Working Paper Series
Department of Economics
University of Verona

Measuring Dissimilarity

Francesco Andreoli, Claudio Zoli

WP Number: 23 December 2014

ISSN: 2036-2919 (paper), 2036-4679 (online)

Measuring dissimilarity*

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

[a] *University of Verona, DSE, Via dell'Artigliere 19, I-37129 Verona, Italy*
[b] *CEPS/INSTEAD, 3 avenue de la Fonte, L4364 Esch-sur-Alzette, Luxembourg*

December 31, 2014

Abstract

The analysis of many social phenomena requires to compare distributions of outcomes achieved by individuals belonging to different social groups, defined for instance by their gender, ethnicity, birthplace, education, age or parental background. When the groups are similarly distributed across classes of realizations, their members have equal chances to achieve any of the attainable outcomes. Otherwise, a form of dissimilarity prevails. We frame dissimilarity comparisons of sets of groups distributions by showing the equivalence between axioms underpinning information criteria, majorization conditions, agreement between dissimilarity indicators and new empirical tests based on Zonotopes and Path Polytopes inclusion. Multi-group comparisons of segregation, discrimination and mobility, as well as inequality evaluations, are embedded within the dissimilarity model.

Keywords: Dissimilarity, informativeness, majorization, Zonotopes.

JEL Codes: J71, D30, D63, C10.

*We are grateful for comments to participants at the 2011 ECINEQ meeting, LAGV 2012, 2012 SSCW meeting, 2012 EEA-ESEM meeting, and seminars at IT2012 Winter School, Aix-Marseille University, THEMA, Palermo University, CORE, Bocconi University, Leicester University, UABarcelona, CERDI/FERDI. Francesco Andreoli acknowledges the financial support by FNR Luxembourg (AFR grant 5932132, FP7-COFUND scheme). An earlier version of this research appeared as ECINEQ wp 2012-274 "On the measurement of dissimilarity and related orders".

†Contacts: francescondrl@gmail.com (F. Andreoli) and claudio.zoli@univr.it (C. Zoli).

1 Introduction

The growing concern of modern policy agenda for social inclusion and equality of opportunity objectives (Atkinson and Marlier 2010) motivates policymakers and researchers to bring about coherent criteria for assessing how overall inequality is shaped by the division of a society into *groups*, gathering together people with common characteristics like gender, ethnicity, birthplace, education, age or parental background. The intuition behind this perspective is that groups distributional heterogeneity across *classes of realizations* might be acceptable, and sometimes even desirable, if it reflects individual free choices. However, dissimilar groups distributional patterns reveal that the group identity plays indeed a role in explaining individual realizations, a situation that may call for compensation.

The literature on the analysis and measurement of segregation, discrimination, inequality of opportunity and mobility has developed specific criteria to evaluate the heterogeneity between groups distributional patterns. We show that all these different perspectives are embedded within a common *multi-group dissimilarity* measurement framework.

The notion of similarity between distributions dates back to the work of Gini (1914, p. 189), who concludes that two distributions (denoted "groups") with identical discrete support (whose realizations are denoted "classes") are *similar* when "the overall populations of the two groups take the same values with the same [relative] frequency."[1] The same logic can be applied to define similarity in the multi-group context. As in Gini (1914), every set of distributions that does not satisfy the similarity condition is said to display some degree of dissimilarity.

This paper develops an axiomatic perspective grounded on the notion of *informativeness* (Blackwell 1953, Torgersen 1992, Grant, Kajii and Polak 1998) to rank pairs of sets of distributions in an "objective" manner, i.e. depending only on a restricted set of intuitive and compelling transformations that, when applied to the data, are assumed to reduce or preserve the overall dissimilarity. These transformations guarantee that the less dissimilar configuration is the one where the knowledge of the group membership is *less informative* on the pattern of realizations attained by the groups. We characterize *model free* dissimilarity evaluations that are coherent with these transformations. Our results can be applied to make comparisons of two- and multi-group distributions across ordered or permutable

[1] The text, translated from Italian, proceeds with a very simple formalization of similarity: "If n is the size of group α, m is the size of group β, n_x the size of group α which is assigned to class x and m_x the size of group β assigned to the same class, then it should hold [under similarity] that, for any value of x, $\frac{n_x}{m_x} = \frac{n}{m}$."

classes, representing qualitative realizations of an underlying outcome variable.

In the spirit of the well-known inequality measurement results by Hardy, Littlewood and Polya (1934), we present two theorems that characterize dissimilarity partial orders when classes are permutable or, alternatively, when classes can be meaningfully ordered. Both theorems illustrate the equivalence of the agreement among dissimilarity orders that are consistent with dissimilarity preserving and reducing transformations, and a more compact representation based on majorization criteria. Moreover, they identify empirical tests for these criteria that are based on Zonotopes and Path Polytopes inclusion.[2] These tests can be implemented making use of standard linear algebra transformations of the data.[3] These two theorems shed lights on sparse, and apparently unconnected results on segregation measurement, decision theory and majorization analysis, and develop different perspectives about dissimilarity comparisons. They clarify how these different perspectives are all connected, and embedded within the informativeness framework, illustrating how to test for changes in informativeness. The novel result of this paper is that all these perspectives express *equivalent* characterizations of the dissimilarity between two or more distributions. The equivalence relations in the ordered case are new. When classes are permutable, only few of the relations where known, mostly in the case of two groups.

The dissimilarity criterion provides a unified framework for the analysis of multi-group segregation, discrimination and mobility, among others. For instance, in a *non-segregated* society, the groups are similarly distributed across non-ordered classes, usually referring to their residential location, their school assignment or their occupational type.[4] Segregation instead arises when the distributions of the groups across neighborhoods, schools or jobs are dissimilar.

With ordered classes, configurations with dissimilar groups distributions anticipate a problem of *discrimination*,[5] arising because the groups have different chances of experiencing good or bad realizations.

[2]See Dahl (1999), Koshevoy and Mosler (1996), McMullen (1971) and Ziegler (1995) for an extended analysis of the geometric properties of these objects.

[3]A user-written Stata routine for checking Zonotopes and Path Polytopes set inclusion is available from the authors upon request.

[4]See Duncan and Duncan (1955), Hutchens (1991), Reardon and Firebaugh (2002), Reardon (2009), Flückiger and Silber (1999), Chakravarty and Silber (2007), Alonso-Villar and del Rio (2010) and Frankel and Volij (2011), for a survey on the methodology. For an analysis of the economic motivations behind segregation comparisons, see for instance Echenique, Fryer and Kaufman (2006) and Borjas (1995).

[5]For a formal treatment see Butler and McDonald (1987), Jenkins (1994), Le Breton, Michelangeli and Peluso (2012) and Fusco and Silber (2013). Lefranc, Pistolesi and Trannoy (2009) and Roemer (2012) refer more explicitly to equality of opportunity.

Also *intergenerational equity assessments* come down to how much dissimilarity exists in the chances of reaching a given class of destination from different groups of origin.[6] In this situation, both groups and classes are ordered, and are often associated to the quantiles of the income distribution of departure and of destination.

Even *inequality* comparisons involve specific types of dissimilarity evaluations, although the converse is not necessarily true. In fact, income inequality prevails if the distribution of income shares across demographic units (such as individuals, families or groups) is dissimilar from the distribution of these units demographic weights. By treating income shares and weights as two distinct distributions across income recipients, we offer a new perspective on income inequality comparisons based on dissimilarity. Similarity, and therefore equality, is reached if every unit receives an income whose size is proportional to her weight. Multivariate extensions are straightforward.[7]

After defining the setting, we investigate dissimilarity orders for permutable and ordered classes in two separate sections. Our main contributions are summarized in two theorems, one for each section. We conclude illustrating the implications of the results for inequality, segregation, discrimination, mobility and distance measurement. All proofs are collected in the Appendix.

2 Setting

2.1 Notation

This paper deals with comparisons of $d \times n$ *distribution matrices*, depicting the relative frequencies distribution of $d \geq 1$ *groups* (indexed by rows) across $n \geq 2$ disjoint *classes* (indexed by columns). The set of distribution matrices with a fixed number d of rows, but possibly variable number of classes, representing the *data*, is:

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

[6]Examples are Dardanoni (1993) and Van de gaer, Schokkaert and Martinez (2001).

[7]See Hardy et al. (1934) and Marshall, Olkin and Arnold (2011), Kolm (1969, 1977), Atkinson and Bourguignon (1982) and Koshevoy and Mosler (1996) for traditional results and multivariate extensions. Ebert and Moyes (2003) treats the case of non-uniform weighting schemes.

where a_{ij} represents the share of group i in class j and $\mathbf{a}_j$ denotes the column vector associated to class j. The set $\mathcal{M}_d$ gathers the distribution matrices that are *stochastic in the rows*, meaning that matrix $\mathbf{A} \in \mathcal{M}_d$ represents a collection of d discrete distributions on the unit simplex Δ^{n_A}. The *cumulative distribution matrix* $\overrightarrow{\mathbf{A}} \in \mathbb{R}_+^{d,n_A}$ is constructed by sequentially cumulating the elements of the classes of $\mathbf{A}$, so that $\overrightarrow{\mathbf{a}}_k := \sum_{j=1}^{k} \mathbf{a}_j$. We also consider *transformation matrices*, representing the transformations that can be applied to the data. These matrices belong either to the set $\mathcal{P}_n$ of $n \times n$ permutation matrices, or to the set $\mathcal{R}_{n,m}$ of $n \times m$ row stochastic matrices whose rows lie in Δ^m.[8] The set of transformation matrices such that $m = n$ is $\mathcal{R}_n$, while $\mathcal{D}_n \subseteq \mathcal{R}_n$ denotes the set of *doubly stochastic* matrices whose rows *and* columns lie in Δ^n.[9] Finally, boldface letters always indicate column vectors, with $\mathbf{e}_n^t := (1, \ldots, 1)$ and $\mathbf{0}_n^t := (0, \ldots, 0)$, where the superscript t denotes transposition.

2.2 Dissimilarity orders

The cases of perfect similarity and maximal dissimilarity can be formalized in matrix notation. The *perfect similarity* matrix $\mathbf{S}$ represents a situation where all groups' distributions coincide across classes and are denoted by the same row vector $\mathbf{s}^t \in \Delta^n$. The *maximal dissimilarity* matrix $\mathbf{D}$ represents instead situations where each class is occupied at most by a group *and* each group occupies adjacent classes. Thus the groups distributions $\mathbf{d}_1^t \in \Delta^{n_1}, \ldots, \mathbf{d}_d^t \in \Delta^{n_d}$ do not overlap across classes. In compact notation:

$$\mathbf{S} := \begin{pmatrix} \mathbf{s}^t \\ \vdots \\ \mathbf{s}^t \end{pmatrix} \quad \text{and} \quad \mathbf{D} := \begin{pmatrix} \mathbf{d}_1^t & \ldots & \mathbf{0}_{n_d} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{n_1} & \ldots & \mathbf{d}_d^t \end{pmatrix}.$$

A *dissimilarity ordering* is a complete and transitive binary relation $\preccurlyeq$ on the set $\mathcal{M}_d$, with symmetric part $\sim$, that ranks $\mathbf{B} \preccurlyeq \mathbf{A}$ whenever $\mathbf{B}$ is *at most as dissimilar as* $\mathbf{A}$.[10] Any dissimilarity ordering should rank $\mathbf{S} \preccurlyeq \mathbf{A} \preccurlyeq \mathbf{D}$ for any $\mathbf{A} \in \mathcal{M}_d$.

[8] An element $\mathbf{X} \in \mathcal{R}_{n,m}$ can be interpreted as a *migration matrix* whose entry x_{ij} is the probability that the population in class i in the distribution of origin "migrates" to class j in the distribution of destination.

[9] Note that $\mathcal{R}_{d,n} \subseteq \mathcal{M}_d$ because both sets consider row stochastic matrices, but $\mathcal{M}_d$ does not impose restrictions on the number of columns.

[10] 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 $\mathbf{B} \sim \mathbf{A}$.

We will characterize the partial orders induced by the intersection of the dissimilarity orderings (Donaldson and Weymark 1998) satisfying desirable properties. Before proceeding to the main results, we present in the next section the geometric criteria that will be proposed to verify empirically the rankings produced by these partial orders.

2.3 Dissimilarity empirical criteria

Given a matrix $\mathbf{A} \in \mathcal{M}_d$, the Monotone Path $MP^*(\mathbf{A}) \subseteq [0,1]^d$ is a graphical arrangement of n_A line segments starting from the origin of the positive orthant, and sequentially connecting the points with coordinates given by the columns of $\overrightarrow{\mathbf{A}}$. The order of the vertices of the Monotone Path $\mathbf{v}_j \in MP^*(\mathbf{A})$ with $j = 0, 1, \ldots, n_A$ coincides with the one of the classes of $\mathbf{A}$, so that $\mathbf{v}_j = \overrightarrow{\mathbf{a}}_j$, $\mathbf{v}_0 = \mathbf{0}_d$ and $\mathbf{v}_{n_A} = \mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{e}_d$.

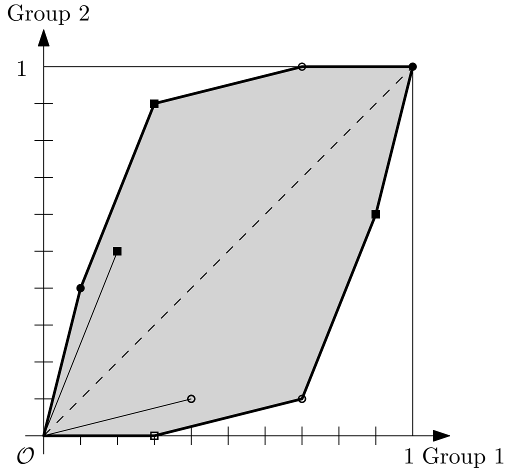
The Monotone Path associated to two groups distributions can be easily represented on a graph. Consider, for instance, the data in matrix $\mathbf{A} \in \mathcal{M}_2$:

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

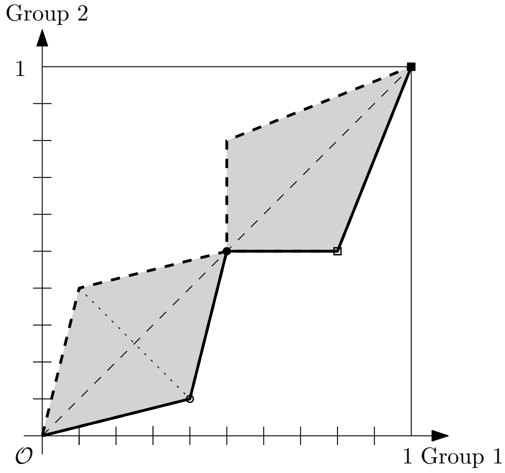
Each column vector associated to a class of $\mathbf{A}$ gives the coordinates of a point on the graph in figure 1(a). These points are represented by different symbols connected to the origin by a segment. For instance, the hollow circle represents $\mathbf{a}_1$. The bold lines in both panels of the figure are Monotone Paths obtained by arranging in different orders these segments. For instance, the upper and lower boundaries of the grey set in panel (a) result from arranging the classes of $\mathbf{A}$ by, respectively, decreasing or increasing values of the *concentration ratio* a_{1j}/a_{2j}. The unique Monotone Path in panel (b), instead, takes the order of the classes of $\mathbf{A}$ as given.

