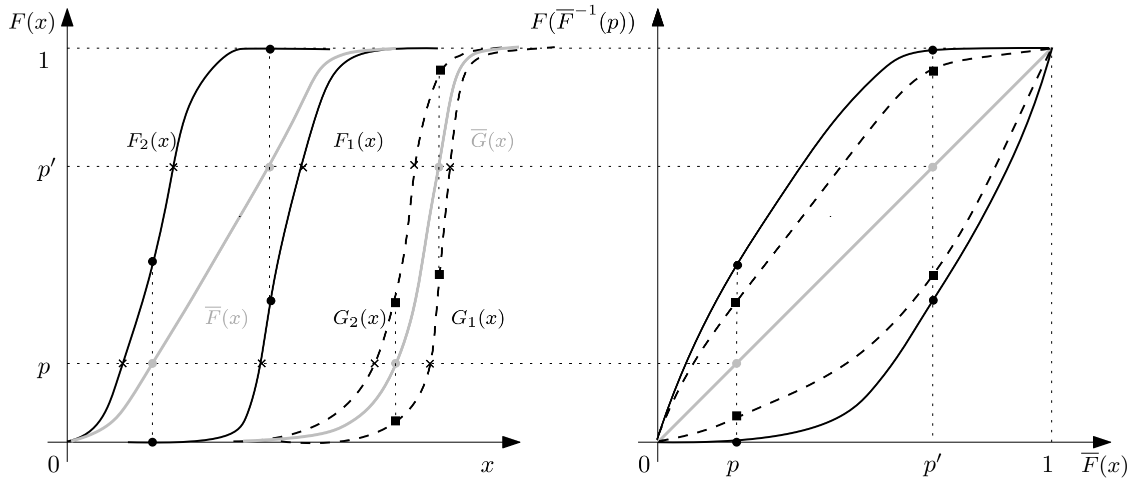


Figure 1: Configurations F (solid lines) and G (dashed lines) on the left panel, and their representations preserving ordinal information (right panel).

All other cases display some dissimilarity, but not maximal dissimilarity. This paper investigates the normative foundations for ranking sets of two or more distributions defined over ordinal outcomes according to the extent of dissimilarity that they exhibit. Sets of multi-group distributions are denoted *configurations*. A configuration could represent, for instance, a country, a year or a certain policy regime. The dissimilarity order of configurations that we envision should value the degree at which groups are disproportionally represented across realizations in each configuration and then rank configurations accordingly. The best-ranked configuration by this order is the one in which groups are equally represented at any realization, i.e. when similarity prevails. The next section provides an operational definition of such criterion and illustrates the way this paper addresses its limitations while contributing to the literature.

1.2 Contribution of the paper

Consider first the basic situation in which configurations consist of two groups $i = 1, 2$, each represented by its cumulative distribution function (cdf) F_i. We assume in this expositional example, for simplicity, that $F_i(x)$ is continuous for $x \in \mathbb{R}$. We are interested in ranking configurations $F = (F_1, F_2)$ and $G = (G_1, G_2)$ according to the degree of disproportionality in groups cdfs that they exhibit. Examples of F and G are reported in panel a) of Figure 1.

Comparing the *vertical* gap between cdfs evaluated at any realization x, that is $|F_1(x)-$

$F_2(x)|$ and $|G_1(x) - G_2(x)|$, yields an intuitive criterion for assessing relative disproportionality in the shares of both groups across configurations F and G.[1] For some problems, however, this criterion may not be appropriate. This is the case, for instance, when the outcomes in two configurations are measured on different scales. Test score achievements, material well-being, and self-reported health or education, could be indeed measured on different scales across countries and time. It could also be the case when the realizations are the residuals of an underlying estimating model that requires adopting non-linear transformations of the outcomes(such as applying the log-scale, see also Athey and Imbens 2006, Bonhomme and Sauder 2011). Furthermore, outcome x may not correspond to the relevant variable according to which dissimilarity between configurations should be assessed. For instance, a policymaker could be interested in comparing two configurations by the dissimilarity they exhibit, but these evaluations should be based on distributions of unknown transformations (such as *utility evaluations*) of the outcomes and not on the distribution of the outcomes. In all these cases the dissimilarity criterion should guarantee consistency of the ranking of the configurations irrespective of monotonic increasing transformations applied to the variables considered in each configuration. For this purpose the comparisons of the gaps in the two configurations could be made at quantiles associated with fixed positions of the inividuals in a given distribution that depends on F_1 and F_2 for configuration F (and on G_1 and G_2 respectively for configuration G) instead of considering the same level of x.

We can make use of reference distributions, denoted $\overline{F}$ and $\overline{G}$, in order to identify comparable realizations across configurations and then compare groups disproportionalties in correspondence to those realizations. A possible candidate for a reference distribution $\overline{F}$ (and $\overline{G}$) could be the *symmetric average groups distribution* $\overline{F}(x) := \frac{1}{2}F_1(x) + \frac{1}{2}F_2(x)$ (computed analogously for $\overline{G}$).[2] Examples of average groups distributions are given by $\overline{F}$ and $\overline{G}$, represented with gray curves in Figure 1.

The quantiles of $\overline{F}$ and $\overline{G}$ at any $p \in [0, 1]$, which are denoted respectively $\overline{F}^{-1}(p)$ and $\overline{G}^{-1}(p)$, identify the comparable realizations across configurations. Evaluating F and G at these quantiles the obtained quantity $F_i(\overline{F}^{-1}(p))$ measures the proportion of group i

[1]A variety of contributions have analyzed distance between distributions (Shorrocks 1982, Ebert 1984, Chakravarty and Dutta 1987), divergence (Magdalou and Nock 2011) as well as welfare gaps (Andreoli, Havnes and Lefranc 2019).

[2]The Symmetric average groups distribution satisfies few relevant properties: $\overline{F}$ is a function of the data (i.e., $\overline{F} : \mathbb{R} \mapsto [0, 1]$ is an onto function mapping $\overline{F}(x) := \overline{F}(F_1(x), F_2(x))$ at any x); $\overline{F}$ is *consistent* with the notion of similarity (if $F_1 = F_2 = F$ then $\overline{F} = F$); $\overline{F}$ treats all groups *symmetrically* irrespectively of their labels (that is, $\overline{F}(F_1, F_2) = \overline{F}(F_2, F_1)$) and responds to *monotone* changes in F along the domain of realizations (that is, $\overline{F}(F_1(x), F_2(x)) < \overline{F}(F_1(x) + \varepsilon_1, F_2(x) + \varepsilon_1)$ $\forall \varepsilon_1, \varepsilon_2 \in [0, 1]$ small enough).

achieving at most the same realizations as the poorest $p100\%$ of the symmetric average group distribution $\overline{F}$. Define $G_i(\overline{G}^{-1}(p))$ accordingly. Dissimilarity cannot be larger in G compared to F if disproportionality in groups cdf is smaller in the former configuration compared to the latter at every quantile of the average distributions $\overline{F}$ and $\overline{G}$, that is if:

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

This condition is illustrated in Panel b) of Figure 1 where one can note that there is less disproportionality between groups in G than in F, implying that G displays at most as much dissimilarity between groups as F.

The condition in (1) defines a robust dissimilarity criterion that relies exclusively on information about disproportionality in groups composition. It has some interesting features. First, dissimilarity comparisons based on (1) are invariant to the shape of $\overline{F}$ and $\overline{G}$, which depend on features of the unequal distribution of realizations within groups across p. The dissimilarity criterion is instead focused on between-group inequalities among F_1 and F_2 at any p separately.

Second, the criterion is naturally bounded by the cases of similarity (when $F_1 = F_2 = F$, the distributions coincide with the diagonal in Panel b)) and of maximal dissimilarity (when F_1 and F_2 do not overlap, the distributions coincide with the unit square in Panel b)). Disproportionality between groups turns out to be irrelevant for comparing configurations once distributions do not overlap. In this case, either $F_i(0.5) = 0$ or $F_i(0.5) = 1$ and its value is constant over realizations not belonging to the joint support, implying that any additional consideration about dissimilarity should necessarily rely on the cardinal scale of the outcome domain. Finally, comparing disproportionality based on reference distributions guarantees invariance to any monotone transformation of the outcomes scale, thereby leading to dissimilarity evaluations that preserve exclusively ordinal information of the data.[3]

The operational definition of dissimilarity embedded in criterion (1) is, to our knowledge, novel, in the sense that its normative justification, its empirical implementation and its extension to the multi-group (i.e. to configurations made of more than two distributions) have not been yet treated in the literature. This paper fills these gaps.

[3] To see this, let h be a monotone mapping of x yielding $u := h(x)$. If x is income, u could be the utility evaluation of an income realization. Let $X_i \sim F_i$ and $h(X_i) = U_i \sim \tilde{F}_i$, yielding $\tilde{F} = (\tilde{F}_1, \tilde{F}_2)$ with $\overline{\tilde{F}}$. Select u, x so that $\overline{\tilde{F}}(u) = p = \overline{F}(x)$. For any i (and similarly for G), one has that $\tilde{F}_i(\overline{\tilde{F}}^{-1}(p)) := \Pr[U_i \leq \overline{\tilde{F}}^{-1}(p)] = \Pr[U_i \leq u] = \Pr[h^{-1}(U_i) \leq x] = \Pr[X_i \leq x] = \Pr[X_i \leq \overline{F}^{-1}(p)] = F_i(\overline{F}^{-1}(p))$. Any increasing transformation applied to outcomes of F and/or of G preserves the graphs in panel b) of Figure 1.

The normative justification for assessing dissimilarity as disproportionality rests on the properties of the set of transformation of the data that, when applied to a given configuration, either preserve or reduce the extent of dissimilarity that it exhibit. Transformations that reshape the outcome scale and the labeling of groups, without affecting disproportionality across groups are regarded to as dissimilarity preserving (see also Andreoli and Zoli 2022). Moreover, dissimilarity evaluations should not depend on the order of the groups, as implied by an anonymity requirement with respect to the labeling of the groups. We also consider a unique dissimilarity-reducing transformation, which consists in *exchanging* deteriorations in outcomes for a better-off group with improvements in outcomes for a worse-off group (for applications in multidimensional inequality analysis, see Epstein and Tanny 1980, Tchen 1980, Atkinson and Bourguignon 1982, Faure and Gravel 2021). A group is better-off (worse-off) if it is represented with higher (lower) intensity at the bottom of the outcomes domain compared to another. By effect of an exchange, the relative cumulative frequency of the worse-off group decreases whereas that of the better-off groups rises, thereby reducing groups disproportionality.

Disproportionality is groups frequencies cannot be increased if one configuration is obtained from another through a sequence of such transformations. Testing for the existence of such a sequence is a daunting task. Our main result, in Theorem 1, identifies a feasible, empirical test that allows to verify whether such sequence of operations exists involving in a finite number of steps. We prove that the testing algorithm applies to configurations with two or more distributions using a robust criterion for comparisons of inequality in groups cumulative distributions -i.e., dominance in Lorenz curves at any quantiles of the symmetric average distribution[4]- in order to conclude on groups disproportionality.

We illustrate the feasibility of the test in a policy evaluation study. In the study, we resort on data and identification strategy in Meghir and Palme (2005) in order to compare actual earnings opportunities of those treated by the 1960s Swedish education reform to counterfactual earnings distributions that would have prevailed in the absence of the reform. We assess the distributional effects of the reform across the population, divided into 32 mutually exclusive groups defined by their gender, parental background, skills and place of birth. Inequalities related to those differences are unfair and deserve compensation. The dissimilarity criterion is used to test the effects of the reform for reducing unfair inequality. There are 1,263 relevant thresholds to be tested, yielding to more than 40,000

[4]Notice that the Lorenz curve coordinates are given by the sequential sum of the proportions $F_i(\overline{F}^{-1}(p))$, $i = 1, \ldots, d$, ranked in ascending order. Disproportionality increases as the share of the least represented group at p decreases.

estimates of Lorenz curves across treatment and counterfactual configurations. Despite complexity, we can represent all such comparisons on a graph and conclude on dominance via visual inspection of it. We find that we cannot conclude that the Swedish reform reduced dissimilarity between earnings opportunities of different groups.

The impossibility to conclude owes to the *incompleteness* of the ordering of configurations implied by the implementation algorithm. In fact, the dissimilarity order of configurations that we investigate is *robust* not only to the way outcomes in each distribution can be transformed, but also to the way in which dissimilarity between distributions is valued. An evaluation could represent, for instance, preferences of a policymaker interested in a policy evaluation assessment. It is shown that the testing algorithm ranks configurations only when the order is backed up by unanimous agreement among all possibly ways of valuing dissimilarity that are consistent with the aforementioned transformations. When the order fails, disagreement exists. Refining the set of admissible evaluations could eventually lead to more conclusive results.

1.3 Organization of the paper

For expositional convenience, the criterion in (1) has utilized continuous distributions to illustrate how the concept of dissimilarity related to disproportionality. The rest of the paper, however, deals with *discrete* distributions of ordered outcomes. This setting is appropriate to analyze with situations in which realizations correspond to ordered categories, such as self-reported health or education levels. The discrete setting is also relevant for *empirical* analysis of dissimilarity when outcomes are discrete or continuous, such as income or test score achievements. In this latter case, outcomes may correspond to groups proportions observed at any realization of an (arbitrarily) fine grid of outcomes, which is later used to estimate the continuous empirical distribution via interpolation methods.[5] Discrete distributions can be organized into matrices, each representing a configuration. Section 2 provides notation and introduces the axiomatic model. Section 3 introduces the dissimilarity empirical test and provides the main result of the paper. The result is then related to existing literature on stochastic orders, multidimensional inequality, mobility and inequality of opportunity. Section 4 provides extensions. First, it shows that the dissimilarity test is robust vis-à-vis the way the algorithm is tested on the data. Second, it characterizes an intuitive family of dissimilarity indices and shows its relation with the dissimilarity robust

[5]Remember that any continuous distribution, such as the cdf of a continuously measurable outcome, can always be seen as the limit case of a stepwise, discrete distribution (see Asplund and Bungart 1966).

criterion. Section 5 reports results from the policy evaluation exercise. Section 6 concludes. Proofs are collected in Appendix A.