When the classes of $\mathbf{A}$ are permutable, there are $n_A!$ alternative Monotone Paths generated by these permutations. All of them are included in the *Zonotope* of $\mathbf{A}$, that is represented by the grey set in panel (a) of figure 1. The Zonotope $Z(\mathbf{A}) \subseteq [0,1]^d$ can be formally defined as:

$$Z(\mathbf{A}) := \left\{ \mathbf{z} := (z_1, \ldots, z_d)^t : \mathbf{z} = \sum_{j=1}^{n_A} \theta_j \mathbf{a}_j,\ \theta_j \in [0,1]\ \forall j = 1, \ldots, n_A \right\}.$$



(a) Zonotope and the segregation curve



(b) Path Polytope and the discrimination curve

Figure 1: Dissimilarity empirical criteria for $\mathbf{A}$ in (1)

It coincides with the set of all possible combinations of the vectors, or proportions of the vectors, corresponding to $\mathbf{A}$'s classes. Its graph is a convex polytope symmetric with respect to the point $\frac{1}{d}\mathbf{e}_d$. The (maximum) *Dissimilarity Zonotope* $Z(\mathbf{D})$ is the d-dimensional hypercube connecting the origin of $\mathbb{R}^d_+$ and the vertex with coordinates $\mathbf{e}_d$. It's diagonal is the *Similarity Zonotope* $Z(\mathbf{S})$. All distribution matrices that display some degree of dissimilarity originate Zonotopes that lie in $Z(\mathbf{D})$ and share the same reference diagonal $Z(\mathbf{S})$. As we will show, the Zonotope inclusion order will be the appropriate criterion for assessing dissimilarity when classes are permutable. Thus, the relation $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ will always indicate that the distributions in $\mathbf{B}$ are closer to similarity than those in $\mathbf{A}$.

In the two groups case, the *segregation curve* (Duncan and Duncan 1955, Hutchens 1991) coincides with the lower boundary of the Zonotope. Therefore, the Zonotopes inclusion criterion is a natural extension to the multi-group setting of the condition of non-intersecting segregation curves.[11]

If, instead, the order of the classes of $\mathbf{A}$ is given, the relevant dissimilarity empirical criterion should be based on the *Path Polytope*. In this case, only a Monotone Path $MP^*(\mathbf{A})$ can be associated to matrix $\mathbf{A} \in \mathcal{M}_d$. The Path Polytope $Z^*(\mathbf{A}) \in [0,1]^d$ is the convex hull of the permutations of $MP^*(\mathbf{A})$ with

[11]The segregation curve displays, for each proportion of group 1 occupying the first j ordered classes, the corresponding proportion of group 2 occupying the same classes, where classes j are arranged by increasing values of the concentration ratio $\frac{a_{2j}}{a_{1j}}$.

respect to the diagonal $Z(\mathbf{S})$ and is formalized as:

$$Z^*(\mathbf{A}) := \left\{\mathbf{z}^* := (z_1^*, \ldots, z_d^*)^t \ : \ \mathbf{z}^* \in \text{conv}\left\{\mathbf{\Pi}_d \cdot \mathbf{p} | \ \mathbf{\Pi}_d \in \mathcal{P}_d\right\}, \ \mathbf{p} \in MP^*(\mathbf{A})\right\},$$

where the conv operator denotes the convex hull of all elements in a set. For instance, the Path Polytope of $\mathbf{A}$ in (1) is the grey set represented in figure 1(b), obtained by taking the convex hull between every point on the Monotone Path and the corresponding symmetric point with respect to the diagonal $Z(\mathbf{S})$ (see for instance the points on the thin dashed segments). As the figure shows, the Path Polytope is not necessarily a convex set. The Dissimilarity Path Polytope and the Similarity Path Polytope coincide with $Z(\mathbf{D})$ and $Z(\mathbf{S})$, respectively. We will show that when the order of the classes is given, $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ will always indicate that the distributions in $\mathbf{B}$ are closer to similarity than those in $\mathbf{A}$.

Within the two groups context, the graph of a Monotone Path defines a *concentration curve* (Mahalanobis 1960, Butler and McDonald 1987) which might, or might not intersect the reference diagonal. When it doesn't, the Monotone Path is the boundary of the Path Polytope and its graph coincides with the *discrimination curve* (Le Breton et al. 2012).[12] The Path Polytope inclusion criterion is therefore a natural extension to the multi-group setting of non-intersecting concentration and discrimination curves.

3 Dissimilarity orders with permutable classes

3.1 Axioms

We introduce a set of axioms defining the change in dissimilarity that should be registered by every dissimilarity ordering when data are transformed according to some specific operations. The first axiom defines an anonymity property with respect to the labels of the classes.

Axiom *IPC (Independence from Permutations of Classes)* *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B = n$, *if* $\mathbf{B} = \mathbf{A} \cdot \mathbf{\Pi}_n$ *for a permutation matrix* $\mathbf{\Pi}_n \in \mathcal{P}_n$ *then* $\mathbf{B} \sim \mathbf{A}$.

[12]The concentration curve maps, for every share of groups 1 occupying the first j ordered classes, the corresponding share of group 2 occupying the same classes, whose order is given. The concentration curve always lies below the diagonal if the proportion of groups 1 in the first j ordered classes is always larger than the corresponding proportion of group 2, for every j.

Axiom *IPC* restricts the focus to evaluations where classes cannot be meaningfully ordered. Therefore, cumulating frequencies across classes does not provide additional information relevant for dissimilarity comparisons. To have a flavor of the implications of the axiom, consider the typical problem of measuring schooling segregation as the dissimilarity in the distributions of students, divided into ethnic groups, across schools. Within this framework, the *IPC* axiom posits that the name of the schools should not matter in the segregation assessment.

Dissimilarity comparisons might also be independent from the label assigned to the groups, while the focus still remains on their distributions. This is the case according to the *IPG* axiom.

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

The following axioms are related to the degree of informativeness of the distribution matrices. The Independence from Empty Classes (*IEC*) axiom states that the *insertion or the elimination of empty classes*, i.e. classes that are not occupied by groups, is irrelevant for dissimilarity comparisons. In the previous example, *IEC* would suggest that adding or eliminating schools in the same district with no students does not affect the overall segregation in the district. The axiom emphasizes dissimilarity originated in the pool of schools effectively interested by the phenomenon.

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 Independence from the Split of Classes (*SC*) axiom postulates that splitting proportionally (the groups populations in) a class into two new classes preserves the overall dissimilarity. As a result, the operation of *splitting of classes* allows to disaggregate a school into two or more smaller institutes without modifying the overall ethnic segregation of students, provided that the ethnic composition of the new institutes reflects the overall ethnic composition in the original school.[13]

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

[13] The transformation corresponds to a particular sequence of *linear bifurcations* for a probability distribution, introduced by Grant et al. (1998).

The operation of *merge of classes* complements the split operation. The Dissimilarity Decreasing Merge of Classes axiom (MC) states that dissimilarity never increases when two classes of a distribution matrix are mixed together. This occurs, for instance, when all the students in a school of departure migrate to another school of destination, so that each ethnic group population in the school of destination is increased by an amount equal to the size of the corresponding group in the school of departure.

Axiom *MC (Dissimilarity Decreasing Merge of Classes)* *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B$*, if* $\mathbf{b}_i = \mathbf{0}_d$*,* $\mathbf{b}_{i+1} = \mathbf{a}_i + \mathbf{a}_{i+1}$ *while* $\mathbf{b}_j = \mathbf{a}_j \; \forall j \neq i, i+1$*, then* $\mathbf{B} \preccurlyeq \mathbf{A}$.

The rationale for the MC axiom is due to the loss of informativeness deriving from the merge operation. In fact, when two schools are merged together, the differences in the ethnic composition of these schools either smooth out, or are preserved by the operation. As a consequence, a merge of schools reduces the informative power that the identity of a school has in predicting the group of origin of any randomly selected student in that school.

3.2 Characterization

When classes are permutable, the notion of dissimilarity is formalized by axioms *IPC*, *IEC*, *SC* and *MC*. Dissimilarity comparisons of distribution matrices, then, require to verify whether one configuration can be obtained from the other through operations underlying these axioms. When this is the case, every dissimilarity indicator consistent with these axioms should not register higher dissimilarity. This is true, in particular, for an interesting family of indicators, measuring dissimilarity as the average *dispersion* of groups compositions within classes.

For this family of indicators, the dispersion within each class is quantified by a function h in the class $\mathcal{H}$ of real valued convex functions defined on Δ^d. The evaluation of the dispersion in groups frequencies within each class contributes to the overall dissimilarity proportionally to the size of the population in that class. Denote by $\overline{a}_j := \mathbf{e}_d^t \cdot \mathbf{a}_j$ the overall population weight of each class, the dissimilarity index D_h with $h \in \mathcal{H}$ is:[14]

$$D_h(\mathbf{A}) := \frac{1}{d} \sum_{j=1}^{n_A} \overline{a}_j \cdot h\left(a_{1j}/\overline{a}_j, \ldots, a_{dj}/\overline{a}_j\right). \tag{2}$$

[14] For notational convenience empty classes receive weight $\overline{a} = 0$ and therefore do not contribute to the overall dissimilarity.

Here, $a_{ij}/\overline{a}_j$ is the share of the population in class j belonging to group i. Dissimilarity is minimal when this ratio is equal to $1/d$ for each of the d groups in all classes. Hence, by setting $h\left(\frac{1}{d}\mathbf{e}_d\right) = 0$ the index can be normalized to 0 when perfect similarity is reached.

A wide range of segregation indicators belongs to this family. The Dissimilarity Index introduced by Duncan and Duncan (1955) for $d = 2$ is one of them.[15] Also the Atkinson-type segregation index and the mutual information index characterized by Frankel and Volij (2011) for $d \geq 2$ are special cases of D_h.[16]

In the multi-groups context, a robust evaluation of the changes in dissimilarity requires to check that the condition $D_h(\mathbf{B}) \leq D_h(\mathbf{A})$ holds for all $h \in \mathcal{H}$.

An alternative interpretation of dissimilarity can be based on the well-known *informativeness* criterion by Blackwell (1953). Within this perspective, the less dissimilar distributions are those where classes are less informative on the group belongings of any randomly selected occupant. When perfect similarity ($\mathbf{S}$) is reached, the knowledge of the class is not informative on the groups occupying it, because each group distribution coincides with the overall population distribution. When maximal dissimilarity ($\mathbf{D}$) is reached, instead, a class identifies exclusively one group.

Operationally, the informativeness criterion requires to check that $\mathbf{B}$ is *matrix majorized* by $\mathbf{A}$, denoted $\mathbf{B} \preccurlyeq^R \mathbf{A}$. This means that there exists a row stochastic matrix $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$ such that $\mathbf{B} = \mathbf{A}{\cdot}\mathbf{X}$. To understand this relation, it is sufficient to recall that $\mathbf{B} \preccurlyeq^R \mathbf{A}$ implies that the rows of $\mathbf{B}$ are closer to independence, and therefore to similarity, than the rows of $\mathbf{A}$.[17]

In the spirit of the classical theorem in inequality analysis by Hardy et al.

[15] The Dissimilarity Index, only valid for $\mathbf{A} \in \mathcal{M}_2$, is $D(\mathbf{A}) := \frac{1}{2}\sum_{j=1}^{n_A} |a_{1j} - a_{2j}|$. It measures dissimilarity as the average absolute distance between the elements $a_{1j}/\overline{a}_j$ and $a_{2j}/\overline{a}_j$ in each class, obtained by setting $h(a_{1j}/\overline{a}_j, a_{2j}/\overline{a}_j) := |a_{1j}/\overline{a}_j - a_{2j}/\overline{a}_j|$.

[16] Within our framework, the Atkinson type index of segregation evaluated over non empty classes is $A_\omega(\mathbf{A}) := 1 - \sum_{j=1}^{n_A}\prod_{i=1}^{d}(a_{ij})^{\omega_i}$ for $\omega_i \geq 0$ s.t. $\sum_{i=1}^{d}\omega_i = 1$. It can be obtained by setting $h(a_{1j}/\overline{a}_j, \ldots, a_{dj}/\overline{a}_j) := 1 - d\prod_{i=1}^{d}(a_{ij}/\overline{a}_j)^{\omega_i}$. The mutual information index evaluated over non empty classes is $M(\mathbf{A}) := \log_2(d) - \sum_{j=1}^{n_A}\left(\frac{\overline{a}_j}{d}\right)\sum_{i=1}^{d}\frac{a_{ij}}{\overline{a}_j}\cdot\log_2\left(\frac{\overline{a}_j}{a_{ij}}\right)$ with $\frac{a_{ij}}{\overline{a}_j}\cdot\log_2\left(\frac{\overline{a}_j}{a_{ij}}\right)$ set equal to 0 if $a_{ij} = 0$. It can be obtained by setting $h(a_{1j}/\overline{a}_j, \ldots, a_{dj}/\overline{a}_j) := \sum_{i=1}^{d}\frac{1}{d}\cdot\log_2(d) - \frac{a_{ij}}{\overline{a}_j}\cdot\log_2\left(\frac{\overline{a}_j}{a_{ij}}\right)$.

[17] Interesting applications of matrix majorization can be found in linear algebra and majorization theory (Dahl 1999, Hasani and Radjabalipour 2007), in inequality analysis (see Chapter 14 in Marshall et al. 2011), in the comparison of statistical experiments (Blackwell 1953, Torgersen 1992), in information theory (Grant et al. 1998) and the study of bivariate dependence orderings for unordered categorical variables (Giovagnoli, Marzialetti and Wynn 2009), as well as in segregation analysis (Frankel and Volij 2011).

(1934), the main result of this section consists in a formalization of the equivalence between data transformations, agreement among dissimilarity orderings and matrix majorization. Comparisons based on these criteria, however, can hardly be verified empirically. The following theorem also shows that there exists an equivalent, but empirically implementable, criterion to assess dissimilarity comparisons which is based on Zonotopes inclusion.

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

(i) $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *through a finite sequence of insertion of empty classes, permutations, splits and merges of classes.*

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

(iii) $D_h(\mathbf{B}) \ \leq\ D_h(\mathbf{A})$ *for all* $h \in \mathcal{H}$*.*

(iv) $\mathbf{B} \ \preccurlyeq^R\ \mathbf{A}$*.*

(v) $Z(\mathbf{B}) \ \subseteq\ Z(\mathbf{A})$*.*

The equivalence between statements *(i)*, *(ii)*, *(iii)* and *(v)* is novel. The link between *(i)* and *(iv)* has been discussed in Grant et al. (1998) and in Frankel and Volij (2011). The equivalence between conditions *(iv)* and *(v)* when $d = 2$ has been shown in Dahl (1999). Theorem 1 extends this equivalence, which plays a key role in applied analysis, to the multi-group setting.

The condition of Zonotopes inclusion in *(v)* offers a novel geometric interpretation of dissimilarity. For a given proportion $p \in (0, 1)$ of the overall population, define an *isopopulation line* (when $d = 2$) or *(hyper)plane* (when $d \geq 3$) as the set of all groups combinations attainable that add up to p. Consider all points $\mathbf{z} \in Z(\mathbf{A})$ such that $\frac{1}{d}\mathbf{e}_d^t \cdot \mathbf{z} = p$. These points correspond to different Monotone Paths truncated at p that can be obtained by combining classes (or proportions of them) in any order. The dissimilarity exhibited by the proportion p of the population is associated to the dispersion of all possible groups compositions that can be reached by gathering together (shares of) different classes of $\mathbf{A}$ up to p. Matrix $\mathbf{B}$ is at most as dissimilar as $\mathbf{A}$ if and only if the set of all possible groups compositions adding up to the proportion p of the population in $\mathbf{B}$ is included in the corresponding set obtained from $\mathbf{A}$. This condition should hold for all proportions p.

We illustrate this interpretation for the two groups case. In figure 2 we depict the Zonotope of $\mathbf{A}$ in (1) along with the Zonotope of $\mathbf{A}'$, obtained from $\mathbf{A}$ by

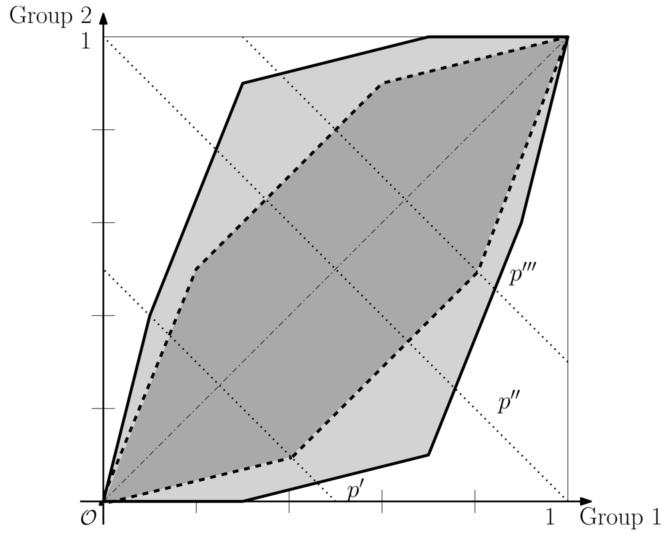


Figure 2: Zonotopes of matrices $\mathbf{A}$ (light grey area) and $\mathbf{A}'$ (dark grey area).

merging classes two and three, thus giving:

$$\mathbf{A}' = \begin{pmatrix} 0.4 & 0 & 0.4 & 0.2 \\ 0.1 & 0 & 0.4 & 0.5 \end{pmatrix} \text{ with } \overrightarrow{\mathbf{A}}' = \begin{pmatrix} 0.4 & 0.4 & 0.8 & 1 \\ 0.1 & 0.1 & 0.5 & 1 \end{pmatrix}. \tag{3}$$

In line with the equivalence between conditions *(i)* and *(v)* in Theorem 1 we have $Z(\mathbf{A}') \subseteq Z(\mathbf{A})$. The three dashed segments correspond to the isopopulation lines at proportions p', p'' and p''' of the overall population. The combinations of groups frequencies delimited by the Zonotopes boundaries on each isopopulation line are all those that can be obtained by splitting, merging and permuting the classes of $\mathbf{A}$ or $\mathbf{A}'$. At any isopopulation line, the set of combinations attainable with $\mathbf{A}$ includes the set of those attainable with $\mathbf{A}'$, thereby indicating that the latter configuration displays less dissimilarity.

3.3 Remarks

According to statement *(iv)* of Theorem 1, the *indifference class* of the dissimilarity partial order is fully characterized by the fact that $\mathbf{B} \sim \mathbf{A}$ for all admissible dissimilarity orderings if and only if there exist $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$ and $\mathbf{X}' \in \mathcal{R}_{n_B,n_A}$ such that $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ and $\mathbf{A} = \mathbf{B} \cdot \mathbf{X}'$. As a consequence $Z(\mathbf{B}) = Z(\mathbf{A})$.

The dissimilarity order is also preserved when some of the d groups in matrices $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ are mixed together with fixed weights, thus generating a new set of $d' < d$ groups distributions. We use row stochastic matrices with at most a non-zero element in each column, denoted by the subset $\widehat{\mathcal{R}}_{d',d} \subset \mathcal{R}_{d',d}$, to represent such mixtures.[18] Next remark is a consequence of the definition of matrix majorization.

Remark 1 *Let* $\widehat{\mathbf{X}} \in \widehat{\mathcal{R}}_{d',d}$ *with* $d' < d$*, if* $\mathbf{B} \preccurlyeq^R \mathbf{A}$*, then* $\widehat{\mathbf{X}} \cdot \mathbf{B} \preccurlyeq^R \widehat{\mathbf{X}} \cdot \mathbf{A}$.

The converse implication, however, is not true. In fact, it is possible to construct counterexamples that show this. Thus, $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ is only sufficient but not necessary for inclusion of the projections of the Zonotopes originated by considering $\widehat{\mathbf{X}} \cdot \mathbf{B}$ and $\widehat{\mathbf{X}} \cdot \mathbf{A}$. When $d' = 2$, it follows that testing whether $\mathbf{B}$ is less dissimilar than $\mathbf{A}$ for any comparison involving different pairs of groups (see Flückiger and Silber 1999) is not sufficient to guarantee that $\mathbf{B}$ can be obtained from $\mathbf{A}$ through a sequence of dissimilarity preserving and reducing operations.

Finally, the axiom *IPG* formally sets the independence from the name of the groups as a property of a dissimilarity order, thus enlarging the indifference set by including all comparisons obtained by relabeling the groups. Next remak formalizes the implications of adding *IPG* to the axioms considered in Theorem 1.

Remark 2 *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$*,* $\mathbf{B} \preccurlyeq \mathbf{A}$ *for all* $\preccurlyeq$ *satisfying axioms IPC, IEC, SC, MC and IPG if and only if* $\exists \mathbf{\Pi} \in \mathcal{P}_d$ *such that* $\mathbf{B} \preccurlyeq^R \mathbf{\Pi} \cdot \mathbf{A}$.

Applications of the results in Theorem 1 to inequality and segregation analysis are discussed in-depth in Section 5.

4 Dissimilarity orders with ordered classes

When classes are associated to ordered realizations such as educational achievements, health levels, income levels or alternatively schools, occupations or neighborhoods classified according to their quality standards, the *cumulative* groups distributions carry the relevant information for making dissimilarity comparisons. In these cases, the axiom *IPC* should be dropped from the analysis. As illustrated in the next example, also axiom *MC* is not appropriate in this context.

From now on, we assume that the classes of the distribution matrices are arranged from the worst to the best realization they represent. Consider, for instance, the cumulative distribution matrix $\overrightarrow{\mathbf{A}}$ in (1). The data show that group 1 suffers a disadvantage compared to group 2, given that at any class the proportion of group 2 members experiencing worse realizations is never larger than the

[18] For instance, the matrices in $\widehat{\mathcal{R}}_{2,3}$ generate comparisons involving two groups. These comparisons involve either all three pairs of groups, or each single group against a mixture of the remaining two groups.

corresponding proportion of group 1 members. A notable exception is class two, where the two groups are equally represented in terms of cumulative frequencies.

The operation of merging classes two and three of $\mathbf{A}$ might lead to counterintuitive conclusions on the change in dissimilarity when classes are ordered. This operation moves 10% of the population of the disadvantaged group 1, along with 40% of the population of group 2, from class two to three, thus giving the new configuration $\mathbf{A}'$ as in (3). Though the operation improves the situation of both groups, it reinforces the advantage of group 2 compared to group 1. In fact, the cumulative distributions in $\overrightarrow{\mathbf{A}}'$ move apart and the gap between the two groups does not "compensate" anymore in class two. This operation can be hardly seen as reducing dissimilarity, although Theorem 1, if adapted for the ordered case, would suggest otherwise. In fact, the inclusion of $Z(\mathbf{A}')$ into $Z(\mathbf{A})$ can be verified in figure 2. Zonotopes comparisons based on merge transformations are therefore not appropriate when classes are ordered. Instead, the discrimination curve of $\mathbf{A}'$, if represented in figure 1(b), would lie nowhere above the discrimination curve of $\mathbf{A}$, already pictured. This comparison highlights an *increase* in dissimilarity when the focus is shifted to discrimination curves analysis.

When classes are ordered, we replace axiom MC with an alternative dissimilarity reducing principle based on the *exchange* of two groups populations across two classes. Every exchange operation makes sure that the improvement of the situation for a proportion of the disadvantaged group is counterbalanced by a deterioration of the situation for an equal proportion of the advantaged group.[19] This was not the case in the previous example where, due to the merge transformation, the improvement experienced by the disadvantaged group 1 was overbalanced by the improvement experienced by group 2. Note that the notion of exchange is meaningful only when classes are ordered.

4.1 Axioms for the ordered case

We define and study the exchange operations within a restricted domain of matrices where *all groups are ordered* according to (first order) *stochastic dominance.* We say that group h dominates group ℓ whenever $\overrightarrow{a}_{hk} \leq \overrightarrow{a}_{\ell k}$ for all classes $k = 1, ..., n$. That is, ℓ is over-represented at the bottom of the realizations domain compared to h.

Unless groups h and ℓ distributions coincide, if h dominates ℓ then there must

[19] For related concepts see Tchen (1980), Van de gaer et al. (2001) and literature cited therein.

exist a class k where $\overrightarrow{a}_{hk} < \overrightarrow{a}_{\ell k}$ such that $a_{\ell k} > 0$. When this is the case, dissimilarity can be reduced through an exchange operation, consisting in an upward movement of a *small enough amount* $\varepsilon > 0$ of the population of ℓ, over-represented at the bottom of the realizations domain, from class k to any other better class $k' > k$. This change is counterbalanced by a downward movement of an equal amount ε of group h from class k' to k. By "small enough" we mean that, after the transfer, the dominance relations between *all* groups (and notably also between h and ℓ) are preserved.

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

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

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

Note that condition *(i)* implies that both matrices have the same size, that is $n_A = n_B = n$, and that their overall population distributions coincide.

For empirical purposes, dissimilarity comparisons involving only ordinal comparable matrices are extremely limited. Further structure imposed on the dissimilarity orders will allow to extend the analysis to all matrices in $\mathcal{M}_d$.

The axiom *IPG*, introduced in Section 3, would shift the focus from groups labels to their distributions, thus extending comparability also to distribution matrices that violate condition *(iii)* in Definition 1. This allows to deal with cases where all groups in $\mathbf{A}$ and $\mathbf{B}$ are ordered according to the stochastic dominance criterion, but their labels do not coincide in the two matrices.

If the focus is on the departure from similarity and not on what group dominates the others, then after a permutation of the groups the informativeness of the classes in a distribution matrix is preserved. In this case, the relevant information on whether one group dominates another is conveyed by the comparison of their cumulated population shares within each class. Thus, permuting the labels of the groups does not affect the differences among these shares in each class. Following this reasoning, it is possible to move one step further and impose that when two or more cumulative distributions coincide in a class then permuting the name of the respective groups from that class on will preserve the informativeness of the matrix and lead to a new distribution matrix that exhibits the same degree of dissimilarity. This is formalized through the *Interchange of Groups (I)* axiom.

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

If all the groups distributions coincide in the initial, say k, classes of $\mathbf{A}$ then an interchange of the groups from class $k+1$ on reproduces the effect of a permutation of the label of the groups. Only matrices in $\mathcal{M}_d$ that satisfy condition *(i)* in Definition 1 can be compared making use of interchange transformations. When axioms E and I are combined together, it emerges that for these matrices dissimilarity is reduced when the gap between any two groups cumulative distributions shrinks in at least one class. Making use of interchange transformations, the groups distributions that are not ordered according to stochastic dominance can be suitably rearranged to restore condition *(ii)* in Definition 1 without affecting the overall dissimilarity.

If instead the groups in two distribution matrices are similarly ordered according to stochastic dominance, but the number and population size of the classes in these matrices do not coincide, then axioms *IEC* and *SC*, whose validity for the dissimilarity model has been already motivated, allow to reshape the overall population distribution such that condition *(i)* in Definition 1 is satisfied. Furthermore, the possibility of adding an empty class at the beginning of a distribution matrix makes always possible to reproduce the effects of a permutation of the groups by applying an interchange of the groups in the first class. This shows that I together with *IEC* imply *IPG*, which is therefore dropped from the analysis.