2 A framework for robust dissimilarity comparisons

2.1 Notation

We represent data through *distribution matrices* of size $d \times n$, depicting sets of distributions (indexed by rows) of $d \geq 2$ *groups* across $n \geq 2$ *classes*. Each class represents one of a finite number of disjoint ordered realizations of an underlying outcomes variable. Notice that n can be very large (for instance, corresponding to a very fine grid of realizations) but it may also hold that $n < d$. Distribution matrices with the same number of groups but variable number of classes are collected in the set

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

where a_{ij} is the proportion (empirical frequency) 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}.

Outcomes are linearly ordered and the arrangement of the classes is consistent with this order (i.e., if $j < k$ then realization j is not better than realization k). The *cumulative distribution matrix*, $\overrightarrow{\mathbf{A}} \in \mathbb{R}_+^{d,n_A}$, is obtained by sequentially cumulating the elements of the classes of $\mathbf{A}$, so that $\overrightarrow{\mathbf{a}}_k := \sum_{j=1}^{k} \mathbf{a}_j$. Moreover, let denote with $p_1, \dots, p_n$ the realizations of the (symmetric) *average cumulative groups distribution* across groups, so that $p_j = \frac{1}{d} \sum_{i=1}^{d} \overrightarrow{a}_{ij} \ \forall j$.

Example. *Let* $\mathbf{A} \in \mathcal{M}_3$ *and its cumulation* $\overrightarrow{\mathbf{A}}$ *be:*

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.1 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0 & 0.5 \\ 0.1 & 0.1 & 0.6 & 0.2 \end{pmatrix} \quad and \quad \overrightarrow{\mathbf{A}} = \begin{pmatrix} 0.4 & 0.5 & 0.8 & 1 \\ 0.1 & 0.5 & 0.5 & 1 \\ 0.1 & 0.2 & 0.8 & 1 \end{pmatrix}. \qquad (2)$$

Here, $a_{13} = 0.3$ *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%*. The average cumulative groups distribution gives*

$p_1 = 0.2$, $p_2 = 0.4$, $p_3 = 0.7$ *and* $p_4 = 1$.

Furthermore, we say that the distribution of group h *first order stochastic dominates* that of groups ℓ (or, in short, 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. 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 $\mathbf{0}_n := (0, \ldots, 0)^t$, where the superscript "t" denotes transposition. We denote $\mathcal{P}_n$ the set of $n \times n$ permutation matrices.

2.2 Axioms

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 distributions in $\mathbf{B}$ display at most as much dissimilarity as distributions in $\mathbf{A}$.[6] The ordering $\preccurlyeq$ may represent, for instance, a particular way to evaluate dissimilarity according to the preferences of a decision maker, who is interested in ranking configurations by the extent of dissimilarity they exhibit.

In this section, we provide few, intuitive properties that every meaningful ordering $\preccurlyeq$ should satisfy. Properties are based on transformations of the data that, when applied to any given distribution matrix $\mathbf{A}$, are bound to produce a new distribution matrix $\mathbf{B}$ that cannot display more dissimilarity than $\mathbf{A}$, thereby leading to $\mathbf{B} \preccurlyeq \mathbf{A}$ by all orderings consistent with such operations.

The first operation that we consider regards every exchange operation as not dissimilarity increasing. Consider a distribution matrix $\mathbf{A} \in \mathcal{M}_d$ in which group h dominates ℓ. Unless the distributions of groups h and ℓ coincide, there must exist a class k where $\overrightarrow{a}_{hk} < \overrightarrow{a}_{\ell k}$ such that $a_{\ell k} > 0$. An *exchange operation* consists of 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

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

consequences for dissimilarity. 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.

The Exchange Axiom E posits that any such exchange operation is bound not to increase dissimilarity.

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* $\overrightarrow{d}_{ij} \leq \overrightarrow{d}_{i'j}$ *then* $\overrightarrow{b}_{ij} \leq \overrightarrow{b}_{i'j}$ *for all groups* $i \neq i'$ *and for all classes* j*, then* $\mathbf{B} \preccurlyeq \mathbf{A}$.

Any exchange operation preserves the "size" of the distribution matrices under comparison, that is $n_A = n_B = n$. Moreover, the exchange always preserves the column margins of a distribution matrix, denoted $\frac{1}{d}\mathbf{1}_d^t \cdot \mathbf{A}$, which should then coincide in matrices $\mathbf{A}$ and $\mathbf{B}$, that is $\frac{1}{d}\mathbf{1}_d^t\mathbf{a}_j = \frac{1}{d}\mathbf{1}_d^t\mathbf{b}_j$ for any class j.

Only matrices with equal size and margins can be compared in terms of exchange transformations. The remaining properties characterize the indifference class of dissimilarity orderings through operations that are capable of expanding the set of admissible matrices.

The axioms *IEC* and *SC* introduce operations that reshape the number and size of the classes of a distribution matrix without affecting the underlying ordinal information about the groups distributions, thereby preserving 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 emphases 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 proportionality (or lack thereof) of groups distributions.

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.

Axiom *ISC* (*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 *ISC* 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 (dis)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 with different size and margins.

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. This makes dissimilarity evaluations independent from the label of the groups. Moreover, 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 makes dissimilarity evaluations capable of valuing disproportionalities in groups composition in each class, irrespectively of the way such groups contribute to dissimilarity in previous or later classes. 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, *such that* $\mathbf{B} = (\mathbf{a}_1, ..., \mathbf{a}_k, \mathbf{\Pi}_{h,\ell} \cdot \mathbf{a}_{k+1}, ..., \mathbf{\Pi}_{h,\ell} \cdot \mathbf{a}_{n_A})$ *whenever* $\overrightarrow{a}_{hk} = \overrightarrow{a}_{\ell k}$, *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. As a consequence, distributions are treated symmetrically in dissimilarity comparisons, meaning that both positive and negative gaps between the distributions contribute equally to measured dissimilarity and do not compensate each others. When the distributions of the groups can be ordered according to *strong* form of stochastic dominance (i.e. groups cumulative distributions never coincide on $p \in (0,1)$), 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, i.e. the groups cumulative distribution functions intersect) 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, per se, sources of indifference in terms of dissimilarity.

When paired with *IEC*, axiom I leads to indifference to *permutations of groups* labels. Orderings consistent with both axioms regards situations in which the underlying groups distributions are the same, but their labels do not coincide across matrices, as indifferent.

The dissimilarity axioms are independent. Axioms *IEC* and *ISC* have been introduced by Andreoli and Zoli (2022) in the context of analysis of dissimilarity with *categorical* outcomes, and are maintained in our setting. Axioms E and I are instead novel. Andreoli and Zoli (2022) have introduced an axiom setting *independence with respect to permutations of the classes*. This axiom is obviously disregarded when classes are ordered. The possibility of permuting classes is intertwined with the consequences of the *Merge* axiom, which is seen as the prevalent dissimilarity-reducing operations when outcomes are unordered categories. A merge operation consists in adding, group by group, the frequencies of each group observed in class k and $k+1$ to form a class of larger size, thereby smoothing groups compositional disparities. A simple example with two groups shows that Exchange and Merge axioms are incompatible, implying a logical distinction between dissimilarity analysis with ordinal and with categorical, non-ordered outcomes.

Example (continued). *Let* $\mathbf{A}' \in \mathcal{M}_2$ *corresponds to rows one and two of* $\mathbf{A}$ *in (2). Consider merging (element by element) classes 2 and 3 of matrix* $\mathbf{A}'$ *and then splitting in proportion* $5/8$ *the obtained class. This gives matrix* $\mathbf{B}' \in \mathcal{M}_d$ *such that:*

$$\mathbf{A}' \xrightarrow{merge} \begin{pmatrix} 0.4 & 0 & 0.4 & 0.2 \\ 0.1 & 0 & 0.4 & 0.5 \end{pmatrix} \xrightarrow{split} \begin{pmatrix} 0.4 & 0.4\frac{5}{8} & 0.4\frac{3}{8} & 0.2 \\ 0.1 & 0.4\frac{5}{8} & 0.4\frac{3}{8} & 0.5 \end{pmatrix} = \mathbf{B}'.$$

Configuration $\mathbf{B}'$ *is unambiguously less dissimilar than* $\mathbf{A}'$ *according to the criterion in Andreoli and Zoli (2022). Yet, for* $\varepsilon = 0.15$*:*

$$\mathbf{A}' = \begin{pmatrix} 0.4 & 0.25-\varepsilon & 0.15+\varepsilon & 0.2 \\ 0.1 & 0.25+\varepsilon & 0.15-\varepsilon & 0.5 \end{pmatrix} \preccurlyeq \mathbf{B}'$$

for all dissimilarity orderings consistent with axiom E. A reversal in the dissimilarity orderings occurs, highlighting the incompatibility of Merge and Exchange axioms.

2.3 The robust dissimilarity partial order

If $\mathbf{B}$ is obtained from $\mathbf{A}$ through transformations that are preserving or reducing dissimilarity, then $\mathbf{B} \preccurlyeq \mathbf{A}$ is agreed upon by all dissimilarity orderings consistent with axioms *SC*, *IEC*, *I* and *E*. We call the *partial* order originated by the intersection of complete dissimilarity orderings $\preccurlyeq$ consistent with these axioms as the *robust dissimilarity (partial) order*. The order is robust in the sense that it regards all orderings $\preccurlyeq$ as equally deserving and is therefore consistent with each of them. Robustness comes, however, at the cost of completeness (see Donaldson and Weymark 1998).

The robust dissimilarity (partial) order is *transitive*, implying that pairwise comparisons can be extended into rankings of multiple distribution matrices. Importantly, the dissimilarity-preserving operations are always source of indifference for the robust partial order.

Remark 1 *Let* $\mathbf{A} \in \mathcal{M}_d$ *and let* $\mathbf{A}^* \in \mathcal{M}_d$ *be obtained from* $\mathbf{A}$ *through insertion/elimination of empty classes, split of classes and interchanges of groups, then* $\mathbf{A}^* \sim \mathbf{A}$ *for all orderings* $\preccurlyeq$ *satisfying axioms SC, IEC and I.*

Second, the robust partial order is bounded by intuitive cases corresponding to similarity or to maximal dissimilarity. A *similarity* matrix $\mathbf{S}$ represents a situation in which the distributions of all groups coincide and can be represented by the same row vector $\mathbf{s}^t \in \Delta^n$. Conversely, a *maximal dissimilarity* matrix $\mathbf{D}$ represents instead situations where each class is occupied at most by one group *and* the groups cumulative distributions do not overlap. Any distribution matrix is always ranked in-between these two cases.

Remark 2 *For any* $\mathbf{S}, \mathbf{D}, \mathbf{A} \in \mathcal{M}_d$*,* $\mathbf{S} \preccurlyeq \mathbf{A} \preccurlyeq \mathbf{D}$ *for all orderings* $\preccurlyeq$ *satisfying axioms SC, IEC, I and E.*

There are infinitely many matrices that can be represented as $\mathbf{S}$ and $\mathbf{D}$. 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, but not maximal dissimilarity.

Remark 3 *Let* $\mathbf{S}, \mathbf{S}' \in \mathcal{M}_d$ *be two distinct similarity matrices and* $\mathbf{D}, \mathbf{D}' \in \mathcal{M}_d$ *be two distinct maximal dissimilarity matrices,* $\mathbf{S} \sim \mathbf{S}'$ *and* $\mathbf{D} \sim \mathbf{D}'$ *for all orderings* $\preccurlyeq$ *satisfying axioms SC, IEC, I and E.*

The robust partial order of dissimilarity cannot be verified empirically, insofar it would require to assess dominance separately for *all* orderings consistent with the dissimilarity axioms. We provide a simple algorithm which is capable of testing for the robust dissimilarity ranking in a finite number of steps. The algorithm is a discrete, multi-group version of the intuitive criterion of dissimilarity (1) outlined in the Introduction.

3 An operational criterion for robust dissimilarity comparisons

3.1 The empirical dissimilarity test

We exploit dissimilarity preserving and reducing operations to derive a testing algorithm. We consider first the possibility of using an arbitrary sequence of transformations applied to matrices $\mathbf{A}$ and $\mathbf{B}$ to generate matrices $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ such that $\mathbf{A}^* \sim \mathbf{A}$ and $\mathbf{B}^* \sim \mathbf{B}$ for all dissimilarity orderings $\preccurlyeq$ satisfying axioms *SC*, *IEC*, *I*. The sequence of transformations can be arbitrarily chosen so that the resulting matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ satisfy the property of *ordinal comparability*.

Definition 1 (Ordinal comparability) *The matrices* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_d$ *are ordinal comparable if (i)* $\mathbf{1}_d^t \cdot \mathbf{A} = \mathbf{1}_d^t \cdot \mathbf{B}$, *(ii) group* h *dominates* $h+1$ $\forall h = 1, \dots, d-1$ *in* $\mathbf{A}$ *and* $\mathbf{B}$, *and (iii) group* h *dominates* ℓ *in* $\mathbf{A}$ *if and only if* h *dominates* ℓ *in* $\mathbf{B}$.

In ordinal comparable matrices (i) margins coincide, (ii) all groups are ordered according to stochastic dominance in both matrices and (iii) the order of the groups coincide across the matrices. Lemma 1 in the appendix provides an algorithm to construct ordinal comparable matrices of size $d \times n^*$ given any pair $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$.

Recall that an exchange operation, when applied to a matrix $\mathbf{A}$, does not affect the margins or the dominance relations between the groups, thereby yielding to another matrix $\mathbf{B}$ that is ordinal comparable to $\mathbf{A}$ by construction. Recall moreover that every exchange transformation gives rise to rank-preserving progressive transfers (Fields and Fei 1978) in the space of groups *cumulative* frequencies within at least one class. Any such class can be seen as a distribution of groups proportions with average size p, where all groups are uniformly weighted $\frac{1}{d}$. Every exchange transfer implies a reduction in the heterogeneity of groups cumulative distributions that is always intercepted by *Lorenz curves dominance*[7] for

[7]Recall that, for any pair of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^d$ such that $\frac{1}{d}\sum_i a_i = \frac{1}{d}\sum_i b_i = p$ and $a_{(i)}, b_{(i)}$ denote the i-th smaller elements, $\mathbf{b}$ Lorenz dominates $\mathbf{a}$ if and only if $\frac{1}{d}\sum_{i=1}^h \frac{1}{p} b_{(i)} \geq \frac{1}{d}\sum_{i=1}^h \frac{1}{p} a_{(i)}$ (or, equivalently, $\sum_{i=1}^h b_{(i)} \geq \sum_{i=1}^h a_{(i)}$) $\forall h = 1, \dots, d$, with equality holding for $h = d$ (see Marshall, Olkin and Arnold 2011).

$d-$dimensional vectors of realizations with equal mean and uniform weights (Kolm 1969, Cowell 2000, Marshall et al. 2011, Andreoli and Zoli 2020).

All together, the transformations implied by the IEC, ISC, I and E axioms suggest an *empirical dissimilarity test* which can be implemented in a finite (but possible large) number of steps based on the data provided by the distribution matrices.[8] The test requires first to map $\mathbf{A}, \mathbf{B}$ into two equivalent (from the dissimilarity perspective) representations $\mathbf{A}^*, \mathbf{B}^*$ that are ordinal comparable. Then, the test resorts on comparisons of relative Lorenz curves for vectors of groups cumulative frequencies corresponding to each of the n^* classes of $\mathbf{A}^*$ and $\mathbf{B}^*$. Each pair of such vectors $\overrightarrow{\mathbf{b}^*}_j$ and $\overrightarrow{\mathbf{a}^*}_j$ that have to be compared has the same size $dp_j := \mathbf{1}_d^t \overrightarrow{\mathbf{b}^*}_j = \mathbf{1}_d^t \overrightarrow{\mathbf{a}^*}_j$, for every $j = 1, \ldots, n^*$. The test yields a partial order of distribution matrices, which is denoted $\preccurlyeq^{\Delta}$.

Definition 2 (Dissimilarity test) *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$ *if and only if for* $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ *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, the condition*

$$\Delta(h, p_j) := \frac{1}{d}\frac{1}{p_j}\sum_{i=1}^{h} \overrightarrow{b}^*_{(i)j} - \frac{1}{d}\frac{1}{p_j}\sum_{i=1}^{h} \overrightarrow{a}^*_{(i)j} \geq 0, \tag{3}$$

holds for all $h = 1, \ldots, d$ *and for all* $j = 1, \ldots, n^*$.

The statistic $\Delta(h, p_j)$ compares one of the d coordinates of Lorenz curves of vectors $\overrightarrow{\mathbf{b}^*}_j$ with the corresponding coordinates of $\overrightarrow{\mathbf{a}^*}_j$. When the statistic is positive, it means that the combined frequency of the h (out of d) less-represented groups observed in the bottom $p_j 100\%$ of the population in configuration $\mathbf{B}^*$ is larger (i.e., the h groups are less under-represented) than the corresponding frequency observed in configuration $\mathbf{A}^*$. There are $d * n$ values of the $\Delta(h, p_j)$ statistic to be estimated. If condition (3) holds for every of the $d * n$ cases, then there is strong evidence that disproportionality in groups composition in $\mathbf{B}^*$ is smaller than in $\mathbf{A}^*$. If the statistic takes on negative values for at least one h, then Lorenz dominance fails and the criterion is not capable of ordering the two distribution matrices. When there are only two distributions (which we conventionally denote $h = 2$), the statistic $\Delta(2, p_j)$ coincides with $\Delta(2, p_j) = \left|\overrightarrow{a^*}_{1j} - \overrightarrow{a^*}_{2j}\right| - \left|\overrightarrow{b^*}_{1j} - \overrightarrow{b^*}_{2j}\right| \geq 0, \ \forall j$, which is a very intuitive measure of disproportionality in groups cumulative shares.

Appendix A.5 provides a simple example that illustrates the implementation of the

[8]The algorithm has been coded and made available in the replication package of this paper.

algorithm and its relation to dissimilarity operations.

3.2 Main result and discussion

The main result of the paper shows that the dissimilarity test $\preccurlyeq^{\Delta}$ is not only a necessary but also a sufficient statistic for the robust partial order of dissimilarity.

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

(i) $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *using operations of split of classes, insertion/elimination of empty classes, interchange and exchange,*

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

(iii) $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$*.*

The theorem delivers multiple contributions. First, it provides an implementable algorithm for testing agreement among all dissimilarity orderings that rank distribution matrices consistently with the dissimilarity-reducing and -preserving operations. The algorithm does not identify the exact sequence of transformations, but it guarantees that such a sequence exists finite. The algorithm supporting $\preccurlyeq^{\Delta}$ is designed in a way that it is capable or retaining exclusively ordinal information about the distribution of the underlying ordered outcome. This is done by using the quantiles of the average population distribution, $p \in [0, 1]$, as a reference. To see this, notice that the statistic $\Delta(h, p_j)$ can be represented on a graph where ps are on the horizontal axis and values of the statistics are reported on the vertical axis. The fact that the statistic is normalized by the average size of a class, dp_j, allows to compare realizations (and the severity of violations of dominance) across the domain of p. As we will show, such a property is useful to produce aggregate evaluations of dissimilarity.

Second, the dissimilarity test is suitable for comparing configurations involving more than two groups and a variable number of outcomes. Transformations that are regarded as dissimilarity preserving allow to change the shape of the distribution matrix while preserving groups proportions across classes. Only Lorenz dominance comparisons based on the Δ statistic allow to conclude about the dissimilarity ranking of configurations.

Third, the testing algorithm is useful to clarify the relation between dissimilarity and groups (dis)proportionality across classes. Lorenz dominance is used as the relevant criterion to assess disproportionality, thereby understood as inequality in groups cumulative frequencies observed in correspondence to classes of the distribution matrices. Any reduction in such inequality is necessarily the result of progressive transfers in the space of groups frequencies, which are implemented through exchange operations.

Results in Theorem 1 relate to different multidimensional dispersion orderings analyzed in the literature. The Exchange transformation is related to correlation-reducing transfers analyzed in Epstein and Tanny (1980), Tchen (1980) and Atkinson and Bourguignon (1982). Axiom E regards distance in cdfs and correlation reduction as two equivalent perspectives for dissimilarity evaluations. When combined with *I*, concerns for correlation are wiped out, insofar any interchange of groups distribution that implements an endogeneous ranking of distribution by mean of stochastic dominance is always bound to rise correlation. However, the same operations are seen as a source of indifference by all orderings consistent with axiom *I*. The dissimilarity criterion emphasises therefore the distance between distributions.

The dissimilarity test is also related to the *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. Differently from Tchen's result, the dissimilarity test is capable of ranking configurations in which 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).

The dissimilarity test can be related as well to the orthant order introduced by Meyer and Strulovici (2013), which is useful to assess supermodularity in matrices with different class margins. Their result decomposes 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.

The dissimilarity test yields new contributions to the field of *unfair inequality measurement*. One example is *income mobility* analysis. In that case, a distribution matrix could represent a mobility matrix, yielding the probabilities of moving to a given class of destination (out of n) conditional on the knowledge of the class of departure (out of d). Jäntti and Jenkins (2015) provide an extensive discussion about mobility orderings to which the dissimilarity criterion relates. One of such criteria is the orthant order in Dardanoni

(1993). In that setting, 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 1 provides a characterization of the orthant order in terms of existence of a sequence of elementary exchange transformations mapping one matrix to another while reducing dissimilarity, which is alternative to Dardanoni's characterization in terms of social welfare functions. In fact, 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 1 applies as well to mobility assessments in which the distribution of departure and that of destination differ. This occurs, for instance, when the distribution of departure is given in terms of income deciles and that of destination in terms of income centiles (so $d < n$). In more general cases, the dissimilarity criterion and that in Dardanoni (1993) do not correspond.

If groups identify percentiles of the parental income distribution while classes correspond to percentiles of the children income distribution, then a distribution matrix could represent an *intergenerational* mobility matrix, where both groups and classes are ordered (evidence on mobility matrices is discussed in Chetty, Hendren, Kline and Saez 2014). When the mobility matrix is *monotone*, that is the group in row $i+1$ stochastic dominates the group in row i for any i, the dissimilarity criterion only requires testing for changes in mobility across selected percentiles in the distribution of destination, provided that a perfectly mobile society could be described as one where the distribution of child 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) and Theorem 1 provides a characterization of it. The dissimilarity partial order extends mobility comparisons when margins differ. This is an important aspect for empirical research, since in many cases either 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), or the parental income is approximated by, or substituted with, observable circumstances of origin (such as parental education or the gender). The *inequality of opportunity* literature emphasizes the latter case, the focus being on unfair inequalities originating from dissimilarities across circumstances group rather than outcomes inequalities (for a review, see Roemer and Trannoy 2016, Ferreira and Peragine 2016). Ferreira and Gignoux (2011) stress the importance of using dissimilarity indices that are sensitive to ordinal information, when opportunities are measured in dimensions other than income, such as cognitive ability or well-being. All these situations could not be compared within the mobility framework or by employing a measurement approach which ignores the ordered feature of the outcomes, although they could be compared in terms of the robust dissimilarity criteria described in Theorem 1.

If the mobility matrices are non-monotone, the dissimilarity criterion imposes stronger conditions than the traditional mobility tests. Empirical evidence suggests that monotonicity of mobility matrices is unlikely to be rejected by the data (Dardanoni, Fiorini and Forcina 2012).

4 Extensions

4.1 Generalizing the validity of the dissimilarity test

The dissimilarity test invoked in Theorem 1 requires performing a finite number of comparisons in order to conclude on $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$. This number crucially depends on the choice of $\mathbf{A}^*$ and $\mathbf{B}^*$ ordinal comparable. There are infinitely many of such ordinal comparable matrices that can be obtained from the same pair $\mathbf{A}$ and $\mathbf{B}$. We provide a geometric representation of dissimilarity that is invariant to the choice of $\mathbf{A}^*$ and $\mathbf{B}^*$ and show its relation with $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$.

For $\mathbf{A} \in \mathcal{M}_d$, consider the mapping $\overrightarrow{a}_i(p)$, which is an onto function defined on $[0,1]$ such that $\overrightarrow{a}_i(0) := 0$, $\overrightarrow{a}_i(p_n) = \overrightarrow{a}_i(1) = 1$ and $\overrightarrow{a}_i(p_j) = \overrightarrow{a}_{ij}$ for any $j = 1, \ldots, n$. Assume further that $\overrightarrow{a}_i(p)$ is piecewise linear for any $p \in (p_{j-1}, p_j)$, $j = 1, \ldots, n$. Its functional form obtains by interpolating linearly the cumulative distributions $\overrightarrow{\mathbf{a}}_{j-1}$ and $\overrightarrow{\mathbf{a}}_j$, yielding the vector of coordinates $\overrightarrow{\mathbf{a}}(p) := (\overrightarrow{a}_1(p), \ldots, \overrightarrow{a}_d(p))^t$ satisfying $\overrightarrow{\mathbf{a}}(p) := (1-\lambda)\overrightarrow{\mathbf{a}}_{j-1} + \lambda\overrightarrow{\mathbf{a}}_j$, where $\lambda \in [0,1]$ is the interpolation parameter. This parameter is identified by the fact that for $p \in (p_{j-1}, p_j)$, the average groups distribution gives $p = \frac{1}{d}\sum_i \overrightarrow{a}_i(p) = (1-\lambda)p_{j-1} + \lambda p_j$ if and only if $\lambda = (p - p_{j-1})/(p_j - p_{j-1})$. Altogether, the compact expression for the functional form of $\overrightarrow{a}_i(p)$ (and hence of $\overrightarrow{a}_i(p)$, $\forall i$) becomes:

$$\overrightarrow{\mathbf{a}}(p) := (1-\lambda)\overrightarrow{\mathbf{a}}_{j-1} + \lambda\overrightarrow{\mathbf{a}}_j = \overrightarrow{\mathbf{a}}_{j-1} + \frac{p - p_{j-1}}{p_j - p_{j-1}}\mathbf{a}_j, \quad \text{with } p \in (p_{j-1}, p_j).$$

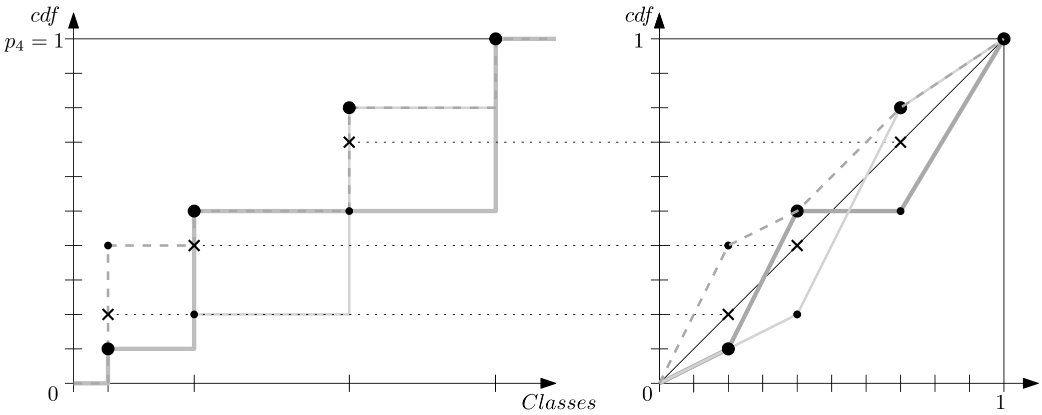


Figure 2: Cdf of distributions in $\mathbf{A}$ (left) and their cumulative group distributions (right).

The plot of $\overrightarrow{a}_i(p)$ lies in the unit square and is piecewise linear over the domain p. The vector $\overrightarrow{\mathbf{a}}(p)$ provides the proportion of each group that achieves a realization that is smaller or equal to that achieved by the $p100\%$ of the uniform average distribution, for all $p \in [0, 1]$.

Example (continued). *The cdfs of groups in $\mathbf{A}$ in (2) are drawn on the left hand-side panel of Figure 2 (group 1 in bold gray, group 2 in dashed, group 3 in light gray). The cardinalizaton of the classes is arbitrary. The cumulative group distributions of $\mathbf{A}$ in (2) are reported on the right-hand side of Figure 2. To obtain such curves, we first plot the levels $\overrightarrow{\mathbf{a}}(p)$ in correspondence to every intercept of the average group distribution $p \in \{0.2, 0.4, 0.7, 1\}$. These originate points lying on the unit square with coordinates $(p, \overrightarrow{a}_i(p))$ for $i = 1, 2, 3$, marked with dots in the figure. Larger dots imply that two or three groups coordinates coincide. We then sequentially connect these points with linear segments, starting from the origin and ending on the point $(1, 1)$, to obtain three separate group cumulative distributions.*

The representation $\overrightarrow{\mathbf{a}}(p)$ of the data is invariant to any dissimilarity-preserving transformation, thereby giving the next remark.