Our notion of dissimilarity for the ordered case is built on the axioms *E*, *I*, *IEC* and *SC*. An example shows the implications of these axioms. The groups in matrices $\mathbf{A}$ and $\mathbf{A}'$ in (1) and (3) are ranked according to stochastic dominance but the two matrices do not satisfy ordinal comparability, because class two in $\mathbf{A}'$ is empty and the size of class three is larger than the size of the corresponding class in $\mathbf{A}$. By eliminating class two in $\mathbf{A}'$ and splitting class three into two classes according to the proportions 0.625 and 0.375, one obtains

$$\mathbf{A}'^{*} = \begin{pmatrix} 0.4 & 0.25 & 0.15 & 0.2 \\ 0.1 & 0.25 & 0.15 & 0.5 \end{pmatrix} \quad \text{with} \quad \overrightarrow{\mathbf{A}'^{*}} = \begin{pmatrix} 0.4 & 0.65 & 0.8 & 1 \\ 0.1 & 0.35 & 0.5 & 1 \end{pmatrix}. \qquad (4)$$

It follows that $\mathbf{A}' \sim \mathbf{A}'^{*}$ holds for all dissimilarity orderings satisfying *IEC* and *SC*. Moreover, $\mathbf{A}$ and $\mathbf{A}'^{*}$ are ordinal comparable. An exchange of $\varepsilon = 0.15$ involving classes two and three can then transform $\mathbf{A}'^{*}$ into $\mathbf{A}$:

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

Thus, we conclude that $\mathbf{A} \preccurlyeq \mathbf{A}'^{*} \sim \mathbf{A}'$, confirming the intuition that the merge operation, transforming $\mathbf{A}$ into $\mathbf{A}'$, leads to an unambiguous *increase* in dissimilarity when classes are ordered.

4.2 Characterization

When classes are ordered, every dissimilarity indicator should register a reduction in dissimilarity if one configuration is obtained from another through a sequence of exchange transformations. If such sequence exists, then the two configurations are represented by ordinal comparable distribution matrices. Within this set of matrices, dissimilarity assessments can be related to an interesting class of indicators, constructed as a weighted average of the cumulative groups frequencies.

Consider the weighting scheme $\mathbf{w}$ that attaches a weight $w_{ij} \geq 0$ to every group $i = 1, \ldots, d$ and class $j = 1, \ldots, n$. Let $\mathcal{W}$ denote the set of weighting schemes where all weights sum to one and are non-decreasing with respect to i, so that $0 \leq w_{ij} \leq w_{i+1\,j}$ for any i and j. For a given $\mathbf{w} \in \mathcal{W}$, a dissimilarity indicator can be written as a *linear rank-dependent* evaluation function:[20]

$$D_{\mathbf{w}}(\mathbf{A}) := \sum_{j=1}^{n} \sum_{i=1}^{d} w_{ij}\, \overrightarrow{a}_{(i)j}, \qquad (6)$$

where $\overrightarrow{a}_{(i)j}$ denotes the i-th smaller element of vector $\overrightarrow{\mathbf{a}}_j$, with $\overrightarrow{a}_{(1)j} \leq \ldots \leq \overrightarrow{a}_{(d)j}$. Since the weights w_{ij} are non-decreasing in i, any dispersion among the elements of $\overrightarrow{\mathbf{a}}_j$ produces, according to (6), larger dissimilarity than a situation where

[20]See Yaari (1987) and Weymark (1981). Indicators admitting a similar representation can be found in Ebert (1984), or in a recent survey by Yalonetzky (2012). By setting $w_{ij} = 1-(1-i/d)^2$ for all j, within class dispersion in (6) is evaluated by the Gini coefficient of $\overrightarrow{\mathbf{a}}_j$ entries.

cumulative frequencies coincide within class j. Overall dissimilarity is measured by linearly aggregating these evaluations across classes.

As we will show, the family of indicators $D_{\mathbf{w}}$ is consistent with axiom E. Robust dissimilarity comparisons will require therefore to check that $D_{\mathbf{w}}(\mathbf{B}) \leq D_{\mathbf{w}}(\mathbf{A})$ for all $\mathbf{w} \in \mathcal{W}$.

An alternative interpretation of the agreement among the dissimilarity orders described above can be formulated in terms of informativeness, intended as the possibility of discriminating between different groups distributions by the knowledge of the class of their realizations. When perfect similarity is achieved, the groups cumulative distributions coincide, while the presence of dissimilarity implies a certain degree of inequality among the cumulative groups frequencies across the classes. This criterion is operationalized by the fact that each column of $\overrightarrow{\mathbf{B}}$ is *majorized* by the corresponding column of $\overrightarrow{\mathbf{A}}$.

Majorization for $d-$dimensional row vectors $\mathbf{a}^t, \mathbf{b}^t \in \mathcal{M}_1$ with $\mathbf{a}^t \cdot \mathbf{e}_d = \mathbf{b}^t \cdot \mathbf{e}_d$ is denoted $\mathbf{b}^t \preccurlyeq^D \mathbf{a}^t$, meaning that there exists a *doubly* stochastic matrix $\mathbf{X} \in \mathcal{D}_d$ such that $\mathbf{b}^t = \mathbf{a}^t \cdot \mathbf{X}$ (see Marshall et al. 2011). The concept of majorization is extensively applied in univariate inequality analysis and is equivalent to the fact that $\mathbf{b}^t$ *Lorenz dominates* $\mathbf{a}^t$, i.e. $\sum_{j=1}^h b_j \geq \sum_{j=1}^h a_j \ \forall h = 1, \ldots, d$ once the elements of $\mathbf{a}^t$ and $\mathbf{b}^t$ are arranged by non-decreasing magnitude.[21]

Dissimilarity comparisons require to verify that the majorization condition holds for all pairwise comparisons of the cumulative groups distributions across all classes of $\mathbf{A}$ and $\mathbf{B}$. That is, lower informativeness is implied by the condition $\overrightarrow{\mathbf{b}}_k^t \preccurlyeq^D \overrightarrow{\mathbf{a}}_k^t$ for all $k = 1, \ldots, n$. This is a consequence of the fact that any exchange of groups frequencies across classes can be equivalently represented by *rank-preserving* transfers across groups, occurring in the space of cumulative frequencies distributions.

Constructing a parallel with Theorem 1, in the next theorem we show that the informativeness criterion applied in the ordered setting is indeed related to agreement for the class of indicators in (6), it is consistent with the majorization criterion and it is implemented empirically by the Path Polytopes inclusion criterion.

[21] See Marshall et al. (2011), Fields and Fei (1978) and Tchen (1980) for relevant applications. Here we use the symbol $\preccurlyeq^D$ to denote majorization via doubly stochastic matrix operations.

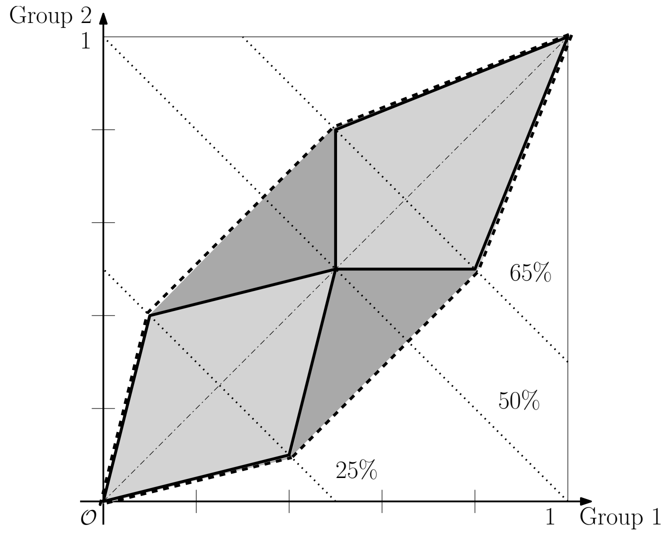


Figure 3: Path Polytopes of matrices $\mathbf{A}$ (light grey area) and $\mathbf{A}'$ and $\mathbf{A}'^*$ (dark grey area).

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

(i) $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *through a finite sequence of exchange operations.*

(ii) $\mathbf{B} \preccurlyeq \mathbf{A}$ *for every ordering* $\preccurlyeq$ *that satisfies axiom E.*

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

(iv) $\overrightarrow{\mathbf{b}}_k^t \preccurlyeq^D \overrightarrow{\mathbf{a}}_k^t$ *(i.e.,* $\overrightarrow{\mathbf{b}}_k^t$ *Lorenz dominates* $\overrightarrow{\mathbf{a}}_k^t$*) for all* $k = 1, \ldots, n$.

(v) $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$.

Claims *(i)-(v)* identify a set of novel equivalences.

The relations *(i)-(iv)* apply exclusively to distribution matrices that are ordinal comparable. This property allows to characterize the dissimilarity order through n Lorenz dominance comparisons of cumulative groups distributions with fixed population shares $\overrightarrow{\overline{b}}_k = \overrightarrow{\overline{a}}_k$ for every k.

The Path Polytope inclusion test in claim *(v)* can apply, instead, to any pair of distribution matrices in $\mathcal{M}_d$. In fact, the dissimilarity preserving transformations underlying axioms *IEC*, *SC* and *I* do not affect the Path Polytope shape and make possible to extend the validity of the equivalences in Theorem 2 to any pair of distribution matrices in $\mathcal{M}_d$. This is illustrated in figure 3, reporting the Path Polytopes of the distribution matrices $\mathbf{A}$ and $\mathbf{A}'^*$ in (1) and (4). The two matrices are ordinal comparable, and from the fact that $Z^*(\mathbf{A}) \subseteq Z^*(\mathbf{A}'^*)$, as shown in the figure, one immediately concludes that $\mathbf{A}'^*$ displays more dissimilarity than $\mathbf{A}$. However, the Path polytope of $\mathbf{A}'^*$ coincides with the one of $\mathbf{A}'$ in (3), which is not ordinal comparable to $\mathbf{A}$. If the operations transforming $\mathbf{A}'$ into $\mathbf{A}'^*$ give $\mathbf{A}' \sim \mathbf{A}'^*$, then it is possible also to compare $\mathbf{A}$ and $\mathbf{A}'$.

More generally, the elementary operations of split, insertion or elimination of classes and interchange of groups can transform any pair of distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ into corresponding distribution matrices $\mathbf{A}^*$, $\mathbf{B}^* \in \mathcal{M}_d$ that are ordinal comparable. We illustrate here the logic behind the various steps that lead to the desired result.

A first sequence of splits applied to $\mathbf{A}$ should be implemented whenever, for any pair of groups h and ℓ, there exists a class k such that $\overrightarrow{a}_{hk-1} > \overrightarrow{a}_{\ell k-1}$ and $\overrightarrow{a}_{hk} < \overrightarrow{a}_{\ell k}$. In this case class k should be split such that $\overrightarrow{a}_{hk} = \overrightarrow{a}_{\ell k}$ is obtained in the new configuration. An analogous sequence of splits should also be applied to $\mathbf{B}$. Then, if necessary, application of interchange operations and/or addition/elimination of empty classes can lead to two distribution matrices that satisfy conditions *(ii)* and *(iii)* in Definition 1. To conclude, further splits and/or addition/elimination of empty classes in both matrices allow to reshape the number and size of the classes, so that the overall population distributions end up to coincide between the classes of both matrices, thereby satisfying condition *(i)* in Definition 1. The outcome of these sequences of transformations will be therefore the desired ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$. Since these operations preserve dissimilarity, we can conclude that $\mathbf{A}^* \sim \mathbf{A}$ and $\mathbf{B}^* \sim \mathbf{B}$.

There are many alternative pairs of ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ that can be obtained from $\mathbf{A}$ and $\mathbf{B}$ respectively through the dissimilarity preserving transformations. Dissimilarity orderings consistent with E are defined on any of these pairs. Assuming that *IEC*, *SC* and *I* hold, the same orderings apply by transitivity to $\mathbf{A}$ and $\mathbf{B}$ in $\mathcal{M}_d$, thus extending the dissimilarity characterization proposed in Theorem 2 to the whole class $\mathcal{M}_d$ of distribution matrices.

Theorem 3 *For any* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_d$ *the following statements are equivalent:*

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

(ii) $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ *for every ordering* $\preccurlyeq$ *that satisfies axiom E and for every pair of ordinal comparable matrices* $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ *that are obtained from* $\mathbf{A}$ *and* $\mathbf{B}$ *through a finite sequence of insertions/eliminations of empty classes, splits of classes and interchanges of groups.*

Theorems 2 and 3 introduce two novelties. The first novelty is that the inclusion of the Path Polytopes of two ordinal comparable matrices is sufficient to guarantee that one distribution matrix has been obtained from the other only through exchanges. The second novelty is that ordinal comparability is not at all a constraining requirement. It can always be restored by conveniently modifying the overall population distribution through dissimilarity preserving operations. The axioms considered in Theorem 3 constitute a rigorous and comprehensive justification for using the Path Polytope inclusion $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ to indicate that the distributions in $\mathbf{B}$ are at most as dissimilar as the distributions in $\mathbf{A}$.

As motivated in the next section, the Path Polytope test breaks down into a sequence of Lorenz dominance evaluations of the dispersion of the cumulative groups distribution at fixed shares of the overall population, rather than at fixed levels of realizations. This is an appealing criterion for comparing situations where the cardinality of the classes is not relevant for the evaluation.

The Lorenz criterion induces, of course, an incomplete ranking. Completeness is customarily achieved by comparing specific dissimilarity indices as special cases of those in (6), or distance indices as those illustrated in Section 5.

4.3 Implementing and interpreting the dissimilarity test

There is an interesting parallel between Zonotopes and Path Polytopes inclusion criteria: both can be checked making use of isopopulation lines. In the ordered case, however, every isopopulation line represents a proportion $p \in [0, 1]$ of the worse-off population. The composition of groups shares making up the proportion p is therefore unique, up to a permutation of the groups labels. The Path Polytopes inclusion criterion can be interpreted as the Lorenz dominance (which is invariant to permutations) of groups proportions distributions at any isopopulation line of size p. In the two groups case, the size of this dispersion is given by the length of the intersection between Path Polytopes and isopopulation lines, as shown in figure 3 for selected population shares. The Lorenz dominance interpretation can also be extended to the multi-group setting.[22]

An illustrative example with three groups shows that the Path Polytope inclusion test has an alternative interpretation in terms of dispersion between *cumulative distribution curves* (cdfs), representing groups distributions. The graphs

[22] For instance, for $d = 3$ consider an isopopulation plane of size p passing through $\overrightarrow{\mathbf{b}^*}_k$ and $\overrightarrow{\mathbf{a}^*}_k$. As shown by Kolm (1969), $\overrightarrow{\mathbf{b}^*}_k$ Lorenz dominates $\overrightarrow{\mathbf{a}^*}_k$ if and only if it is included in the hexagon generated by all the permutation of $\overrightarrow{\mathbf{a}^*}_k$ (the *Kolm triangle*), a subset of the simplex with vertices $(p \cdot d, 0, 0)$, $(0, p \cdot d, 0)$ and $(0, 0, p \cdot d)$. An hexagon corresponds to the "contour curves" of the Path Polytope $Z^*(\mathbf{A})$ identified by the intersection with the isopopulation plane p. Checking inclusion for every $p \in [0, 1]$ leads to Path Polytope inclusion.

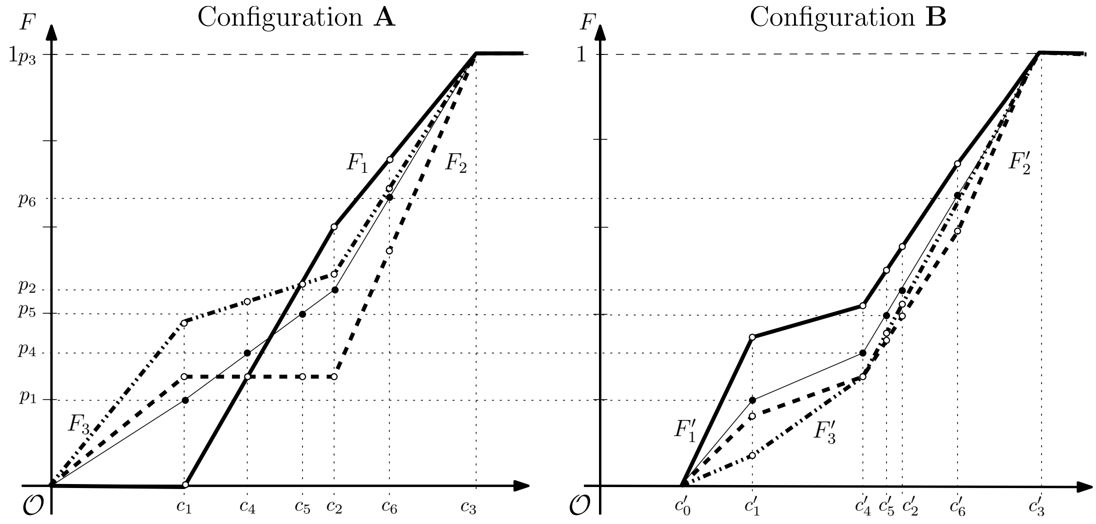


Figure 4: Cdfs of group 1 (solid lines), group 2 (dashed lines) and group 3 (dotted lines) for two configurations, with overall population distributions (thin solid line).

of the cdfs F_1, F_2, F_3 and F'_1, F'_2, F'_3, associated respectively to configurations $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_3$, are reported in figure 4. The graphs are continuous piecewise linear, a representation that can be supported by a sequence of infinite splits of each class.[23]

Configuration **A** is defined on three classes, while configuration **B** is defined on five classes of realizations of an underlying ordered outcome variable, giving the scale of the horizontal axis. The boundaries of these classes are arbitrarily set to coincide with c_1-c_3 for **A**, and with c'_0, c'_1, c'_4, c'_6 and c'_3 for **B**, where the class bounded by the origin and c'_0 is empty.

The dissimilarity comparison of the two configurations develops in two steps. The first step consists in calculating the distribution of the overall population. In the example, it corresponds to the vertical average between the three cdfs, and it is represented by a thin solid line in the figure.

The second step involves fixing proportions $p \in [0, 1]$ of the overall population (on the vertical axis), determining at every p the corresponding quantile of the overall population distribution (on the horizontal axis) in each of the two configurations, and finally checking the degree of dispersion of the groups cumulative distributions composition (white dots) around the common average (black dot),

[23]Through split operations, the original data can be represented in a graph by a finer step function. At the limit, this step function, if properly defined, can become piecewise linear, since any monotonic continuous function is the limit of a sequence of step functions (see ch. 1 in Asplund and Bungart 1966).

evaluated in correspondence of the respective quantiles in **A** and **B**. This corresponds to check that at any p the Lorenz curve of the groups proportions identified by the white dots in configuration **B** dominates the one in **A**.

As the figure shows, piecewise linearity of the cdfs only calls for a *finite* number of comparisons at population levels $p_1 - p_6$, that can be easily identified from the patterns of the cdfs graphs. This makes the criterion empirically tractable, and disconnected from the cardinality of the underlying realizations domain.

5 Related orders

5.1 Inequality

Consider two n-variate vectors $\mathbf{a}^t, \mathbf{b}^t \in \mathcal{M}_1$. These vectors may represent income shares distributions across n individuals. A well known result in inequality measurement is that an income distribution $\mathbf{b}^t$ displays less inequality than another distribution $\mathbf{a}^t$ if it can be obtained from the latter through a finite sequence of progressive (Pigou-Dalton, *PD*) transfers of income from rich to poor income recipients, without switching their relative positions in the income ranking (Hardy et al. 1934, Marshall et al. 2011). If this is the case, then $\mathbf{b}^t \preccurlyeq^D \mathbf{a}^t$ or, equivalently, the Lorenz curve of $\mathbf{b}^t$ lies nowhere below the Lorenz curve of $\mathbf{a}^t$.

Recall that the the Lorenz curve is a joint plot of the cumulative income shares, arranged by increasing income magnitude, and the cumulative empirical frequency at which these income shares are observed in the data. It follows that every inequality comparison involves the assessment of the dissimilarity between the distribution of income shares and the weights of the income recipients. The intuition extends also to the analysis of multivariate inequality, where income shares vectors are replaced by d-variate bundles of goods shares distributed across the n recipients. The following result, derived from Theorem 1, highlights the implications for inequality measurement.

Corollary 1 *Let* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *and let* $\preccurlyeq$ *be a dissimilarity ordering satisfying axioms IEC, IPC, SC and MC:*

$$\mathbf{B}' := \begin{pmatrix} \frac{1}{n_B}\mathbf{e}^t_{n_B} \\ \mathbf{B} \end{pmatrix} \quad \preccurlyeq \quad \mathbf{A}' := \begin{pmatrix} \frac{1}{n_A}\mathbf{e}^t_{n_A} \\ \mathbf{A} \end{pmatrix} \tag{7}$$

if and only if (i) $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ *with* $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$ *and (ii)* $\mathbf{e}^t_{n_A} \cdot \mathbf{X} = \frac{n_A}{n_B}\mathbf{e}^t_{n_B}$.

When $n_A = n_B = n$, matrix $\mathbf{X}$ in the corollary should be *doubly stochastic.* The majorization condition $\mathbf{B}' = \mathbf{A}' \cdot \mathbf{X}$ with $\mathbf{X} \in \mathcal{D}_n$ is commonly used for uni- and multivariate inequality analysis. In the univariate case ($d = 1$), $\mathbf{B}' \preccurlyeq \mathbf{A}'$ indicates that the *PD* transfers applied to $\mathbf{a}^t$ to obtain $\mathbf{b}^t$ involve precise sequences of splits and merges of the classes of $\mathbf{A}'$. Hence, splits and merges are regarded to as more elementary inequality reducing transformations than *PD* transformations.

The interesting result is that there always exists a sequence of splits and merges that support majorization even in the multidimensional case when $d \geq 2$ (denoted *uniform* majorization), while this is not the case for *PD* transfers (Kolm 1977).[24] Hence, any inequality comparison is a dissimilarity comparison, while the converse is not true.

Alternative criteria for assessing multivariate inequality are also embedded within the dissimilarity model. Koshevoy and Mosler (1996) have studied the properties of the *Lorenz Zonotopes* inclusion order. A Lorenz Zonotope $LZ(.)$ in $\mathbb{R}^{d+1}_+$ is the plot, for each population fraction, of the associated set of possible bundles of attributes that this fraction of population may achieve. When $d = 1$, LZ is consistent with the Lorenz curve order. By Corollary 1, $LZ(\mathbf{A}) := Z(\mathbf{A}')$ and the dominance relation $LZ(\mathbf{B}) \subseteq LZ(\mathbf{A})$ indicates, according to Theorem 1, that there exists a sequence of merge, split, permutation and insertion of empty classes transforming $\mathbf{B}'$ into $\mathbf{A}'$.

Ebert and Moyes (2003) analyze the relation between welfare evaluations, Lorenz dominance and equivalence scales for incomes when population weights may differ among units and across distribution matrices. In line with Corollary 1, inequality comparisons in this framework can be made in terms of the dissimilarity between the income distribution and the population weights distribution. A direct application of Theorem 1 shows that any welfare consistent measure of inequality should be based on convex transformations of equivalized incomes, scaled by their demographic weights. In fact, robust inequality comparisons in this framework coincide with the agreement of the indicators $D_h = \sum_{j=1}^{n} w_j h(a_j/w_j)$ with h convex, where w_j is individual j's weight and a_j/w_j is her equivalent income.[25]

An increasingly popular alternative notion of inequality, based on *inequality of*

[24] Marshall et al. (2011, Ch. 15.A) show that every *PD* transfer can be equivalently represented in matrix notation through *T-transforms*. Every sequence of T-transforms, either applied to vectors or matrices, generates uniform majorization, while the converse is true only in the univariate setting. Every T-transform can be decomposed into a precise sequence of splits and merges, however in the multivariate setting there can be sequences of splits and merges that support uniform majorization that cannot be represented through T-transforms.

[25] The result, based on Lemma 1 in the appendix, follows by the homogeneity and convexity of $g : \mathbb{R}^2 \to \mathbb{R}$, that gives $g(a_j, w_j) = w_j g(a_j/w_j, 1) = w_j h(a_j/w_j)$ with h convex.

opportunity analysis (Roemer 2012), suggests to compare the distribution of different socioeconomic groups with a reference distribution (possibly non-degenerate) incorporating fairness concerns. Inequality of opportunity evaluations may involve dissimilarity comparisons with ordered classes. The theorems in Section 4 suggest alternative models to measure it, that are particularly interesting when realizations convey only an ordinal meaning.

5.2 Segregation

In the two groups case, the lower bound of the Zonotope $Z(.)$ is the segregation curve associated to the two groups distributions. This is illustrated in figure 1. For $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$, if $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ then the segregation curve of $\mathbf{B}$ lies nowhere below the segregation curve of $\mathbf{A}$. Hence, the operations of merge, split, permutation and insertion of empty classes characterize the ranking produced by non-intersecting segregation curves (Hutchens 1991). This order is naturally extended to the multi-group case by looking at d-variate Zonotopes inclusion, which assures an empirical test for the informativeness perspective in segregation undertaken by Frankel and Volij (2011).

5.3 Discrimination

Within the ordered setting, consider instead $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$ in which row two stochastic dominates row one in both matrices. As shown in Figure 1, the lower bound of the Path Polytope $Z^*(.)$ is the discrimination curve. Hence, if $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ then the discrimination curve of $\mathbf{B}$ lies nowhere below the curve of $\mathbf{A}$. Le Breton et al. (2012) investigates the relation between this dominance and the traditional statistical discrimination analysis. The Path Polytopes inclusion in Theorem 2 extends results to the multi-group case and to situations where groups may not be ordered in terms of stochastic dominance.

5.4 Mobility

If groups identify percentiles of parental incomes, while classes correspond to percentiles of children incomes, then any distribution matrix might represent an intergenerational *mobility matrix*, where both groups and classes are ordered. If the mobility matrix is *monotone*[26], the dissimilarity model in Section 4 provides implementable criteria for assessing changes in economic mobility across percentiles in the distribution of destination, provided that perfect mobility is associated to lack of informativeness of the group of origin with respect to the class of destination (this is a recurring idea in many applied works, including Stiglitz 2012). For any pair of monotone mobility matrices with fixed marginals both for rows and columns, the dissimilarity criterion in Theorem 2 coincides with the *test of the orthants* (Tchen 1980, Dardanoni 1993).[27] If marginals differ, the dissimilarity order extends mobility comparisons, and can be tested through Path Polytopes inclusion. If the mobility matrices are non-monotone, the dissimilarity criterion imposes stronger conditions than the traditional mobility test. Empirical evidence suggests, however, that monotonicity of mobility matrices is unlikely to be rejected by the data (Dardanoni, Fiorini and Forcina 2012).

5.5 Distance

Dissimilarity analysis is also related to the measurement of the distance between two or more distributions (Ebert 1984). Recall that any distribution matrix in $\mathcal{M}_d$ can be equivalently represented by d cumulative distribution functions (cdfs) defined on a realizations domain $\mathcal{X}$. This can be done by assuming that population masses are uniformly distributed within classes, a consequence of admitting the possibility of splitting and adding empty classes without affecting the overall dissimilarity (see Section 4.3).