Remark 4 *Let $\mathbf{A} \in \mathcal{M}_d$ and $\mathbf{A}^*$ be obtained from $\mathbf{A}$ through split of classes, insertion/elimination of empty classes and interchanges. Then $\overrightarrow{\mathbf{a}}(p) = \overrightarrow{\mathbf{a}^*}(p)$ for any $p \in [0, 1]$ up to a permutation of the vectors.*

It is intuitive that adding empty classes (so that $\mathbf{a}_j = \mathbf{0}_d$) does not affect $\overrightarrow{\mathbf{a}}(p)$. Splitting proportionally all entries in a class j generates instead the piecewise linear arrangement. The interchange axiom could rearrange the order of the distributions, but does not affect the ordered vector $(\overrightarrow{a}_{(1)}(p), \ldots, \overrightarrow{a}_{(d)}(p))$.

The following corollary provides a criterion in claim (ii) that generalizes to the continuum the sequential Lorenz dominance test advocated in $\preccurlyeq^{\Delta}$. The new criterion ranks $\mathbf{B}$ is at most as dissimilar as $\mathbf{A}$ whenever the proportions of the groups adding up to the bottom p100% of the average of the cumulative distributions across groups in $\mathbf{B}$ (i.e $\overrightarrow{\mathbf{b}}(p)$) dominate, in the sense of Lorenz curve dominance, the corresponding proportions in $\mathbf{A}$ (i.e. $\overrightarrow{\mathbf{a}}(p)$), for any $p \in [0,1]$. Owing to Remark 4, if the condition is verified for all p, then the partial order $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$ is also verified irrespectively of the choice of $\mathbf{A}^*$ and $\mathbf{B}^*$ ordinal comparable.

The interesting result is that $\preccurlyeq^{\Delta}$, which simply requires to select one pair $\mathbf{A}^*$ and $\mathbf{B}^*$ ordinal comparable among all potential alternatives, is also sufficient to grant dominance using $\overrightarrow{\mathbf{a}}(p)$ and $\overrightarrow{\mathbf{b}}(p)$ representations. The choice of the ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ is hence irrelevant to conclude on $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$.

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

i $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$*;*

ii For any $p \in [0,1]$ it holds that:

$$\sum_{i=1}^{h} \overrightarrow{b}_{(i)}(p) \geq \sum_{i=1}^{h} \overrightarrow{a}_{(i)}(p), \quad \textit{for any } h = 1, \ldots, d, \tag{4}$$

Remark 5 *If $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$, condition (4) is equivalent to:*

$$|\overrightarrow{b}_1(p) - \overrightarrow{b}_2(p)| \leq |\overrightarrow{a}_1(p) - \overrightarrow{a}_2(p)|, \quad \forall p \in [0,1].$$

The remark establishes a relation with the criterion in (1) and provides an additional characterization of dissimilarity in terms of inequalities in groups composition at *any* proportion of the average population distribution. The result is useful for characterizing empirically-relevant indices of dissimilarity.

4.2 Dissimilarity indices

The criterion outlined in Corollary 1 offers an intuitive argument for constructing measures of dissimilarity that do not depend on the cardinalization of the outcome scale or the specific choice of the underlying ordinal comparable matrices. Claim (ii) offers, however, only a partial order of distribution matrices. In this section, we propose a novel family of dissimilarity measures consistent with the criterion in claim (ii) and with $\preccurlyeq^{\Delta}$. The family corresponds to linear rank-dependent evaluation functions $D_w(\mathbf{A})$, each of which is an average, taken over all proportions p, of the inequality displayed by vectors $\overrightarrow{\mathbf{a}}(p)$. Inequality is measured 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

$$D_w(\mathbf{A}) := \int_0^1 \sum_{i=1}^{d} w_i(p) \overrightarrow{a}_{(i)}(p) dp, \tag{5}$$

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 guarantees to focus on distributional concerns. The dissimilarity test proves to be a necessary and sufficient criterion for assessing agreement in dissimilarity evaluation among all indices belonging to the family D_w.

Corollary 2 *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *the following statements are equivalent:*

i $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$*;*

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

Weymark (1981) has 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. Following notation in Maccheroni, Muliere and Zoli (2005), the discrete counterpart of the S-Gini weights is obtained by setting

$$w_i(p) = \frac{1}{d \cdot p} \left[\frac{1}{d} - \left(\left(1 - \frac{i-1}{d}\right)^k - \left(1 - \frac{i}{d}\right)^k \right) \right]$$

with $k \geq 2$ a positive integer. When $k = 2$, the weights coincide with those of the Gini inequality index. As k grows, the index attributes increasing weight to the share of the most under-represented groups at any average groups share p. Inequalities across ps contribute uniformly to overall dissimilarity. In general, weights in $\mathcal{W}$ depend on p, thereby providing enough flexibility for assigning different relevance to dissimilarities in groups proportions occurring in correspondence of p-proportions of the average population. The family of indices D_w provides a representation of dissimilarity orderings consistent with axioms IEC, ISC, I and E. Corollary 2 provides the rationale for using D_w as a more discriminatory criterion for ranking configurations when $\preccurlyeq^{\Delta}$ is not capable of concluding.

5 Dissimilarity and evaluations of unfair inequality: Evidence from the Swedish education reform

Reducing unfair inequality is increasingly seen as the relevant social objective by experts and policymakers alike. Unfair inequality amounts to the extent of inequality that is attributable to circumstances of origin for which individuals cannot be held responsible for(for a review, see Roemer and Trannoy 2016, Ferreira and Peragine 2016). The dissimilarity criterion provides a useful tool to assess unfair inequality when a society consists of multiple circumstance groups and when evaluations have to be robust with respect to the metric that is used to define inequality, which can be measured in the space of incomes or for instance, alternatively, in the space of utility evaluations of incomes(as in Andreoli et al. 2019).

We make use of the dissimilarity criterion to investigate the implications of a large scale education reform on unfair income inequality in Sweden. The Swedish education reform increased compulsory education duration, abolished streaming after grade six and introduced a uniform national curriculum. The reform was gradually introduced across a selected group of Swedish municipalities in 1949 until 1962. Afterwards, the reform was gradually extended to the universe of municipalities.

The literature has focused on the average effects of the Swedish reform on earnings (Meghir and Palme 2005, Fischer, Karlsson, Nilsson and Schwarz 2019), education (Holmlund 2008), mortality (Lager and Torssander 2012) and health (Meghir, Palme and Simeonova 2018). Educational interventions have been shown to generate a leveling-the-playing-field effect on incomes across parental background groups (Andreoli et al. 2019). We investigate the consequences of the Swedish reform for unfair inequality by comparing counterfactual life-cycle earnings distributions under different policy regimes for a representative sample of Swedish people born around the period of implementation of the policy. Meghir and Palme (2005) provides an exhaustive description of the reform, the identification strategy and the data that are also used in our policy evaluation exercise.

	mean	sd	min	max	N
Outcome variable					
Life-cycle monthly earnings PPP (SEK)	1879.7	796.7	349.8	5239.8	22,644
Controls and treatment definition					
Cohort 1953 (Post)	0.477	0.499	0	1	22,644
Treatment municipality (Treat)	0.828	0.378	0	1	22,644
Always treated municipality	0.212	0.409	0	1	22,644
Always control municipality	0.172	0.378	0	1	22,644
Reform indicator (Reform)	0.563	0.496	0	1	22,644
Circumstances					
Female	0.469	0.499	0	1	22,644
Father education:					
Primary	0.814	0.389	0	1	22,644
Vocational	0.079	0.271	0	1	22,644
Secondary	0.069	0.254	0	1	22,644
Higher	0.037	0.190	0	1	22,644
High ability	0.513	0.500	0	1	22,644
Urban	0.166	0.372	0	1	22,644

Table 1: Table of descriptives for the using sample

5.1 Data and estimation

The sample cosnidered covers about 10% of the Swedish population born in 1948 and 1953.[9] The data consist of a balanced longitudinal sample of 22,644 boys and girls born in 1948 and 1953, for which income during adulthood is observed in 1985 through 1996, gathering a total of 209,155 observations. As an outcome, we use life-cycle monthly earnings in SEK at base year values. We obtain such estimates by averaging earnings observed at individual level over the 1985-1996 period, when cohorts 1948 and 1953 were about 40 years old. Summary statistics are reported in Table 1.

The cohort 1948 roughly corresponds to the group of individuals which were already completing compulsory education in the early stages of the implementation of the reform in 1962, and thus were likely experiencing the old compulsory education system in their municipality of residence. The cohort 1953 gathers instead individuals who were entering secondary education when the reform was already in place in most of Sweden. Table 1 shows that 47.7% of the sample is from the post-treatment cohort, whereas 82.2% of the sample is from municipalities that were offering the reformed education system (defining a treatment group) at some point. Not all municipalities adopted the reform at the same time. In the sample, 21.2% of observations lived in municipalities that adopted the reform early, whereas 17.2% of the sample lived in municipalities that never adopted the reformed

[9] Anonymized data are accessible online from the American Economic Association website.

system. All other cases correspond to individuals who lived in municipalities that switched towards the reformed education regime between cohort 1948 and 1953. We use these cases to identify a counterfactual distribution of earnings that would have prevailed in the absence of the reform.

Following Meghir and Palme (2005) and Holmlund (2008), identification rests on the quasi-random assignment of the reform across municipalities that switch into the reformed system over time. We identify such observations through a *reform indictor* R, which takes value $R = 1$ for those that were either in municipalities that were early implementers (hence always treated) irrespectively of the cohort of birth, or for those that were living in switching municipalities but were born in post-reform cohorts. Table 1 shows that 56.3% of the sample received the treatment ($R = 1$).

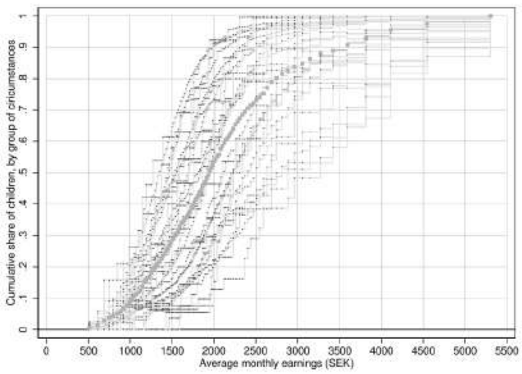
The differences between the distribution of earnings of those that received the treatment ($R = 1$) compared to those that did not ($R = 0$) is informative of the policy effects of the education reform, all else equal. We are interest in the distributional impacts of the reform along the lines of circumstances of origin. We consider $d = 32$ mutually exclusive circumstances groups denoted $C = 1, \ldots, 32$, each gathering observations with similar gender (two categories), father education (four categories), ability score collected while in education (two categories) and being born in Stockholm, Gotheborg or Malmo (the urban group, two categories). Our goal is to retrieve reliable estimates of the conditional distributions $F_{C,R}$, and then use these estimates to compare dissimilarity in $\{F_{C,R=1}\}_{C=1}^{32}$ to that in $\{F_{C,R=0}\}_{C=1}^{32}$.

We estimate conditional income distributions semi-parametrically using distributional regressions (see for instance Foresi and Peracchi 1995, Firpo et al. 2009, Chernozhukov et al. 2013). The underlying specification of the estimating model is:

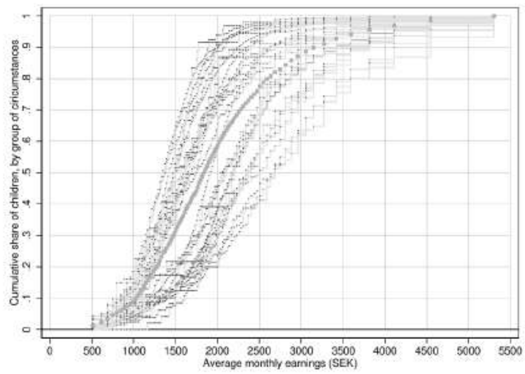
$$F_{C,R}(y) = \Pr[Y \leq y|C, R] = E[\pi(\mathbf{x}\boldsymbol{\theta}_y)|C, R], \tag{6}$$

where π is a binary choice model, $\mathbf{x}$ includes indicators for circumstance groups, an indicator of the reform status, fixed effects by municipality and cohort and their interactions, whereas $\boldsymbol{\theta}_y$ is a vector of parameters. Estimates are conditional to a non-stochastic income threshold y. We assume $\pi(\mathbf{x}\boldsymbol{\theta}_y)$ to be the linear probability model and we consider the following specification of the outcome equation:

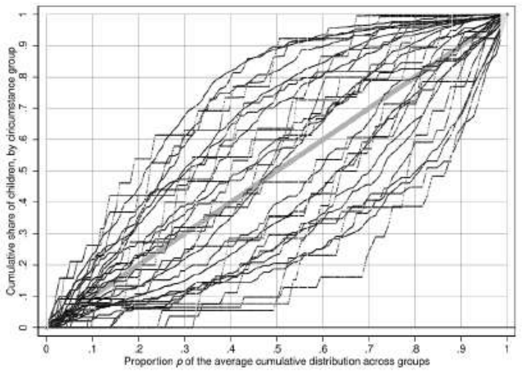
$$\mathbf{1}[y_{icm} \leq y] = \alpha_y + \mathbf{x}_i\boldsymbol{\beta}_y + (\gamma_y + \mathbf{x}_i\boldsymbol{\delta}_y) * R_i + \mathbf{x}_{cm}\boldsymbol{\phi}_y + \varepsilon_{icmy},$$



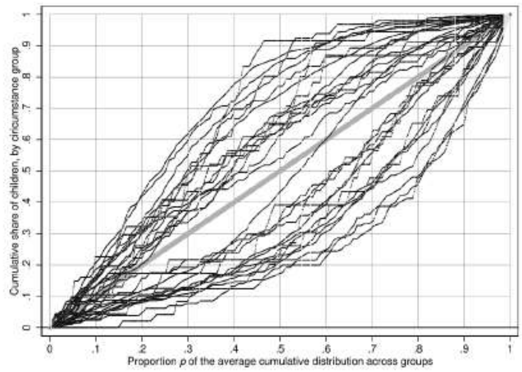
(a) Matrix $\overrightarrow{\mathbf{C}}$.