Denote these cdfs as $F_1,\ F_2,\ldots,F_d$, with overall population distribution $\overline{F} := \sum_i \frac{1}{d} F_i(x)$. The dissimilarity of $F_1,\ F_2,\ldots,F_d$ can be seen as the average vertical distance (according to some metrics) between these cdfs, as calculated at the realization $\overline{F}^{-1}(p)$ corresponding to the worse-off fraction p of the overall population. When $d = 2$, a natural metric for distance is $\left|F_1(\overline{F}^{-1}(p)) - F_2(\overline{F}^{-1}(p))\right|$. An aggregate distance indicator $D_2^*(F_1, F_2)$, consistent with the dissimilarity model for

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

[27] After groups proportions have been cumulated first by row and then by column, the test of the orthants requires to verify that the entries of the resulting matrix expressing higher mobility are nowhere smaller than the entries of the resulting matrix expressing lower mobility. Given that the matrices are monotone with fixed margins, they satisfy ordinal comparability and the test coincides with the Lorenz dominance criterion in statement *(iv)* of Theorem 2.

ordered classes, can be defined as the integral of these distances on $[0, 1]$, that is:

$$D_2^*(F_1, F_2) = \int_0^1 \left| F_1(\overline{F}^{-1}(p)) - F_2(\overline{F}^{-1}(p)) \right| dp. \quad (8)$$

After a change in variable, it rewrites:

$$D_2^*(F_1, F_2) := \int_{\mathcal{X}} |F_1(x) - F_2(x)| \, d\overline{F}(x). \quad (9)$$

Compared to the Manhattan distance index $D(F_1, F_2) := \int_{\mathcal{X}} |F_1(x) - F_2(x)| \, dx$, often employed in discrimination analysis, the dissimilarity indicator $D^*(F_1, F_2)$ is independent from any cardinal interpretation attributed to the classes. In fact, it follows from (9) that $D_2^*(F_1, F_2)$ is invariant to monotone transformations of the variable defined on the domain $\mathcal{X}$. Thus, $D^*(F_1, F_2)$ is suitable to measure discrimination not only with cardinal realizations such as income, but also with ordinal categories, such as health, education, well-being or life satisfaction outcomes.

The index reaches its maximum $D_2^*(F_1, F_2) = 1$ when F_1 and F_2 do not overlap, a situation where dissimilarity is maximal since any given realization can be attained exclusively by one of the groups. Furthermore, it can be shown that $D_2^*(F_1, F_2)$ is proportional to the area of the Path Polytope. For any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$, the condition $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ is therefore sufficient (but not necessary) for $D_2^*(F_1^B, F_2^B) \leq D_2^*(F_1^A, F_2^A)$.

Extensions of the distance index to $d \geq 3$ can be obtained from (9), assuming that distance is an average of the inequality within each class weighted by the overall population distribution across these classes. Let $I_d : [0, 1]^d \to [0, 1]$ be an onto function representing an inequality indicator consistent with the Lorenz criterion, such that $I_2(F_1, F_2) = |F_1 - F_2|$. The multi-group distance index is $D_d^*(F_1, \ldots, F_d) := \int_{\mathcal{X}} I_d(F_1(x), \ldots, F_d(x)) d\overline{F}(x)$. Compared to the family of indicators discussed in statement *(iii)* of Theorem 2, $D_d^*(F_1, \ldots, F_d)$ can be used to rank *any* pair of distribution matrices irrespectively from their number and size of classes, since it is invariant to splits and insertion of empty classes.

6 Concluding remarks

A large and sparse literature on segregation, discrimination, mobility, inequality and distance measurement has proposed criteria for ranking sets of ordered and non-ordered distributions according to the dissimilarity they exhibit. Two separate theorems show that a suitable set of transformations decreasing or preserving dissimilarity have a natural and meaningful interpretation in the analysis of the aforementioned phenomena, and prove that the existence of these transformations can be tested by looking at inclusions of either Zonotopes or Path Polytopes, that can be seen as multidimensional generalizations of the segregation, discrimination and concentration curves.

These results are of practical use in policy evaluation analysis. For instance, a policymaker interested in reducing ethnic segregation of students across institutes in a given school district might propose a portfolio of policy measures, none of which has to do with more "elementary" transformations such as splitting, merging, permuting or adding schools. Nonetheless, these "elementary" transformations might still be targeted as obviously segregation-preserving/reducing. If the implementation of the "complex" policy measures reshapes the students distribution across schools in a way that is consistent with sequences of more "elementary" transformations, then the policymaker can safely conclude that his objective has been achieved. The exact sequence of "elementary" transformations, however, cannot be easily inferred by looking at the data on the ethnic composition of the schools before and after the application of the policy measures. However, the policymaker can conclude that such sequence *exists* upon verification of the Zonotopes inclusion empirical test, based on the available data. Routines are made available to facilitate this task.

The same procedure applies to the analysis of dissimilarity with ordered classes, that can be related to problems of discrimination, mobility or equality of opportunity. In this case, the Path Polytope inclusion test is the relevant one for the policymaker. Routines are made available also in this case.

There might be cases, however, where Zonotopes or Path Polytopes inclusion fails. Making use of a dissimilarity indices lying in the classes studied in the two theorems, it is possible to produce an evaluation of the changes in dissimilarity that is consistent with the implications of the "elementary" transformations. Evaluations based on one or few dissimilarity indicators are, however, not robust since they can always be challenged on the perspective offered by alternative measures. The characterization of the dissimilarity indicators presented in the paper is therefore a policy-relevant priority that is left for future research.

A Proofs

We first introduce three lemmas and some additional notation to facilitate the demonstrations of the main theorems.

A.1 Preliminary results

Preliminary results for the proof of Theorem 1 The first result shows that matrix majorization admits an equivalent representation in a well defined class of convex functions.

Lemma 1 *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$*,* $\mathbf{B} \preccurlyeq^R \mathbf{A}$ *if and only if*

$$\sum_{j=1}^{n_B} g(b_{1j}, \ldots, b_{dj}) \;\leq\; \sum_{j=1}^{n_A} g(a_{1j}, \ldots, a_{dj}), \tag{10}$$

for all functions $g : \mathbb{R}^d \to \mathbb{R}$ *that are convex and homogeneous such that* $g(\mathbf{0}_d^t) = 0$*.*

For a formal proof, see Lemma 15.C.11 in Marshall et al. (2011).

The second result shows that the insertion of empty classes, split and merges operations can be represented through linear transformations involving row stochastic matrices. An operation of *insertion of empty classes* transforms $\mathbf{A}$ into $\mathbf{B}$ with $n_B > n_A$ by augmenting $\mathbf{A}$ of $n_B - n_A$ columns with zero entries. We denote $\mathcal{R}^{IEC}_{n_A,n_B} \subset \mathcal{R}_{n_A,n_B}$ the set of al matrices reproducing an insertion of empty classes. Hence $\mathbf{Y} \in \mathcal{R}^{IEC}_{n_A,n_B}$ is an identity matrix augmented by $n_B - n_A$ columns with zero entries.

Let $\mathcal{M}_d^0 \subset \mathcal{M}_d$ define the set of matrices exhibiting at least one column of zeroes. For $\mathbf{A} \in \mathcal{M}_d^0$, let $\mathcal{J}_A^0$ denote the index set of all columns in $\mathbf{A}$ with all zeroes and $\mathcal{J}_A$ denote the index set of all non-zero columns of $\mathbf{A}$. Let $j \in \mathcal{J}_A$ such that $j+1 \in \mathcal{J}_A^0$. The matrix $\mathbf{Z}_{[j]}$ incorporates an operation of *split of classes* applied to matrix $\mathbf{A} \in \mathcal{M}_d^0$ that leads to matrix $\mathbf{B} \in \mathcal{M}_d$ with $\mathbf{b}_j = \lambda \mathbf{a}_j$ and $\mathbf{b}_{j+1} = \mathbf{a}_{j+1} + (1-\lambda)\mathbf{a}_j = (1-\lambda)\mathbf{a}_j$. Let $k \neq k' \neq j$, the set of all transformation matrices $\mathbf{Z}_{[j]}$ reproducing a split of classes is denoted by:

$$\mathcal{R}_A^{SC} := \left\{ \mathbf{Z}_{[j]} \in \mathcal{R}_n : \begin{array}{c} z_{jj} := \lambda \ , \ z_{j\ j+1} := (1-\lambda), \ z_{kk} := 1, \ z_{kk'} := 0, \\ \lambda \in [0,1], j \in \mathcal{J}_A, j+1 \in \mathcal{J}_A^0 \end{array} \right\}.$$

Also the *merge of classes* operation originates $\mathbf{B} = \mathbf{A} \cdot \mathbf{M}_{[j]}$, where the matrix $\mathbf{M}_{[j]}$ performs a merge of class j towards $j+1$. It belongs to the set:

$$\mathcal{R}_n^{MC} := \left\{ \mathbf{M}_{[j]} \in \mathcal{R}_n \ : \ m_{j\,j+1} = m_{kk} = 1 \ \forall k \neq j, \ m_{ij} = 0 \text{ in all other cases} \right\}.$$

Preliminary results for the proof of Theorem 2 We develop a rank-preserving version of Tchen's (1980) algorithm to show that claim *(iv)* of Theorem 2 is always supported by the existence of a finite sequence of exchange transformations mapping the distribution matrix $\mathbf{A}$ into the less dissimilar one $\mathbf{B}$. The algorithm applies to ordinal comparable matrices, a subset of the matrices with fixed marginals analyzed in Tchen (1980). As a consequence, Tchen's algorithm is not valid in our setting because it may generate matrices where the rank of the groups is not preserved, as required in conditions *(ii)* and *(iii)* of Definition 1.

Additional notation We focus here on ordinal comparable matrices, where the order of the groups coincides with the one of the rows, so that group i dominates group $i-1$, for any i. Hence, for $\mathbf{A} \in \mathcal{M}_d$, $\overrightarrow{a}_{ij} \leq \overrightarrow{a}_{i-1\,j}$, $\forall i, j$. Moreover, let (x, y) identify the cell corresponding to row x and column y of a distribution matrix, with $x \in \{1, \ldots, d\}$ and $y \in \{1, \ldots, n\}$. The lexicographic order on $\{1, \ldots, d\} \times \{1, \ldots, n\}$ that we consider is denoted by $(x, y) < (x', y')$ if $y < y'$ or if $y = y'$ and $x > x'$. We also use $i \in [x, x']$ to denote $i \in \{x, \ldots, x' | x < \ldots < x'\}$. Furthermore, the *doubly cumulative distribution matrix* of $\mathbf{A}$ is denoted by $\overrightarrow{\overrightarrow{\mathbf{A}}}$, with $\overrightarrow{\overrightarrow{a}}_{ij} = \sum_{x \geq i} \overrightarrow{a}_{xj}$. Using this compact notation, the Lorenz dominance criterion in claim *(iv)* of Theorem 2 rewrites $\overrightarrow{\overrightarrow{\mathbf{B}}} \geq \overrightarrow{\overrightarrow{\mathbf{A}}}$.

Strategy of the proof The algorithm is built in two steps, illustrated respectively in Lemma 2 and 3. The *first step* delimits the building blocks of the analysis by developing a rank-preserving version of Tchen (1980) algorithm, from where the notation is taken. Given two ordinal comparable matrices $\mathbf{H}$, $\mathbf{H}' \in \mathcal{M}_d$ satisfying $h_{ij} < h'_{ij}$, $\overrightarrow{\overrightarrow{\mathbf{H}}} \leq \overrightarrow{\overrightarrow{\mathbf{H}'}}$ and $\overrightarrow{\overrightarrow{h}}_{xy} = \overrightarrow{\overrightarrow{h'}}_{xy}$ for all $(x, y) < (i, j)$, 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 $h_{ij} < k^m_{ij} \leq h'_{ij}$. The 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. Throughout the sequence reducing $h'_{ij} - h_{ij}$, the rank of the groups is always preserved.

Given two distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ ordinal comparable such that $\overrightarrow{\overrightarrow{\mathbf{A}}} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$, the *second step* develops the sequences of transfers of groups masses transforming $\mathbf{A}$ into $\mathbf{B}$ in a way that preserves, at each step, 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.[28] 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 transfer sequentially mass to groups $i-1, i-2, ...$ before affecting group i in class j. This subsequence is defined by $p \in \{1, \dots, P\}$. The two sequences together induce transfers that are bounded and guarantee:

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

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. 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 For any pair $\mathbf{H}, \mathbf{H}' \in \mathcal{M}_d$ ordinal comparable, consider the sequence $\mathbf{K}^m$ 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 maintain 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 $\overrightarrow{k}'_{ij} \geq \overrightarrow{k}'_{i+1j}, \forall i, j$).

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

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

$k_{i^*j} > 0$, and consider $\mathbf{K}$ satisfying the following conditions:

$$k_{ij} < h'_{ij} \text{ and } \overset{\rightrightarrows}{k}_{xy} = \overset{\rightrightarrows}{h'}_{xy} \text{ for all } (x,y) < (i,j), \tag{11}$$

$$\delta = \min\left\{ \overrightarrow{k}_{i-1\,j} - \overrightarrow{k}_{ij},\ \overrightarrow{k}_{i^*j} - \overrightarrow{k}_{i^*+1\,j},\ \frac{1}{2}(\overrightarrow{k}_{i^*j} - \overrightarrow{k}_{ij}) \right\} > 0. \tag{12}$$

Condition (11) is as in Tchen (1980), while condition (12) is new. It secures that there is enough mass that can be moved from cell (i^*, j) and added to (i, j) so that the rank of the groups is preserved. Given $\mathbf{K}$, define the sequence $S(\mathbf{K}, \mathbf{H}'|i^*) := (x_s, y_s)_{s\in\{1,\ldots,S\}}$ by setting

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

if $s < S$, while $(x_S, y_S) = (i^*, j)$. This sequence is nonempty with $x_S = i^* < x_{S-1} < \ldots < x_1 = i$ and $y_1 > y_2 > \ldots > y_S = j$, and leads to $\mathbf{K}'$. Define $\mathbf{K}^1 = \mathbf{K}$ and $\mathbf{K}^s$ as the distribution matrix obtained from $\mathbf{K}^{s-1}$ where at most a mass $\Delta > 0$ is subtracted from (i, y_{s-1}) and (x_s, y_s) and added to (x_s, y_{s-1}) and (i, y_s). The mass Δ that can be moved should coincide with the smallest quantity between (i) $h'_{ij} - k_{ij}$ (the quantity that should be compensated), (ii) the frequency of group x_s in class y_s (this guarantees the monotonicity), (iii) the gap between the cumulative distributions of group i and group $i-1$, and (iv) the gap between group x_s and group x_s+1. These two latter conditions guarantee that the rank of the groups is preserved by the transfer. When $x_s = i-1$, at most half of the gap $\overrightarrow{k}_{x_s j} - \overrightarrow{k}_{ij}$ can be transferred. By construction of the sequence, at every step s $k_{x_s y} = 0\ \forall x_s$, $\forall y \in [y_s, y_{s-1}-1]$. Thus, conditions (iii) and (iv) are always satisfied when (12) holds. Altogether these conditions give:

$$\Delta := \min\left\{ h'_{ij} - k_{ij},\ \min_{S(\mathbf{K},\mathbf{H}'|i^*)}\{k^s_{x_s,y_s}\},\ \delta \right\}. \tag{13}$$