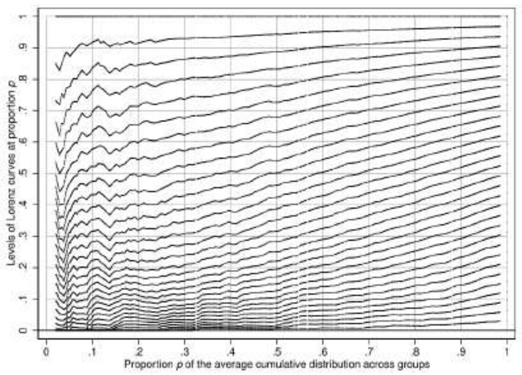
(b) Matrix $\overrightarrow{\mathbf{T}}$



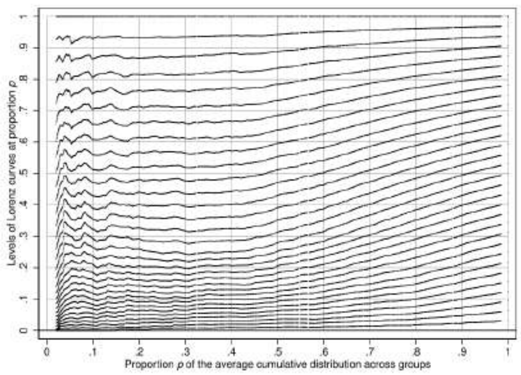
(c) $\overrightarrow{\mathbf{C}^*}$ (black dots)



(d) $\overrightarrow{\mathbf{T}^*}$ (black dots)



(e) $\sum_{i=1}^{h} \frac{1}{dp_j} \overrightarrow{c^*}_{(i)j}$



(f) $\sum_{i=1}^{h} \frac{1}{dp_j} \overrightarrow{t^*}_{(i)j}$

Figure 3: Cumulative groups distributions and their transformations, for control (left panels) and treatment (right panels) groups.

Note: Authors' computations based on Meghir and Palme (2005) data. Groups formed by interacting information on parents education, gender, ability and location. Gray curves in panel a) and b) are for average groups distributions.

where y_{icm} are permanent earnings of individual i from cohort c living in municipality m at the moment of the reform, R_i is the reform indicator for i, $\mathbf{x}_i$ includes indicators for 32 mutually exclusive circumstances groups and $\mathbf{x}_{cm}$ controls for fixed effects by cohort, municipality of birth, treatment group and for always treated municipalities. The estimates of γ_y provide evidence of the average effects of the policy. Estimates of $\boldsymbol{\delta}_y$ uncover evidence about the distributional impacts of the policy across circumstance groups. Estimated coefficients at income threshold y are collected in the vector $\hat{\boldsymbol{\theta}}_y = (\hat{\alpha}_y, \hat{\boldsymbol{\beta}}_y, \hat{\gamma}_y, \hat{\boldsymbol{\delta}}_y, \hat{\boldsymbol{\phi}}_y)$.

We estimate the binary choice model over a finite grid of $n = 100$ reference values $y \in \{y_1, \ldots, y_n\}$, corresponding to the percentiles of the unconditional sample distribution of lifetime earnings. Estimates of the model parameters in correspondence of the chosen grid, $\{\hat{\boldsymbol{\theta}}_{y_j}\}_{j=1}^n$, are then plugged into (6) to generate predictions of the conditional distributions $\{\hat{F}_{C,R}(y_j)\}_{j=1}^n$ at every step of the grid, for every group of circumstances C and for every policy regime R. The obtained point estimates of the conditional distributions can be linearly interpolated to obtain full estimates of the relevant cumulative distribution functions. Predictions are organized into two 32×100 matrices denoted respectively by $\mathbf{T}$ for the treatment situation (i.e., when $R = 1$) and $\mathbf{C}$ for the control situation (i.e., when $R = 0$). The elements of these matrices are $t_{hj} := \hat{F}_{C=h,R=1}(y_j) - \hat{F}_{C=h,R=1}(y_{j-1})$ and $c_{hj} := \hat{F}_{C=h,R=0}(y_j) - \hat{F}_{C=h,R=0}(y_{j-1})$ for any $h = 1, \ldots, 32$ and $j = 1, \ldots, 100$, where it is assumed that $\hat{F}_{C,R}(y_0) = 0$. Estimates of the groups conditional distributions $\overrightarrow{\mathbf{T}}$ and $\overrightarrow{\mathbf{C}}$ are depicted in Figure 3, panel a) for the control group and panel b) for the treatment group. Any difference between $\mathbf{T}$ and $\mathbf{C}$ should be attributed exclusively to the reform and hence has a causal interpretation.

5.2 Results

We use matrices $\mathbf{T}$ and $\mathbf{C}$ to test the null hypothesis $\mathbf{T} \preccurlyeq^{\Delta} \mathbf{C}$, that is, that the Swedish education reform has a causal impact in reducing unfair inequalities generated along the lines of circumstances of origin. The matrices $\mathbf{C}$ and $\mathbf{T}$ display inequality in two dimensions: first along the domain of realizations (captured by the position each units occupies on the sample distribution of income); second, across circumstances of origin. Unfair inequality originates from differences between the 32 groups at any realization.

Matrices $\mathbf{T}$ and $\mathbf{C}$ are not ordinal comparable. Testing for the dissimilarity criterion $\preccurlyeq^{\Delta}$ requires first to identify ordinal comparable matrices $\mathbf{T}^*$ and $\mathbf{C}^*$ (using the algorithm developed in Lemma 1) and then applying the test statistic $\Delta(h, p_j)$ to these estimates. Values of $\overrightarrow{t}^*_{ij}$ and $\overrightarrow{c}^*_{ij}$, issued from $\mathbf{T}^*$ and $\mathbf{C}^*$, are reported in panels c) and d) in Figure 3. Each dot in the figure represents one of the $32 \cdot 1,263 = 40,416$ estimates drawn from these matrices. In order to apply the $\Delta(h, p_j)$ statistic, we estimate for every p_j,

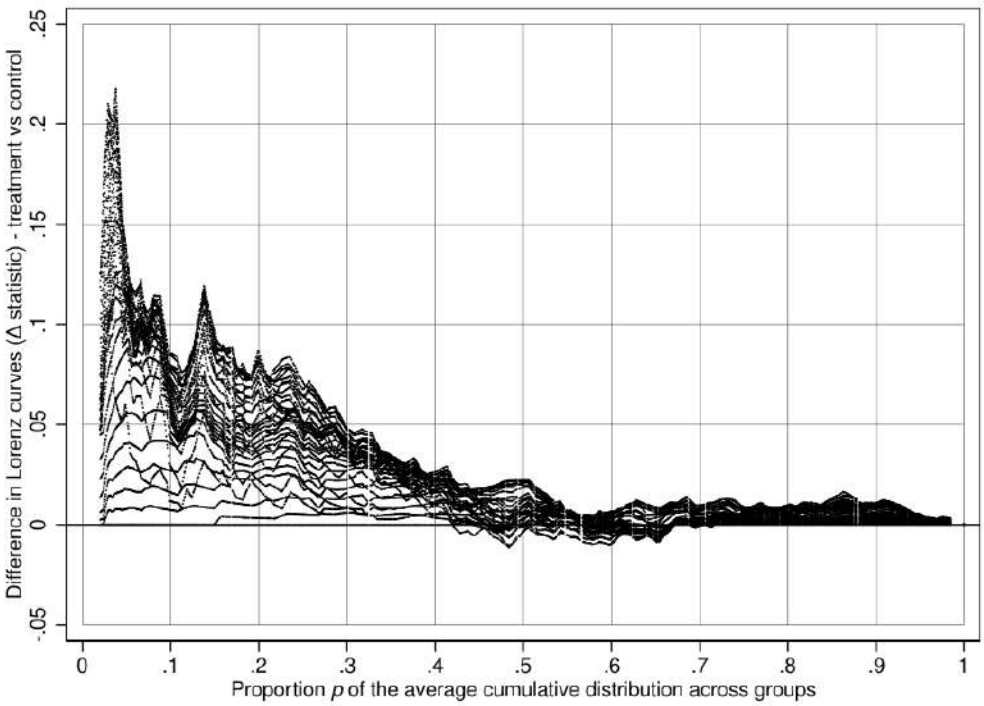


Figure 4: Empirical test for $\mathbf{T} \preccurlyeq^{\Delta} \mathbf{C}$.

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.

$j = 1, \ldots, 1236$ the relative Lorenz curve coordinates of vectors $\overrightarrow{\mathbf{t}}^*_j$ and $\overrightarrow{\mathbf{c}}^*_j$. These are obtained by ordering entries in each vector from the smallest to the largest and then taking the sequential sum of these estimates, namely computing for $\overrightarrow{\mathbf{t}}^*_j$ the values $\frac{1}{d \cdot p_j} \sum_{i=1}^{h} \overrightarrow{t}^*_{(i)j}$ for each $h = 1, 2, ..., 32$. The relative Lorenz curve coordinates are marked with small dots in panels e) and f) of Figure 3 for treated and control groups. In the absence of dissimilarity across cumulative distribution functions, dots in each panel should appear aligned along parallel, equally-spaced horizontal lines, indicating lack of inequality in groups cumulative proportions at any p. Such a configuration is clearly rejected by the data, revealing instead that relative inequalities in groups composition are stronger at the bottom of the average groups distribution than elsewhere, both in the treatment and control configurations. The pattern of the relative Lorenz curve coordinates vary slightly across the two configurations, with stronger evidence of inequality in group composition at the bottom of the population distributions in the control configuration. The $\Delta(h, p)$ statistic is useful to represent the direction and magnitude of such differences.

Values of the $\Delta(h, p_j)$ statistic are reported in Figure 5.2. The statistic takes on positive values almost everywhere. At the bottom of the average population distribution (for $p <$

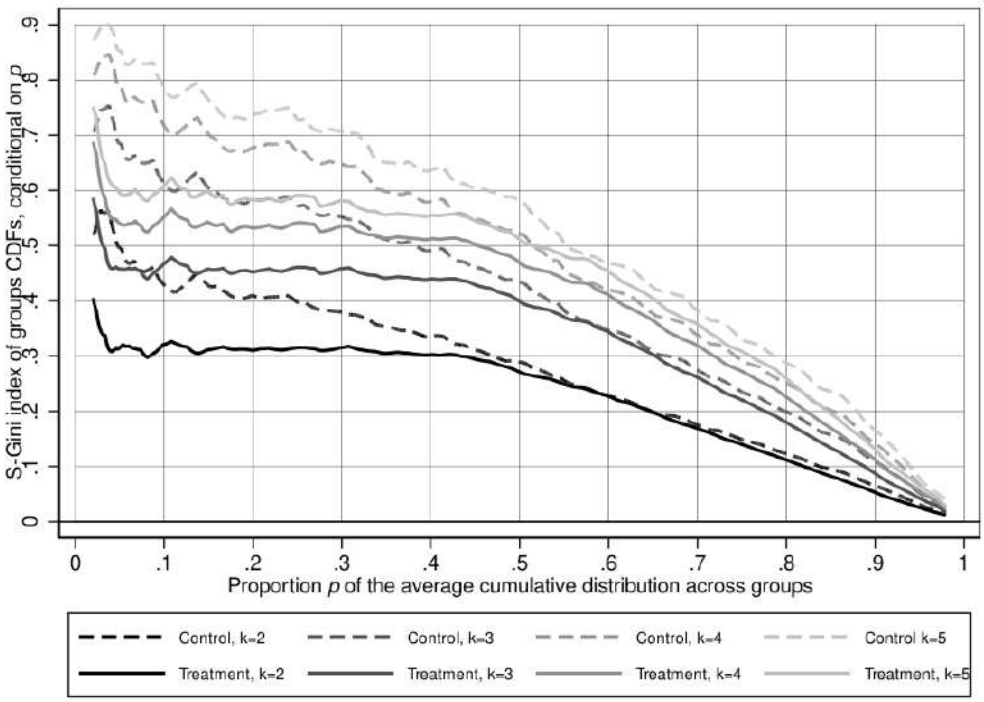


Figure 5: Comparing distributions by mean of S-Gini indices at selected proportions of the average groups' cdfs.

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 $\sum_{i=1}^{32} w_i \overrightarrow{t}^*_{(i)}(p)$ and $\sum_{i=1}^{32} w_i \overrightarrow{c}^*_{(i)}(p)$ for selected $p \in [0,1]$, where w_i is the S-Gini weighting function (parametrized by $k = 1, \ldots, 5$).

0.4), the magnitude of the gaps is relatively large, reaching a peak of 0.2 on the unit scale. Such an estimate originates from the pattern of Lorenz curves in Figure 3, where the coordinates ranging between the 18th to 28th position are, at the bottom of the domain p, at least 0.15 points larger (i.e. there is Lorenz dominance) in the treatment group compared to the control group. Conversely, the effects of the reform have been modest elsewhere. Violations of the empirical dissimilarity criterion are concentrated in the middle of the average probability distribution, implying an ambiguous effect of the reform for unfair inequality. Relative gaps are, nonetheless, small in size and possibly indistinguishable from zero from a statistical perspective.

The data reject the conclusion that $\mathbf{T} \preccurlyeq^{\Delta} \mathbf{C}$. Dissimilarity indices are useful to quantify the extent of inequality in groups cumulative distributions and to produce less partial evaluations that are consistent with the dissimilarity axiomatic model. Figure 5.2 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 aversion to the disproportional composition of groups. The graph provides compelling evidence that dissimilarity in groups distributions is smaller in the treatment setting compared to the control setting at any p, being the solid line never above the dashed line across levels of k. After aggregating these assessments into the dissimilarity index D_k (a specification of the index in (5) where the weighting function $w \in \mathcal{W}$ is parametrized by the S-Gini weights), we find evidence that the education reform has reduced dispersion across earnings profiles. In fact, the difference $D_k(\mathbf{T}) - D_k(\mathbf{C})$, which takes values $0.228 - 0.273$ for $k = 2$, $0.337 - 0.399$ for $k = 3$, $0.399 - 0.473$ for $k = 4$ and $0.439 - 0.521$ for $k = 5$, is always negative for reasonable selections of the inequality aversion parameter.

6 Concluding remarks

A large and sparse literature analyzing discrimination, mobility and unfair inequality has brought about criteria for ranking multi-groups distributions of an ordered outcome according to the dissimilarity they exhibit. This paper develops the normative underpinnings of robust dissimilarity comparisons with ordinal outcomes. The main result of the paper provides an empirical dissimilarity test which is capable of ranking configurations in a way that guarantees unanimous agreement among all possible ways of valuing dissimilarity consistently with a minimal set of relevant transformations.

The dissimilarity test that we invoke is not affected by the cardinality of the measurement scale of the outcome. The model only requires outcomes to be ordered in a meaningful and stable way. In fact, dissimilarity comparisons are concerned with inequalities *between* groups distributions and not with the way outcomes are unequally distributed *within* each group. This feature is of practical interest in cases where the underlying raw data have to be transformed through monotone, but not necessarily linear, transformations, thus distorting the cardinal interpretation attached to them.

We make use of the dissimilarity criterion to assess the distributional impact of the Swedish education reform on income. The reform has affected the treatment group in many complex ways, modifying the pattern of inequalities across individuals that differ in gender, family background, skills and place of birth. Reasonably, none of the revealed policy effects can be attributed to operations that intentionally exchange individuals across outcomes levels, or interchange groups labels.

Nonetheless, these "elementary" transformations might still be regarded as obviously reducing unfair inequality. If the "complex" features of the education reform reshape the students income opportunities in a way that is consistent with the existence of sequences of more "elementary" transformations, then the policymaker can safely conclude that the unfair inequality reduction objective has been attained. The dissimilarity criterion informs the policymaker that such a sequence of "elementary" transformations *exists* and is finite, thereby implying that every admissible evaluation of dissimilarity that is consistent with the implications of such transformations will also agree about the effect of the policy. Routines are made available to facilitate this task.

In our application, we are forced to reject that the Swedish reform reduces unfair inequalities, even if we find strong evidence in the direction of equalization at the bottom and at the top of the average distribution. Violations of the test mostly occur at the center of the distribution. When aggregating evaluations by mean of dissimilarity indices, we are bound to conclude that such violations are of lesser importance and overall, we find evidence that the reform has reduced dissimilarity for a large number of dissimilarity indices consistent with the axioms.

The results developed in this paper provide a complete characterization of dissimilarity through equivalent, yet separate, perspectives about dissimilarity, echoing results in the famous Hardy, Littlewood and Polya (1934) theorem for inequality analysis (see also Marshall et al. 2011, Andreoli and Zoli 2020). We show that the possibility of transforming a configuration $\mathbf{A}$ into $\mathbf{B}$ resorting exclusively to dissimilarity-preserving and -reducing operations generates agreement in ranking $\mathbf{B} \preccurlyeq \mathbf{A}$ among all orderings consistent with these transformations (statements (i) and (ii), Theorem 1), and in particular within the family of dissimilarity indices ranking $D_w(\mathbf{B}) \leq D_w(\mathbf{A})$ irrespectively of the choice of the weighting scheme w (statement (ii) in Corollary 2). Agreement can be empirically verified with the implementable dissimilarity test $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$ (statement (iii) in Theorem 1), which can be readily visualized as a comparison of dispersion among cumulative group distributions $\overrightarrow{\mathbf{a}}(p)$ and $\overrightarrow{\mathbf{b}}(p)$ (Claim (ii) in Corollary 1). Such a criterion provides an intuitive multi-group extension of the criterion (1) discussed in the Introduction.

A Proofs

A.1 Preliminary results

An algorith to obtain ordinal comparable matrices. Recall that a pair of distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ are said to be *ordinal comparable* whenever i) $n_A = n_B = n$ and $\mathbf{1}_d^t\mathbf{A} = \mathbf{1}_d^t\mathbf{B}$; ii) all groups are ordered according to stochastic dominance in $\mathbf{A}$ and $\mathbf{B}$, that is $\forall i, h$ either $\overrightarrow{a}_{ij} \leq \overrightarrow{a}_{hj}$ or $\overrightarrow{a}_{hj} \leq \overrightarrow{a}_{ij}$ for all $j = 1, \ldots, n_A$, and $\forall i', h'$ either $\overrightarrow{b}_{i'j} \leq \overrightarrow{b}_{h'j}$ or $\overrightarrow{b}_{h'j} \leq \overrightarrow{b}_{i'j}$ for all $j = 1, \ldots, n_B$; iii) the order of the groups is the same in $\mathbf{A}$ and $\mathbf{B}$, that is, if $\overrightarrow{a}_{ij} \leq \overrightarrow{a}_{hj}$ then $\overrightarrow{b}_{ij} \leq \overrightarrow{b}_{hj}$ for all groups $i \neq h$ and for all classes j.

Our first Lemma shows that the dissimilarity preserving operations allow to map any pair of distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ into ordinal comparable matrices $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$. Such matrices are ranked $\mathbf{A} \sim \mathbf{A}^*$ and $\mathbf{B} \sim \mathbf{B}^*$ by all dissimilarity orderings consistent with axioms IEC, ISC and I. The proof of the next result offers an *algorithm* through which any pair of distribution matrices can be transformed into a pair of ordinal comparable matrices.

Lemma 1 *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *there exist* $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ *ordinal comparable that are obtained from* $\mathbf{A}$ *and* $\mathbf{B}$ *respectively through operations of split of classes, insertion/elimination of empty classes and interchange of groups transformations.*

Proof We sketch an algorithm for transforming matrix $\mathbf{A}$ into $\mathbf{A}^*$. The same algorithm can be applied to $\mathbf{B}$ in order to obtain $\mathbf{B}^*$. Our procedure starts from the representation of the dissimilarity criterion based on the distribution functions $\overrightarrow{\mathbf{a}}(p)$ and $\overrightarrow{\mathbf{b}}(p)$.

Denote with $\mathbf{p}$ a realization of $\overrightarrow{\mathbf{a}}(p)$, so that $\mathbf{p} := \overrightarrow{\mathbf{a}}(p)$ where $p = \frac{1}{d}\mathbf{1}_d^t \cdot \mathbf{p}$. We consider two sets, denoted $\mathcal{S}_1$ and $\mathcal{S}_2$, whose elements are proportions p_j of average cumulative group distributions obtained from $\mathbf{A}$ and from the transformed matrix $\mathbf{A}^*$. (I) We define the first set as $\mathcal{S}_1 := \{p_j : p_j = \frac{1}{d}\mathbf{1}_d^t \cdot \overrightarrow{\mathbf{a}}_j, j = 1, \ldots, n_A\}$ with n_A denoting the number of classes of $\mathbf{A}$.

(II) To identify the second set we consider all the shares p and thus the corresponding vectors $\mathbf{p} := \overrightarrow{\mathbf{a}}(p)$ where it occurs a re-ranking of the proportions of the groups across $\mathbf{A}$'s classes. We illustrate here the procedure to derive these shares associated with the re-ranking involving two groups. Define the indices $j = 1, \ldots, n_A^c$ associated with points $\mathbf{p}_j$ that are *comonotonic*, i.e., such that for every group i, the element p_{ij} of $\mathbf{p}_j$ is ordered with respect to any other element $p_{i'j}$ of $\mathbf{p}_j$ with $i \neq i'$, in the same way as the element p_{ij+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 $\mathbf{p}_1 = \overrightarrow{\mathbf{a}}_1$. In order to identify $j = 2$, solve the following:

$$\mathbf{p}_2 := \arg\max_{\mathbf{p}} \left\{ \mathbf{1}_d^t \cdot \mathbf{p} : \mathbf{p} = \overrightarrow{\mathbf{a}}(\frac{1}{d}\mathbf{1}_d^t \cdot \mathbf{p}),\ \mathbf{1}_d^t \cdot \mathbf{p} > \mathbf{1}_d^t \cdot \mathbf{p}_1,\ \mathbf{p} \text{ is comonotonic to } \mathbf{p}_1 \right\}.$$

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 $\mathbf{p}_2 = \overrightarrow{\mathbf{a}}_{n_A} = \mathbf{1}_d$ and $n_A^c = 2$. Else, if $\overrightarrow{\mathbf{a}}(p_j)$ and $\overrightarrow{\mathbf{a}}(p_{j+1})$ are not comonotonic, then $\mathbf{p}_2$ identifies the smallest proportion $\delta \in [0, 1]$ such that $\overrightarrow{\mathbf{a}}(p_j)$ and $\delta\overrightarrow{\mathbf{a}}(p_j) + (1-\delta)\overrightarrow{\mathbf{a}}(p_{j+1})$ are comonotonic. Once δ is identified, then the corresponding vector of proportions is given by $\mathbf{p}_2 = \delta\overrightarrow{\mathbf{a}}(p_j) + (1-\delta)\overrightarrow{\mathbf{a}}(p_{j+1})$, which is such that $1 > \mathbf{1}_d^t \cdot \mathbf{p}_2 > \mathbf{1}_d^t \cdot \mathbf{p}_1$. We can then reiterate the procedure to derive $\mathbf{p}_3$, setting $\mathbf{p}_2$ as the reference.

Recursively, step j of the algorithm yields:

$$\mathbf{p}_j = \arg\max_{\mathbf{p}} \left\{ \mathbf{1}_d^t \cdot \mathbf{p} : \mathbf{p} = \overrightarrow{\mathbf{a}}(\frac{1}{d}\mathbf{1}_d^t \cdot \mathbf{p}),\ \mathbf{1}_d^t \cdot \mathbf{p} > \mathbf{1}_d^t \cdot \mathbf{p}_{j-1},\ \mathbf{p} \text{ is comonotonic to } \mathbf{p}_{j-1} \right\},$$

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 $\mathcal{S}_2 := \{p_j : \ p_j = \frac{1}{d}\mathbf{1}_d^t \cdot \mathbf{p}_j,\ j = 1, \ldots, n_A^c\}$.

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 $n_A \le n_A^* \le n_A + n_A^c$.

An analogous procedure identifies a set of proportions $\mathcal{S}_B := \{p_j : j = 1, \ldots, n_B^*\}$. 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 $\max\{n_A^*, n_B^*\} \le n^* \le n_A^* + n_B^* - 1$, $p_1 = \frac{1}{d}\min\{\overline{a}_1, \overline{b}_1\}$ and $p_{n^*} = 1$.

The sequence of indices $j = 1, \ldots, n^*$ and the associated p_j's yield the partition in n^* classes of the two ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ obtained from $\mathbf{A}$ and $\mathbf{B}$. We show that such matrices are obtained using exclusively dissimilarity preserving operations.

Consider first matrix $\mathbf{A}$. Note that for every $p_j \in \mathcal{S}$ there exists a $\mathbf{p}_j^A = \overrightarrow{\mathbf{a}}(\frac{1}{d}\mathbf{1}_d^t \cdot \mathbf{p}_j^A)$ such that $p_j = \frac{1}{d}\mathbf{1}_d^t \cdot \mathbf{p}_j^A$. The vector $\mathbf{p}_j^A$ can be obtained from the vectors associated with the classes of $\mathbf{A}$ through *splits and elimination/insertion of empty classes*. First, consider deleting all empty classes from $\mathbf{A}$. In what follows we assume that $\mathbf{A}$ has no empty classes.

Let $\mathcal{S}_1^A$ denote the set $\mathcal{S}_1$ computed for matrix $\mathbf{A}$. By construction we have that all $p_j \in \mathcal{S}_1^A$ coincide with $\frac{1}{d}\mathbf{1}_d^t \cdot \overrightarrow{\mathbf{a}}_h$ for $h = 1, \ldots, n_A$. We obtain therefore a first set of vectors $\mathbf{p}_j^A$ that coincide with the vectors $\overrightarrow{\mathbf{a}}_h$ for $h = 1, \ldots, n_A$. We focus then on the p_j's in $\mathcal{S}$

that do not belong to $\mathcal{S}_1^A$, that is, those in $\mathcal{S}\backslash\mathcal{S}_1^A$.

For any $p_j \in \mathcal{S}\backslash\mathcal{S}_1^A$, let $k \in \{1, \ldots, n_A\}$ be the index of a class of $\mathbf{A}$ such that $\frac{1}{d}\mathbf{1}_d^t \cdot \overrightarrow{\mathbf{a}}_k < p_j < \frac{1}{d}\mathbf{1}_d^t \cdot \overrightarrow{\mathbf{a}}_{k+1}$, then it can be re-written as $\mathbf{p}_j^A = \overrightarrow{\mathbf{a}}_k + \lambda \mathbf{a}_{k+1}$, which is a specific set of coordinates of $\overrightarrow{\mathbf{a}}(p)$. In this case, $\lambda \in (0, 1)$ can be interpreted as a split parameter. Iterating such procedure for every $p_j \in \mathcal{S}\backslash\mathcal{S}_1^A$, yields a sequence of split operations.

The sequence of vectors $\mathbf{p}_1^A, \ldots, \mathbf{p}_{n^*}^A$ computed for all the $p_j \in \mathcal{S}$ displays comonotonic elements, in the sense that $\forall p_j, p_{j+1} \in \mathcal{S}$, $\overrightarrow{\mathbf{a}}(p_j)$ and $\overrightarrow{\mathbf{a}}(p_{j+1})$ are comonotonic. However, it is possible that $\exists j$ such that $\mathbf{p}_j^A$ is not comonotonic with $\mathbf{p}_{j+k}^A$ 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 $\mathbf{p}_1^A$. 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 $j = 1, \ldots, n^*$.

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 $\mathbf{z}_0^A := \mathbf{0}_d$. The definition of $\mathbf{A}^*$ clarifies that the group permutations mapping vectors $\mathbf{p}_j^A$ 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 interchanges 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 involves 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 $\mathbf{B}^* := (\Delta\mathbf{z}_1^B, \ldots, \Delta\mathbf{z}_{n^*}^B)$. By construction $\mathbf{A}^*$ and $\mathbf{B}^*$ satisfy (i) $n_{A^*} = n_{B^*} = n^*$ and $\mathbf{1}_d^t \cdot \mathbf{A}^* = \mathbf{1}_d^t \cdot \mathbf{B}^*$; (ii) the group distributions in matrix $\mathbf{A}^*$ and in matrix $\mathbf{B}^*$ are ordered by stochastic dominance, and (iii) the order of the groups coincides in both matrices up to an independent permutations of the rows of the matrices. We conclude that $\mathbf{A}^*$ 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.

Q.E.D.

An algorithm showing the relation between exchange operations and the dissimilarity criterion for ordinal comparable matrices. We develop a rank-preserving version of Tchen's (1980) algorithm to show that the sequential majorization condition in statement (iii) in 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 $\mathbf{B}$ as claimed in statement (i) of the theorem. 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, implying that conditions *(ii)* and *(iii)* of Definition 1 might be violated by Tchen's algorithm.