Lemma 2 *Let* $\mathbf{K}$ *satisfy conditions (11) and (12), there exists* $\mathbf{K}' \in \mathcal{M}_d$ *obtained from* $\mathbf{K}$ *through a sequence of exchanges, such that* $\overset{\rightrightarrows}{\mathbf{K}'} \leq \overset{\rightrightarrows}{\mathbf{H}'}$ *and* $k'_{ij} = k_{ij} + \Delta$, *with* $\Delta > 0$ *as in (13).*

Proof Let $S(\mathbf{K}, \mathbf{H}'|i^*)$ defined as above and define $\mathbf{K}^1 = \mathbf{K}$. 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 (13) this quantity must be lower than k_{iy_1} and $k_{x_2y_2}$, which guarantees that $\overrightarrow{k}_{iy_1} - \Delta \geq 0$ and $\overrightarrow{k}_{x_2y_2} - \Delta \geq \overrightarrow{k}_{x_2+1\,y_2}$. This operation leads to $\mathbf{K}^2$. Then, a mass Δ is subtracted from (i, y_2) (x_3, y_3) and added to (i, y_3) and (x_3, y_2) giving $\mathbf{K}^3$. By (13), also this operation is supported by an exchange. The last step of this sequence involves moving mass from (i^*, j) to (i, j) where $k_{i^*j} > 0$ by definition. Recall that $i \geq 2$, hence i^* always exists. To show that $\overrightarrow{\overrightarrow{\mathbf{K}}}^s \leq \overrightarrow{\overrightarrow{\mathbf{H}}}'$ for any s, assume by recurrence that $\overrightarrow{\overrightarrow{\mathbf{K}}}^{s-1} \leq \overrightarrow{\overrightarrow{\mathbf{H}}}'$. For $(x, y) < (i, y_s)$, $k^s_{xy} = k^{s-1}_{xy}$. By definition, Δ is such that the order of groups i and $i-1$ is preserved, hence $k^s_{iy_s} > k^{s-1}_{iy_s}$ and $\overrightarrow{\overrightarrow{k}}{}^s_{iy_s} > \overrightarrow{\overrightarrow{k}}{}^{s-1}_{iy_s}$ while $\overrightarrow{\overrightarrow{k}}{}^s_{iy_s} < \overrightarrow{\overrightarrow{k}}{}^s_{i-1\,y_s}$. Moreover, $k^s_{xy} = k^{s-1}_{xy}$ for $x \in [x_s+1, i-1]$ and $y \in [y_s, y_{s-1}]$, hence $\overrightarrow{\overrightarrow{k}}{}^s_{xy} > \overrightarrow{\overrightarrow{k}}{}^{s+1}_{xy}$. Finally, $k^s_{x_sy_s} < k^{s-1}_{x_sy_s}$ and $\overrightarrow{\overrightarrow{k}}{}^s_{x_sy_s} = \overrightarrow{\overrightarrow{k}}{}^{s-1}_{x_sy_s}$, as well as $\overrightarrow{\overrightarrow{k}}{}^s_{xy_{s-1}} = \overrightarrow{\overrightarrow{k}}{}^{s-1}_{xy_{s-1}}$ for $x \in [x_s, i]$. Combining these conditions, the required result is obtained. *Q.E.D.*

Under (11) and (12), 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 $(x_s, y_s) \in S(\mathbf{K}, \mathbf{H}'|i^*)$. 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 $S(\mathbf{K}^1, \mathbf{H}'|i^*)$ leading to $\mathbf{K}^2$. A second iteration of the Lemma would give the sequence $S(\mathbf{K}^2, \mathbf{H}'|i^*)$ leading to $\mathbf{K}^3$, and so on. Generally, repeated iterations of the Lemma lead to a sequence of distribution matrices $\mathbf{K}^m$, $m \in \{1, \ldots, M\}$ where $h'_{ij} - k^{m+1}_{ij} < h'_{ij} - k^m_{ij}$. Each of these matrices is supported by a sequence $S(\mathbf{K}^m, \mathbf{H}'|i^*)$ so that if $\Delta = k^m_{x_sy_s}$ for some $(x_s, y_s) \in S(\mathbf{K}^m, \mathbf{H}'|i^*)$, then $S(\mathbf{K}^{m+1}, \mathbf{H}'|i^*)$ must contain all the points of $S(\mathbf{K}^m, \mathbf{H}'|i^*)$ 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 $S(\mathbf{K}^m, \mathbf{H}'|i^*)$ 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^*. Lemma 3, demonstrated in the second step of the main algorithm, shows how to iteratively construct matrices $\mathbf{H}$ and $\mathbf{H}'$ where either situation (i) or (ii) can occur.

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

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

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

The sequence then has reached cell (i,j), giving by construction $\overrightarrow{\overrightarrow{\mathbf{A}}}^{q,1} \leq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p-1} \leq \overrightarrow{\overrightarrow{\mathbf{A}}}^{q,p} \leq \overrightarrow{\overrightarrow{\mathbf{B}}}$. If $\overrightarrow{a}^{q,p}_{ij} = \overrightarrow{b}_{ij}$, the sequence is completed and $p = P$. Otherwise $\overrightarrow{a}^{q,1}_{i^*j} - (i-i^*)\Delta_{ij}(i^*) = \overrightarrow{a}^{q,p}_{i^*+1\,j} = \ldots = \overrightarrow{a}^{q,p}_{ij} < \overrightarrow{b}_{ij}$. In this case 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 exists at least a group 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 iteration, 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

$$\Delta_{ij}(i^*, \widetilde{i}) \;=\; \min\left\{ \overrightarrow{b}_{ij} - \overrightarrow{a}^{q}_{ij},\ \overrightarrow{a}^{q}_{\widetilde{i}j} - \overrightarrow{a}^{q}_{\widetilde{i}+1\,j},\ \frac{1}{i-\widetilde{i}}\left(\overrightarrow{a}^{q}_{i^*j} - \overrightarrow{a}^{q}_{i^*+1\,j}\right)\right\}. \tag{15}$$

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, one of the following constraints is binding:

$$\overrightarrow{a}^{q,p}_{\widetilde{i}j} = \overrightarrow{a}^{q,p}_{\widetilde{i}+1\,j} = \ldots = \overrightarrow{a}^{q,p}_{ij} < \overrightarrow{b}_{ij}, \tag{16}$$

$$\overrightarrow{a}^{q,p}_{i^*j} - (i - \widetilde{i})\Delta_{ij}(i^*, \widetilde{i}) = \overrightarrow{a}^{q,p}_{i^*+1\,j}. \tag{17}$$

If (16) holds but (17) does not, 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 (15), 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 (14) to obtain $\mathbf{A}^{q,p+1}$ only by letting $\overrightarrow{a}^{q,p+1}_{\widetilde{i}'+1\,j} = \overrightarrow{a}^{q,p}_{\widetilde{i}'+1\,j} + \Delta_{ij}(\widetilde{i}')$ and $\overrightarrow{a}^{q,p+1}_{\widetilde{i}'j} = \overrightarrow{a}^{q,p}_{\widetilde{i}'j} - \Delta_{ij}(\widetilde{i}')$. If, instead, (17) holds but (16) does not, 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 exchanges, so that P must be finite as well. For any given $\mathbf{A}^{q,p}$ associated to 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 (14) or by (15), depending on which case respectively prevails. 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 (11). The two matrices also satisfy condition (12) 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 operations. 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 operations.

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 Proofs of the theorems

Proof of Theorem 1. We show that *(i)⇒(ii)⇒(iii)⇒(iv)⇒(v)⇒(i).*

(i) ⇒ (ii). This is a consequence of the definition of the axioms.

(ii) ⇒ (iii). The ordering induced by the family of indicators D_h is consistent with axioms *IPC* (D_h is symmetric with respect to classes) and *IEC* (empty classes receive weight $\overline{a}_j = 0$). Consistency with axiom *SC* follows from homogeneity with respect to $\overline{a}_j$: a split of class j into j and j' gives $\lambda \mathbf{a}_j/(\lambda \overline{a}_j) = \mathbf{a}_j/\overline{a}_j = (1-\lambda)\mathbf{a}_j/((1-\lambda)\overline{a}_j)$. Finally, if D_h is consistent with *MC* then h is subadditive that, along with homogeneity, gives that h is also convex (see Proposition B.9.b at p.651 in Marshall et al. 2011).

(iii) ⇒ (iv). Note that $\mathbf{B} \preccurlyeq^R \mathbf{A}$ is equivalent to $\begin{pmatrix} \mathbf{B} \\ \overline{\mathbf{b}}^t \end{pmatrix} \preccurlyeq^R \begin{pmatrix} \mathbf{A} \\ \overline{\mathbf{a}}^t \end{pmatrix}$, where $\overline{\mathbf{b}}^t$ and $\overline{\mathbf{a}}^t$ are row vectors depicting the overall population in each class. Hence condition (10) in Lemma 1 rewrites $\sum_j g(\mathbf{b}_j^t, \overline{b}_j) \leq \sum_j g(\mathbf{a}_j^t, \overline{a}_j)$ with g defined on $\mathbb{R}^{d+1}$. Since g is convex and homogeneous, then $g(\mathbf{a}_j^t, \overline{a}_j) = \overline{a}_j g(\mathbf{a}_j^t/\overline{a}_j, 1) = \overline{a}_j h(\mathbf{a}_j^t/\overline{a}_j)$ for all $h \in \mathcal{H}$, while for convenience empty classes receive weight $\overline{a} = 0$ and do not count in the index computation. Moreover, adding $|n_A - n_B|$ empty classes preserves the relation in (10). By Lemma 1, $D_h(\mathbf{B}) \leq D_h(\mathbf{A})\ \forall h \in \mathcal{H}$ is equivalent to (10) and implies *(iv).*

(iv) ⇒ (v). Recall that $\mathbf{B} \preccurlyeq^R \mathbf{A}$ means that $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ for $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$. The set $\mathcal{R}_{n_A,n_B}$ describes a polytope in $\mathbb{R}_+^{n,m}$. Every $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$ can be written as the convex combination of its vertices, given by all the $H = (n_B)^{n_A}$ (0,1)-matrices of dimension $n_A \times n_B$ with exactly one nonzero element in each row, hereafter denoted as $\mathbf{X}(1), \ldots, \mathbf{X}(h), \ldots, \mathbf{X}(H)$. Hence $\mathbf{B} = \sum_h \lambda_h \mathbf{A} \cdot \mathbf{X}(h)$ with weights $\lambda_h \geq 0\ \forall h$ and $\sum_h \lambda_h = 1$, where h ranges from 1 to H. Following this notation, any column k of $\mathbf{B}$ rewrites $\mathbf{b}_k = \sum_h \lambda_h \mathbf{A} \cdot \mathbf{x}_k(h)$. Using Tonelli's theorem, the weighted sum $\mathbf{z} = \sum_{k=1}^{n_B} \theta_k \mathbf{b}_k$ becomes:

$$\mathbf{z} = \sum_{k=1}^{n_B} \theta_k \left(\sum_h \lambda_h \sum_j \mathbf{a}_j \cdot x_{jk}(h) \right) = \sum_{j=1}^{n_A} \mathbf{a}_j \left(\sum_h \lambda_h \sum_k \ \theta_k x_{jk}(h) \right) = \sum_{j=1}^{n_A} \widetilde{\theta}_j \mathbf{a}_j.$$

Thus,

$$\begin{aligned} Z(\mathbf{B}) &= \left\{ \mathbf{z} := (z_1, \ldots, z_d)^t \ : \ \mathbf{z} = \sum_{k=1}^{n_B} \theta_k \mathbf{b}_k, \quad \theta_k \in [0,1] \ \forall k = 1, \ldots, n_B \right\} \\ &:= \left\{ \mathbf{z} = \sum_{j=1}^{n_A} \widetilde{\theta}_j \mathbf{a}_j, \quad \widetilde{\theta}_j \in \mathcal{I} \subset [0,1] \ \forall j = 1, \ldots, n_A \right\} \subseteq Z(\mathbf{A}), \end{aligned}$$

where the interval $\mathcal{I}$ lies on the $[0,1]$ segment because $x_{jk}(h) = 1$ and $x_{jk'}(h) = 0$ for all $k' \neq k$, and given the restrictions on θ_k. The elements of $\mathcal{I}$, that is the new weights $\widetilde{\theta}_j$, are obtained for a given weighting scheme $(\lambda_1, \ldots, \lambda_H)$. If $\mathcal{I} \subset [0,1]$, matrix majorization implies that every element of $Z(\mathbf{B})$ can be written as an element of $Z(\mathbf{A})$, or equivalently $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$. When $\mathcal{I} = [0,1]$, $Z(\mathbf{B}) = Z(\mathbf{A})$.