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 according to first order stochastic dominance group $i-1$, for any i. That is, for $\mathbf{A} \in \mathcal{M}_d$, $\overrightarrow{a}_{ij} \leq \overrightarrow{a}_{i-1j}$, $\forall i, j$. 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 $(x, y) < (x', y')$ and is obtained either if $y < y'$ or if $y = y'$ and $x > x'$. We also use $i \in [x, x']$ to denote $i \in \{x, \ldots, x' : x < \ldots < x'\}$. We will say that the vector of cumulative groups shares $\overrightarrow{\mathbf{b}}_j$ corresponding to class j *Lorenz dominates* $\overrightarrow{\mathbf{a}}_j$ whenever $\sum_{i=1}^{h} \overrightarrow{b}_{ij} \geq \sum_{i=1}^{h} \overrightarrow{a}_{ij}$ for every h, a condition which builds on the fact that entries in $\overrightarrow{\mathbf{b}}_j$ and in $\overrightarrow{\mathbf{a}}_j$ are ordered in increasing magnitude and $\overrightarrow{\mathbf{b}}_j$ and in $\overrightarrow{\mathbf{a}}_j$ have the same average.

Denote the *doubly cumulative distribution matrix* of $\mathbf{A}$ by $\overrightarrow{\overrightarrow{\mathbf{A}}}$, with $\overrightarrow{\overrightarrow{a}}_{ij} = \sum_{x \geq i} \overrightarrow{a}_{xj}$. Using this compact notation, the Lorenz dominance condition applied to every class j rewrites equivalently as $\overrightarrow{\overrightarrow{\mathbf{B}}} \geq \overrightarrow{\overrightarrow{\mathbf{A}}}$.

Strategy of the proof. The algorithm is built in two steps that are illustrated respectively in Lemma 2 and Lemma 3. 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 $\overrightarrow{\overrightarrow{\mathbf{H}}} \leq \overrightarrow{\overrightarrow{\mathbf{H}'}}$ and $\overrightarrow{\overrightarrow{h}}_{xy} = \overrightarrow{\overrightarrow{h'}}_{xy}$ for all $(x, y) < (i, j)$, Lemma 2 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 $\overrightarrow{\overrightarrow{\mathbf{H}}} \leq \overrightarrow{\overrightarrow{\mathbf{K}}}^m \leq \overrightarrow{\overrightarrow{\mathbf{K}}}^{m+1} \leq \overrightarrow{\overrightarrow{\mathbf{H}'}}$ with element k^m_{ij} such that $h_{ij} < k^m_{ij} \leq h'_{ij}$. Lemma 2 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 $\overrightarrow{\overrightarrow{\mathbf{A}}} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$, 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 $q \in \{1, \dots, Q\}$, 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.[10] 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 $p \in \{1, \dots, P\}$. The two sequences together induce transfers that are bounded and guarantee that:

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

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 3.

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 $m \in \{1, \dots, M\}$ 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 $s \in \{1, \dots, S\}$ leading to $\mathbf{K}'$ from $\mathbf{K}$. We present a version of this subsequence that is also rank-preserving (i.e., that guarantees that $\overrightarrow{k}'_{ij} \geq \overrightarrow{k}'_{i+1\,j}, \forall i, j$).

[10] 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 $\mathbf{b}_j$.

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:

$$k_{ij} < h'_{ij} \text{ and } \overrightarrow{\overrightarrow{k}}_{xy} = \overrightarrow{\overrightarrow{h'}}_{xy} \text{ for all } (x, y) < (i, j), \tag{7}$$

$$\delta = \min\left\{ \overrightarrow{k}_{i-1j} - \overrightarrow{k}_{ij}, \overrightarrow{k}_{i^*j} - \overrightarrow{k}_{i^*+1j}, \frac{1}{2}(\overrightarrow{k}_{i^*j} - \overrightarrow{k}_{ij}) \right\} > 0. \tag{8}$$

Condition (7) is as in Tchen (1980), while condition (8) 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 gap that should be reduced), (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, $k_{x_s y} = 0\ \forall x_s$ and $\forall y \in [y_s, y_{s-1} - 1]$. Thus, conditions (iii) and (iv) are always satisfied when (8) holds. Altogether these conditions give:

$$\Delta := \min\left\{ h'_{ij} - k_{ij},\ \min_{S(\mathbf{K},\mathbf{H}'|i^*)} \{k^s_{x_s,y_s}\},\ \delta \right\}. \tag{9}$$

Lemma 2 *Let* $\mathbf{K}$ *satisfy conditions (7) and (8), 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*

(9).

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 (9), 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) and (x_3, y_3) and added to (i, y_3) and (x_3, y_2) giving $\mathbf{K}^3$. By (9), 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 $\overset{\rightrightarrows}{\mathbf{K}}{}^s \leq \overset{\rightrightarrows}{\mathbf{H}}{}'$ for any s, assume by recurrence that $\overset{\rightrightarrows}{\mathbf{K}}{}^{s-1} \leq \overset{\rightrightarrows}{\mathbf{H}}{}'$. 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 $\overset{\rightrightarrows}{k}{}^s_{iy_s} > \overset{\rightrightarrows}{k}{}^{s-1}_{iy_s}$ while $\overset{\rightrightarrows}{k}{}^s_{iy_s} < \overset{\rightrightarrows}{k}{}^s_{i-1\,y_s}$. 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 $\overset{\rightrightarrows}{k}{}^s_{xy} > \overset{\rightrightarrows}{k}{}^{s+1}_{xy}$. Finally, $k^s_{x_sy_s} < k^{s-1}_{x_sy_s}$ and $\overset{\rightrightarrows}{k}{}^s_{x_sy_s} = \overset{\rightrightarrows}{k}{}^{s-1}_{x_sy_s}$, as well as $\overset{\rightrightarrows}{k}{}^s_{xy_{s-1}} = \overset{\rightrightarrows}{k}{}^{s-1}_{xy_{s-1}}$ for $x \in [x_s, i]$. Combining these conditions, the required result is obtained. *Q.E.D.*

Under (7) and (8), the iteration of the sequence $S(\mathbf{K}, \mathbf{H}'|i^*)$ in Lemma 2 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 2 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 2 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. 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 $k^M_{ij} = h'_{ij}$ 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 2 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 3, 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), but not (iii), 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 $\overrightarrow{\overrightarrow{\mathbf{B}}} \geq \overrightarrow{\overrightarrow{\mathbf{A}}}$. Every transfer of mass is constructed in such a way that Lemma 2 always applies. Thus, each transfer breaks down into a finite number of exchange transformations.

Lemma 3 *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) $\overrightarrow{\overrightarrow{\mathbf{B}}} \geq \overrightarrow{\overrightarrow{\mathbf{A}}}$.*

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 the 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 $\overrightarrow{\overrightarrow{\mathbf{B}}} \geq \overrightarrow{\overrightarrow{\mathbf{A}}}$. 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 the 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

$$\Delta_{ij}(i^*) := \min\left\{\overrightarrow{b}_{ij} - \overrightarrow{a}^{q}_{ij},\ \frac{1}{i-i^*+1}\left(\overrightarrow{a}^{q}_{i^*j} - \overrightarrow{a}^{q}_{ij}\right)\right\}. \tag{10}$$

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 we have $\overrightarrow{a}^{q,1}_{i^*j} - (i-i^*)\Delta_{ij}(i^*) = \overrightarrow{a}^{q,p}_{i^*+1\,j} = \ldots = \overrightarrow{a}^{q,p}_{ij} < \overrightarrow{b}_{ij}$. 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

$$\Delta_{ij}(i^*, \widetilde{i}) := \min\left\{\overrightarrow{b}_{ij} - \overrightarrow{a}^{q}_{ij},\ \overrightarrow{a}^{q}_{\widetilde{i}j} - \overrightarrow{a}^{q}_{\widetilde{i}+1\,j},\ \frac{1}{i-\widetilde{i}}\left(\overrightarrow{a}^{q}_{i^*j} - \overrightarrow{a}^{q}_{i^*+1\,j}\right)\right\}. \tag{11}$$

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:

$$\overrightarrow{a}^{q,p}_{\widetilde{i}j} = \overrightarrow{a}^{q,p}_{\widetilde{i}+1\,j} = \ldots = \overrightarrow{a}^{q,p}_{ij} < \overrightarrow{b}_{ij}, \tag{12}$$

$$\overrightarrow{a}^{q,p}_{i^*j} - (i - \widetilde{i})\Delta_{ij}(i^*, \widetilde{i}) = \overrightarrow{a}^{q,p}_{i^*+1\,j}. \tag{13}$$

If (12) holds but (13) 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 (11), 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 (10) 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, (13) holds but (12) 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 (10) or by (11), 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 (7). The two matrices also satisfy condition (8) 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 2 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 2. 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 $a_{d1}^{1,P} = b_{d1}$; then set $\mathbf{A}^2 = \mathbf{A}^{1,P}$ and $(i,j) = (d-1,1)$ to obtain $\mathbf{A}^{2,P}$ with $\overrightarrow{\overrightarrow{\mathbf{A}}}^{2,P} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$, $a_{d1}^{2,P} = b_{d1}$ and $a_{d-1\,1}^{2,P} = b_{d-1\,1}$; and so on. *Q.E.D.*

A.2 Proof of Theorem 1.

Proof It is shown that $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$.

$(i) \Rightarrow (ii)$. Given $\mathbf{A}$ and $\mathbf{B}$, consider using insertion of empty classes, split of classes and interchange to obtain a pair of ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$. According to Lemma 1, the two matrices always exist and are ranked $\mathbf{A} \sim \mathbf{A}^*$ and $\mathbf{B} \sim \mathbf{B}^*$ by all orderings that are consistent with axioms ISC, IEC and I. Obtain now $\mathbf{B}^*$ making use of exchange transformations that map $\mathbf{A}^*$ into $\mathbf{B}^*$. Consequently, $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ by all orderings satisfying ISC, IEC, I and E. The intersection of these orderings gives rise to a transitive partial order ranking $\mathbf{B} \sim \mathbf{B}^* \preccurlyeq \mathbf{A}^* \sim \mathbf{A}$, which is $\mathbf{B} \preccurlyeq \mathbf{A}$ as in (ii).

$(ii) \Rightarrow (iii)$. If (ii) holds, then there exist ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ obtained from $\mathbf{A}$ and $\mathbf{B}$ respectively, through split of classes, insertion/elimination of empty classes and interchanges that are ranked $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ by all orderings consistent with axiom E. Consider a representation of such orderings based on indices $\sum_{j=1}^{n^*} \sum_{i=1}^{d} w_{ij} \overrightarrow{a}^*_{(i)j}$ where w_{ij} is a weighting scheme so that $w_{ij} \leq w_{i+1j}$ $\forall i,j$. To see that such representation is consistent with axiom E, consider $\mathbf{A}^*$ and an exchange operation that moves ε from group ℓ to h with $\ell > h$ involving adjacent classes j and $j+1$ of matrix $\mathbf{A}^*$. The net effect of the operation on measured dissimilarity is $(w_{hj} - w_{\ell j})\varepsilon \leq 0$ since $w_{hj} \leq w_{\ell j}$ for all j, implying that the measured dissimilarity always decreases by effect of an exchange operation. If (ii) holds, then every measure such as $\sum_{j=1}^{n^*} \sum_{i=1}^{d} w_{ij} \overrightarrow{a}^*_{(i)j}$ for any weighting scheme w_{ij} such that $w_{ij} \leq w_{i+1j}$ $\forall i,j$ ranks $\mathbf{B}^*$ at most as much dissimilar as $\mathbf{A}^*$, whatever the choice of the two ordinally comparable matrices. In particular, this holds when $\sum_{i=1}^{d} w_{ij} = 0$ and $w_{ik} = 0$ for any $k \neq j$ and for any i. Hence, according to this latter weighting functions comparisons should be made for each class j. That is, $\sum_{i=1}^{d} w_{ij} \overrightarrow{b}^*_{(i)j} \leq \sum_{i=1}^{d} w_{ij} \overrightarrow{a}^*_{(i)j}$ for all j should hold, or equivalently: $\sum_{i=1}^{d} (1 - w_{ij}) \overrightarrow{b}^*_{(i)j} \geq \sum_{i=1}^{d} (1 - w_{ij}) \overrightarrow{a}^*_{(i)j}$ $\forall j$ (since by construction $\sum_{i=1}^{d} \overrightarrow{b}^*_{(i)j} = \sum_{i=1}^{d} \overrightarrow{a}^*_{(i)j}$ from ordinal comparability). For any j, let $w_{ij} = 1 - d/h$ for $i = 1, 2, \ldots, h$ and $w_{ij} = 1$ for $i = h+1, \ldots, d$, for any $h = 1, \ldots, d-1$. The latter inequality in combination with $\sum_{i=1}^{d} \overrightarrow{b}^*_{(i)j} = \sum_{i=1}^{d} \overrightarrow{a}^*_{(i)j}$ gives $\frac{d}{h} \sum_{i=1}^{h} \overrightarrow{b}^*_{(i)j} \geq \frac{d}{h} \sum_{i=1}^{h} \overrightarrow{a}^*_{(i)j}$ $\forall h$ and $\forall j$, or equivalently $\sum_{i=1}^{h} \overrightarrow{b}^*_{(i)j} \geq \sum_{i=1}^{h} \overrightarrow{a}^*_{(i)j}$ $\forall h$ and $\forall j$, which is: $\overrightarrow{\mathbf{b}}^*_j$ Lorenz dominates $\overrightarrow{\mathbf{a}}^*_j$ $\forall j$. Since j can be any class, inequality evaluations can be separated across classes. Repeating through $j = 1, \ldots, n^*$ and computing $\Delta(h, p_j) = \frac{d}{h} \sum_{i=1}^{h} \overrightarrow{b}^*_{(i)j} - \frac{d}{h} \sum_{i=1}^{h} \overrightarrow{a}^*_{(i)j}$

always yields $\Delta(h, p_j) \geq 0$ at any i and j. This is claim *(iii)*.

(iii) $\Rightarrow$ *(i)*. Let $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$, which is equivalent to $\sum_{i=1}^{h} \overrightarrow{b}^{*}_{ij} \geq \sum_{i=1}^{h} \overrightarrow{a}^{*}_{ij}$ $\forall h$ and $\forall j = 1, \ldots, n^*$, that is $\overrightarrow{\overrightarrow{\mathbf{B}^*}} \geq \overrightarrow{\overrightarrow{\mathbf{A}^*}}$, for given $\mathbf{A}^*$ and $\mathbf{B}^*$. Lemma 3 guarantees that there exists a finite sequence of exchange operations mapping $\mathbf{A}^*$ into $\mathbf{B}^*$. Since the two matrices are obtained by construction from $\mathbf{A}$ and $\mathbf{B}$ respectively exclusively through split of classes, insertion/elimination of classes and interchanges (see Lemma 1) then condition (i) must hold. *Q.E.D.*

A.3 Proof of Corollary 1.

Proof *(i)* $\Rightarrow$ *(ii)*. From Theorem A.2 in Marshall et al. (2011, p.30), $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$ is equivalent to

$$conv\{\mathbf{\Pi}_d \overrightarrow{\mathbf{b}^*}_j : \mathbf{\Pi}_d \in \mathcal{P}_d\} \subseteq conv\{\mathbf{\Pi}_d \overrightarrow{\mathbf{a}^*}_j : \mathbf{\Pi}_d \in \mathcal{P}_d\}, \quad \text{for any } j = 1, \ldots, n^*, \tag{14}$$

where the *conv* operator indicates the *convex hull* and $\mathbf{\Pi}_d$ is a $d-$dimensional permutation matrix. Since $\mathbf{A}^*, \mathbf{B}^*$ are ordinal comparable then $\overrightarrow{\mathbf{a}^*}_j$ and $\overrightarrow{\mathbf{a}^*}_{j+1}$ are comonotone and $\overrightarrow{\mathbf{b}^*}_j$ and $\overrightarrow{\mathbf{b}^*}_{j+1}$ are also comonotone for any $j = 1, \ldots, n^*$. It follows that $\forall \alpha \in [0, 1]$ and $j = 1, \ldots, n^*$ it also holds that:

$$conv\{\alpha \mathbf{\Pi}_d \overrightarrow{\mathbf{b}^*}_j + (1-\alpha)\mathbf{\Pi}_d \overrightarrow{\mathbf{b}^*}_{j+1} : \mathbf{\Pi}_d \in \mathcal{P}_d\} \subseteq conv\{\alpha \mathbf{\Pi}_d \overrightarrow{\mathbf{a}^*}_j + (1-\alpha)\mathbf{\Pi}_d \overrightarrow{\mathbf{a}^*}_{j+1} : \mathbf{\Pi}_d \in \mathcal{P}_d\}. \tag{15}$$

Recall that $\overrightarrow{\mathbf{a}^*}_j = \overrightarrow{\mathbf{a}^*}(p_j)$ $\forall j$ with $p_j = \frac{1}{d}\mathbf{1}_d^t \overrightarrow{\mathbf{a}^*}_j$. It then follows that, by definition of $\overrightarrow{\mathbf{a}^*}(p)$,

$$\alpha \overrightarrow{\mathbf{a}^*}_j + (1-\alpha)\overrightarrow{\mathbf{a}^*}_{j+1} = \alpha \overrightarrow{\mathbf{a}^*}(p_j) + (1-\alpha)\overrightarrow{\mathbf{a}^*}(p_{j+1}) = \overrightarrow{\mathbf{a}^*}(\alpha p_j + (1-\alpha)p_{j+1}),$$

for any $\alpha \in [0, 1]$, for any p_j, p_{j+1}. The same condition holds for $\overrightarrow{\mathbf{b}^*}(p)$. Substituting in (15) yields:

$$\begin{aligned} conv\{\mathbf{\Pi}_d \overrightarrow{\mathbf{b}^*}(\alpha p_j + (1-\alpha)p_{j+1}) \quad &: \quad \mathbf{\Pi}_d \in \mathcal{P}_d\} \subseteq conv\{\mathbf{\Pi}_d \overrightarrow{\mathbf{a}^*}(\alpha p_j + (1-\alpha)p_{j+1}) : \mathbf{\Pi}_d \in \mathcal{P}_d\}, \\ \text{for any } j \quad &= \quad 1, \ldots, n^*, \end{aligned}$$

which can be written as $\sum_{i=1}^{h} \overrightarrow{b}_{(i)}(p) \geq \sum_{i=1}^{h} \overrightarrow{a}_{(i)}(p)$ for $p := \alpha p_j + (1-\alpha)p_{j+1}$, $\forall h$ and $\forall j$, given that $\overrightarrow{b}_{(i)}(p) = \overrightarrow{b}^{*}_{(i)}(p)$ and $\overrightarrow{a}_{(i)}(p) = \overrightarrow{a}^{*}_{(i)}(p)$, which is (ii).

(ii) $\Rightarrow$ *(i)*. If (ii) holds, then consider testing condition (ii) at pre-determined thresholds p_j for $j = 1, \ldots, n^*$. Obtain p_j such that $p_j = \frac{1}{d}\sum_i \overrightarrow{a}^{*}_{ij} = \frac{1}{d}\sum_i \overrightarrow{b}^{*}_{ij}$ where $\mathbf{A}^*$ and $\mathbf{B}^*$

are two ordinal comparable matrices of size $d \times n^*$ issued from $\mathbf{A}$ and $\mathbf{B}$ respectively. This condition is sufficient to grant $\Delta(h, p_j) \geq 0$ for all i and j, which is (i). *Q.E.D.*

A.4 Proof of Corollary 2

Proof $(i) \Rightarrow (ii)$. As shown in Lemma 1, for any pair $\mathbf{A}, \mathbf{A}^* \in \mathcal{M}_d$ such that $\mathbf{A}^*$ is obtained from $\mathbf{A}$ through insertion/deletion of empty classes, split of classes and interchanges (including permutations of groups as a special case of interchange), then $\overrightarrow{a^*}_{(i)}(p) = \overrightarrow{a}_{(i)}(p)$ for every $p \in [0, 1]$. Moreover, from Corollary 1, $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$ is equivalent to $\overrightarrow{\mathbf{b}}(p)$ Lorenz dominates $\overrightarrow{\mathbf{a}}(p)$ $\forall p \in [0, 1]$. By definition, the index D_w is invariant to split of classes, insertion/deletion of empty classes and interchange transformations. 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 $p_j := \frac{1}{d}\mathbf{1}_d^t \cdot \overrightarrow{\mathbf{a}}_j$ 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 any p. Thus, the index is consistent as well with the implications of exchange operations. This verifies that (ii) must be true.

$(ii) \Rightarrow (i)$. Recall that from Corollary 1, condition *(i)* is equivalent to $\sum_{i=1}^{h} \overrightarrow{b}_{(i)}(p) \geq \sum_{i=1}^{h} \overrightarrow{a}_{(i)}(p)$ 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 *(i)* does not hold then also claim *(ii)* 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 *(i)* 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 $\sum_{i=1}^{h^*} \overrightarrow{b}_{(i)}(p) - \sum_{i=1}^{h^*} \overrightarrow{a}_{(i)}(p) < 0$ for all $p \in (q_L, q^H)$. Denote $\Delta_{(i)}(p) := \overrightarrow{b}_{(i)}(p) - \overrightarrow{a}_{(i)}(p)$, then the condition can be rewritten as $\sum_{i=1}^{h^*} \Delta_{(i)}(p) < 0$ for all $p \in (q_L, q^H)$. 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 $\sum_{i=1}^{h^*} \Delta_{(i)}(p) < 0$ for all $p \in (q_L, q^H)$, it follows that $D_w(\mathbf{B}) - D_w(\mathbf{A}) > 0$, thereby violating claim *(ii)*. This establishes by contradiction that $(ii) \Rightarrow (i)$. *Q.E.D.*

A.5 An example illustrating the dissimilarity test

We provide an application of the criterion $\preccurlyeq^{\Delta}$ for comparing $\mathbf{A}$ in (2), which is 3×4 in size, to another matrix $\mathbf{B}$ which is 3×5 in size, with the following entries:

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.1 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0 & 0.5 \\ 0.1 & 0.1 & 0.6 & 0.2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0.3 & 0.2 & 0.3 & 0.05 & 0.15 \\ 0.2 & 0.3 & 0 & 0.35 & 0.15 \\ 0.1 & 0.1 & 0.6 & 0.05 & 0.15 \end{pmatrix}.$$

The two matrices are not ordinal comparable: they exibith a different number of classes, thus margins do not coincide, morevoer groups are not ordered. To achieve ordinal comparability, consider splitting class 3 of both matrices using a splitting parameter 1/2. Then interchange groups 2 and 3 distributions from that class onward in matrix $\mathbf{B}$. Moreover, split class 4 of $\mathbf{A}$ using a parameter 1/2. This sequence of operations leads to $\mathbf{A}^*$ and $\mathbf{B}^*$ which are 3×6 in size and are ordinal comparable. The obtained matrices are:

$$\mathbf{A}^* = \begin{pmatrix} 0.4 & 0.1 & 0.15 & 0.15 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0 & 0 & 0.25 & 0.25 \\ 0.1 & 0.1 & 0.3 & 0.3 & 0.1 & 0.1 \end{pmatrix}, \quad \mathbf{B}^* = \begin{pmatrix} \underline{0.3} & \underline{0.2} & 0.15 & 0.15 & \underline{0.05} & \underline{0.15} \\ \underline{0.2} & \underline{0.3} & 0 & 0.3 & \underline{0.05} & \underline{0.15} \\ 0.1 & 0.1 & 0.3 & 0 & \underline{0.35} & \underline{0.15} \end{pmatrix}. \tag{16}$$

The average cumulated population shares across the six classes are $(0.2, 0.4, 0.55, 0.7, 0.85, 1)$ in both matrices.

In order to test for dominance according to the criterion $\preccurlyeq^{\Delta}$, consider the Lorenz curves coordinates obtained from columns of the matrices $\mathbf{A}^*$ and $\mathbf{B}^*$. For matrix $\mathbf{A}^*$, one has that for $p = 0.2$ (class 1), $\sum_i \frac{1}{3 \cdot 0.2} \overrightarrow{a}^*_{(i)1}$ takes values 1/6 for $h = 1$, 2/6 for $h = 2$ and 1 for $h = 3$, whereas for $p = 0.4$ (class 2), $\sum_i \frac{1}{3 \cdot 0.4} \overrightarrow{a}^*_{(i)2}$ takes values 1/6 for $h = 1$, 7/12 for $h = 2$ and 1 for $h = 3$, and so on for the other classes. Coordinates of the Lorenz curve $\sum_i \frac{1}{3 \cdot p_j} \overrightarrow{b}^*_{(i)j}$ $\forall j$ are obtained in the same way.

The dissimilarity test requires verifying that $\mathbf{\Delta}(p_j) = (\Delta(1, p_j), \Delta(2, p_j), \Delta(3, p_j))^t \geq \mathbf{0}_3$ for all $j = 1, \ldots, 6$. We find: $0.6 \cdot \mathbf{\Delta}(0.2) = (0.1 - 0.1, 0.3 - 0.2, 0.6 - 0.6)^t \geq \mathbf{0}_3$, $1.2 \cdot \mathbf{\Delta}(0.4) = 1.65 \cdot \mathbf{\Delta}(0.55) = 2.1 \cdot \mathbf{\Delta}(0.7) = \mathbf{0}_3$ and $2.55 \cdot \mathbf{\Delta}(0.85) = (0.85 - 0.75, 1.7 - 1.65, 2.55 - 2.55)^t \geq$

$\mathbf{0}_3$. As expected, the dissimilarity criterion supports dominance and $\mathbf{B} \preccurlyeq^{\Delta} \mathbf{A}$ is verified.

From Theorem 1 we conclude $\mathbf{B} \preccurlyeq \mathbf{A}$ for all dissimilarity orderings consistent with axioms *ISC*, *IEC*, *I*, *E*. In fact, matrix $\mathbf{B}^*$ is obtained from $\mathbf{A}^*$ through a sequence of exchange operations that involve the numbers underlined in (16).

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.

Andreoli, F. and Zoli, C. (2022). Robust dissimilairty comparisons with categorical outcomes, *Social Choice and Welfare* **forthcoming**.

Asplund, E. and Bungart, L. (1966). *A First Course in Integration*, Holt Rinehart & Winston, New York.

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.

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.

Chakravarty, S. R. and Dutta, B. (1987). A note on measures of distance between imcome distributions, *Journal of Economic Theory* **41**(1): 185 – 188.

Chernozhukov, V., Fernandez-Val, I. and Melly, B. (2013). Inference on counterfactual distributions, *Econometrica* **81**(6): 2205–2268.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States, *The Quarterly Journal of Economics* **129**(4): 1553–1623.

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.

Faure, M. and Gravel, N. (2021). Reducing inequalities among unequals, *International Economic Review* **62**(1): 357–404.

Ferreira, F. H. G. and Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and an application to Latin America, *Review of Income and Wealth* **57**(4): 622–657.

Ferreira, F. H. G. and Peragine, V. (2016). *Individual Responsibility and Equality of Opportunity*, Vol. The Oxford Handbook of Well-Being and Public Policy, Oxford University Press, chapter 25.

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

Firpo, S., Fortin, N. M. and Lemieux, T. (2009). Unconditional quantile regressions, *Econometrica* **77**(3): 953–973.

Fischer, M., Karlsson, M., Nilsson, T. and Schwarz, N. (2019). The Long-Term Effects of Long Terms Compulsory Schooling Reforms in Sweden, *Journal of the European Economic Association* **18**(6): 2776–2823.

Fleurbaey, M. (2008). *Fairness, Responsability and Welfare*, Orxford University Press.

Foresi, S. and Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis, *Journal of the American Statistical Association* **90**(430): 451–466.

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.

Jäntti, M. and Jenkins, S. P. (2015). Chapter 10 - income mobility, *in* A. B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution*, Vol. 2 of *Handbook of Income Distribution*, Elsevier, pp. 807 – 935.

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.

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

Magdalou, B. and Nock, R. (2011). Income distributions and decomposable divergence measures, *Journal of Economic Theory* **146**(6): 2440 – 2454.

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.

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

Roemer, J. E. and Trannoy, A. (2016). Equality of opportunity: Theory and measurement, *Journal of Economic Literature* **54**(4): 1288–1332.

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). Inequality decomposition by factor components, *Econometrica* **50**(1): 193–211.

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.