(v) ⇒ (i). Let consider $\mathbf{B}, \mathbf{A} \in \mathcal{M}_d$ and use the indices k and j to denote the columns of $\mathbf{B}$ and $\mathbf{A}$ respectively. We show that if $\mathbf{b}_k \in Z(\mathbf{A}) \ \forall k$, then there exists a finite sequence of insertion of empty classes, permutation, split and merge of classes of $\mathbf{A}$ that gives $\mathbf{B}$. Since $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B}$, the columns of $\mathbf{B}$ can be rewritten as:

$$\begin{aligned} \mathbf{b}_k &:= \sum_j \theta_j(k) \mathbf{a}_j, \quad \text{for all } k \in \{1, \ldots, n_B\} \backslash k' \quad \text{and} \\ \mathbf{b}_{k'} &:= \sum_j \theta_j(k') \mathbf{a}_j \ = \ \mathbf{A} \cdot \mathbf{e}_{n_A} - \sum_{k \neq k'} \sum_j \theta_j(k) \mathbf{a}_j = \sum_j \left(1 - \sum_{k \neq k'} \theta_j(k) \right) \mathbf{a}_j \end{aligned}$$

for a generic class k' of $\mathbf{B}$. In shorthand notation

$$\mathbf{B} = \sum_{j=1}^{n_A} (\theta_j(1)\mathbf{a}_j, \ldots, \theta_j(n_B)\mathbf{a}_j). \tag{18}$$

Given that $\theta_j(k) \in [0,1]$ and $\theta_j(k') := \left(1 - \sum_{k \neq k'} \theta_j(k)\right) \in [0,1]$, this implies that $\sum_k \theta_j(k) = 1$. So, each addendum in (18) rewrites:

$$\left(\lambda_{j1}\mathbf{a}_j, \lambda_{j2}(1-\lambda_{j1})\mathbf{a}_j, \ldots, \lambda_{j(n_B-1)} \prod_{1 \leq k < n_B - 1} (1-\lambda_{jk})\mathbf{a}_j, \prod_{1 \leq k \leq n_B - 1} (1-\lambda_{jk})\mathbf{a}_j \right), \tag{19}$$

where $\lambda \in [0,1]$. In fact, every sequence of n_B random numbers $\{\theta(k)\}_{k=1}^{n_B}$ with support in $[0,1]$ satisfying $\sum_k \theta(k) = 1$ can be written as:

$$\begin{aligned} &\theta(1) = \lambda_1 \ \in \ [0,1] \\ &\theta(k) = \lambda_k \left(1 - \sum_{j=1}^{k-1} \theta(j)\right) \quad \text{with} \ \ \lambda_j \in [0,1] \ \forall j = 2, \dots, n_B. \end{aligned} \tag{20}$$

The constraint $\sum_k \theta(k) = 1$ imposes that there must exist an index k such that $\lambda_k = 1$. If $\lambda_k = 1$, then the series is completed and $\lambda_j = 0 = \theta(j)$ for any $j > k$. Note that $\theta(k) = 0$ also if $\lambda_k = 0$, thus the sequence of $\theta(k)$ may also include elements equal to 0 even if it is not yet completed. Solving backward the sequence in (20) leads to (19) given that $\theta(k) = \lambda_k \cdot \prod_{j=1}^{k-1} (1 - \lambda_j)$ with $\lambda_j \in [0,1] \ \forall j$ and $\lambda_k \in [0,1] \ \forall k = 2, \dots, n_B$.

Consider a sequence of matrices $\mathbf{Z}_{[k]} \in \mathcal{R}_A^{SC}$. Matrix $\mathbf{Z}_{[1]}$ performs the first split of vector $\mathbf{a}_j$ according to proportion λ_{j1}. Matrix $\mathbf{Z}_{[2]}$ performs a split on the residual component $(1 - \lambda_{j1})\mathbf{a}_j$ according to the proportion λ_{j2}. The Iteration of these arguments leads to matrix $\mathbf{Z}_{[n_B-1]}$, representing the last split of vector $\mathbf{a}_j$ out of a sequence of $n_B - 2$ splits. It follows that (19) can be equivalently written as:

$$\left(\lambda_{j1}\mathbf{a}_j, \dots, \prod_{1 \le k < n_B - 1} (1 - \lambda_{jk})\mathbf{a}_j, \mathbf{0}_d\right) \cdot \mathbf{Z}_{[n_B-1]} = (\mathbf{a}_j, \mathbf{0}_d, \dots, \mathbf{0}_d) \cdot \prod_{1 \le k \le n_B - 1} \mathbf{Z}_{[k]}. \tag{21}$$

Extending the representation in (21) to all addends in (18) leads to a total of $n_A(n_B - 1) = n$ splits of $\mathbf{A}$'s classes. The split operation preserves the number of classes, therefore it can be operationalized only if there exists a matrix $\mathbf{Y} \in \mathcal{R}_{n_A,n}^{IEC}$ adding a sufficient amount of empty class to $\mathbf{A}$ to perform the n splits. The summation in (18) reveals that the order of the classes of $\mathbf{A}$ is irrelevant. Thus operations of *permutations of classes* are admitted.[29] By combining all the operations in a single row we obtain $\mathbf{A} \cdot \widehat{\mathbf{X}}$, where the $n_A \times n$ matrix $\widehat{\mathbf{X}}$ rewrites:

$$\widehat{\mathbf{X}} \ := \ \mathbf{\Pi}_{n_A} \cdot \mathbf{Y} \cdot \text{diag}\left(\prod_{k=1}^{n_B-1} \mathbf{Z}_{[k]}(1), \dots, \prod_{k=1}^{n_B-1} \mathbf{Z}_{[k]}(n_A)\right) \tag{22}$$

$$= \ \mathbf{\Pi}_{n_A} \cdot \mathbf{Y} \cdot \prod_{j=1}^{n_A} \left(\prod_{k=1}^{n_B-1} \widetilde{\mathbf{Z}}_{[k]}(j)\right), \tag{23}$$

[29]The two operations of permutation and insertion of classes transform $\mathbf{A}$ into

$$\mathbf{A} \cdot \mathbf{\Pi}_{n_A} \cdot \mathbf{Y} := (\mathbf{a}_1, \underbrace{\mathbf{0}_d, \dots, \mathbf{0}_d}_{n_B-1 \text{ times}}, \dots, \mathbf{a}_{n_A}, \underbrace{\mathbf{0}_d, \dots, \mathbf{0}_d}_{n_B-1 \text{ times}}).$$

where $\mathbf{Z}_{[k]}(j)$ is indexed for j to show the relation with the class j in $\mathbf{A}$. Here $\widetilde{\mathbf{Z}}_{[k]}(j) := \text{diag}\left(\mathbf{I},\ \mathbf{Z}_{[k]}(j), \mathbf{I}'\right)$ and $\mathbf{I}$ and $\mathbf{I}'$ are two identity matrices of size $(j-1)n_B$ and $(n_A - j)n_B$ respectively. Line (23) comes from the fact that every block diagonal matrix can be represented as the product of the matrices associated to each block, obtained substituting the remaining blocks with identity matrices.

To conclude, it is possible to perform permutations of $n_A n_B$ classes to rearrange the entries in $\mathbf{A} \cdot \widehat{\mathbf{X}}$ to accommodate the definition of a *merge of classes transformation* through a matrix $\mathbf{\Pi}_{n_A n_B}$. A convenient permutation rearranges n_B groups of n_A-tuples of classes of $\mathbf{A} \cdot \widehat{\mathbf{X}}$, so that the j-th group consists of the sequence of classes $(\lambda_{1j}\mathbf{a}_1, \ldots, \lambda_{n_A j}\mathbf{a}_{n_A}, \ldots)$.[30] Consider a sequence of merges of classes, so that class 1 in the new configuration is merged with class 2, then the resulting class 2 is merged with class 3 and so on, up to the first n_A classes. The sequence of merge transformations can be modeled by matrices $\mathbf{M}_{[1]} \in \mathcal{R}^{MC}_{n_A n_B}$, $\mathbf{M}_{[2]} \in \mathcal{R}^{MC}_{n_A n_B}$ and so on, up to $\mathbf{M}_{[n_A - 1]} \in \mathcal{R}^{MC}_{n_A n_B}$, respectively. Given the order of the classes, the same procedure can be extended to all the $n_B - 1$ remaining n_A-tuples of classes. This operation leaves many empty classes, that can be eliminated using a matrix $\mathbf{Y}'$, incorporating the *elimination of empty classes* operation. As a result:

$$\mathbf{B} = \mathbf{A} \cdot \widehat{\mathbf{X}} \cdot \mathbf{\Pi}_{n_A n_B} \cdot \prod_{1 \le k \le n_B} \left(\prod_{(k-1)n_A < j < k n_A} \mathbf{M}_{[j]} \right) \cdot \mathbf{Y}'.$$

Hence, the condition $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ is mapped into matrix operations transforming $\mathbf{A}$ into $\mathbf{B}$ that can be decomposed into a finite sequence of permutations, insertion or elimination of empty classes, split and merge transformations, which concludes the proof. *Q.E.D.*

Proof of Theorem 2. We show that *(i)⇒(ii)⇒(iii)⇒(iv)⇒(v)⇒(i).*

(i) ⇒ (ii). This is a consequence of the definition of the axioms.

(ii) ⇒ (iii). It is sufficient to verify the consistency of $D_{\mathbf{w}}$ $\forall \mathbf{w} \in \mathcal{W}$ with axiom *E.* This is assured by the compatibility of the index with rank preserving transfers (see Lemma 3): moving ε from group ℓ to h with $\ell > h$ in class j implies a net effect on the overall measured dissimilarity of $(w_{hj} - w_{\ell j})\varepsilon < 0$, since $0 \le w_{hj} \le w_{\ell j}$ for all j.

(iii) ⇒ (iv). The condition in *(iii)* holds, in particular, when $\sum_{j=1}^{n} w_{ij} = 1$ and $w_{ij} = 0$ for any $j \neq k$ and i. Since j can be any class, inequality evaluations can be separated across classes. Hence, $\sum_{i=1}^{d} w_{ij} \overrightarrow{b}_{(i)j} \le \sum_{i=1}^{d} w_{ij} \overrightarrow{a}_{(i)j}$ for all j holds

[30] Formally: $\mathbf{A} \cdot \mathbf{X} \cdot \mathbf{\Pi}_{n_A n_B} = (\lambda_{11}\mathbf{a}_1, \ldots, \lambda_{n_A 1}\mathbf{a}_{n_A}, \ldots, \lambda_{1 n_B}\mathbf{a}_1, \ldots, \lambda_{n_A\ n_B}\mathbf{a}_{n_A})$.

as well, 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 $\sum_i \overrightarrow{b}_{(i)j} = \sum_i \overrightarrow{a}_{(i)j}$). For any j, let $w_{ij} = 0$ for $i = 1, 2, \ldots, k$ and $w_{ij} = 1/(d-k)$ for $i = k+1, \ldots, d$, for any $k = 1, \ldots, d$. The latter inequality gives $\sum_{i=1}^{k} \overrightarrow{b}_{(i)j} \geq \sum_{i=1}^{k} \overrightarrow{a}_{(i)j}$ $\forall k$ and $\forall j$, which is $\overrightarrow{\mathbf{b}}_j^t$ Lorenz dominates $\overrightarrow{\mathbf{a}}_j^t$ for all classes j. This is claim *(iv)*.

(iv) $\Rightarrow$ *(v)*. Statement *(iv)* implies (see Theorem A.2 in Marshall et al. 2011, p.30) that $conv\{\mathbf{\Pi}_d \cdot \overrightarrow{\mathbf{b}}_k : \mathbf{\Pi}_d \in \mathcal{P}_d\} \subseteq conv\{\mathbf{\Pi}_d \cdot \overrightarrow{\mathbf{a}}_k : \mathbf{\Pi}_d \in \mathcal{P}_d\}$ for every $k = 1, \ldots, n$, where the *conv* operator indicates the convex hull. This condition can be extended to all vectors originating from liner combinations of adjacent classes of the cumulative distribution matrices. Recall the definition of Lorenz dominance. For any k and $k+1$, $\overrightarrow{\mathbf{b}}_k^t$ Lorenz dominates $\overrightarrow{\mathbf{a}}_k^t$ and $\overrightarrow{\mathbf{b}}_{k+1}$ Lorenz dominates $\overrightarrow{\mathbf{a}}_{k+1}$ implies $\theta\sum_i^j \overrightarrow{b}_{ik} + (1-\theta)\sum_i^j \overrightarrow{b}_{ik} \geq \theta\sum_i^j \overrightarrow{a}_{ik} + (1-\theta)\sum_i^j \overrightarrow{a}_{ik}$ $\forall j = 1, \ldots, n$ and $\forall \theta \in [0,1]$. This condition is Lorenz dominance for combinations of vectors. The entries of the vectors are always arranged in increasing order, hence one has that the Lorenz dominance condition gives

$$conv\left\{\mathbf{\Pi}_d \cdot \left(\theta\overrightarrow{\mathbf{b}}_k + (1-\theta)\overrightarrow{\mathbf{b}}_{k+1}\right) : \mathbf{\Pi}_d \in \mathcal{P}_d\right\} \subseteq conv\left\{\mathbf{\Pi}_d \cdot \left(\theta\overrightarrow{\mathbf{a}}_k + (1-\theta)\overrightarrow{\mathbf{a}}_{k+1}\right) : \mathbf{\Pi}_d \in \mathcal{P}_d\right\}$$

$\forall k$ $\forall \theta \in [0,1]$. By definition of the Path Polytope, this condition implies $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$.

(v) $\Rightarrow$ *(i)*. Statement *(v)* implies that $MP^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$, i.e. $\forall \mathbf{p} \in MP^*(\mathbf{B})$ there exists a vertex $\mathbf{v} \in Z^*(\mathbf{A})$ satisfying $\mathbf{e}_d \cdot \mathbf{p} = \mathbf{e}_d \cdot \mathbf{v}$ such that $\mathbf{p} = conv\{\mathbf{\Pi}_d \cdot \mathbf{v} : \mathbf{\Pi}_d \in \mathcal{P}_d\}$. Consider a sequence $\mathbf{p}_1, \ldots, \mathbf{p}_n$ such that $\mathbf{p}_k = \overrightarrow{\mathbf{b}}_k$ $\forall k = 1, \ldots, n$. The vertices $\mathbf{v}_k$ corresponding to $\overrightarrow{\mathbf{b}}_k$ $\forall k$ can be arranged into a matrix $\overrightarrow{\mathbf{V}} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$, interpreted as a cumulative distribution matrix. The underlying distribution matrix $\mathbf{V} = (\Delta\mathbf{v}_1, \ldots, \Delta\mathbf{v}_n)$ is such that $\Delta\mathbf{v}_k = \mathbf{v}_k - \mathbf{v}_{k-1} \geq \mathbf{0}$ and $\Delta\mathbf{v}_1 = \mathbf{v}_1$. If every $\mathbf{v}_k$ is properly selected to be comonotonic with $\overrightarrow{\mathbf{b}}_k$ $\forall k$, then $\mathbf{V}$ and $\mathbf{B}$ are ordinal comparable, thus $\mathbf{V} = \mathbf{A}$ by construction. It follows that $\overrightarrow{\mathbf{b}}_k = conv\left\{\mathbf{\Pi}_d \cdot \overrightarrow{\mathbf{a}}_k : \mathbf{\Pi}_d \in \mathcal{P}_d\right\}$ or, equivalently, $\overrightarrow{\mathbf{b}}_k^t$ Lorenz dominates $\overrightarrow{\mathbf{a}}_k^t$ (see Theorems 1.A.2 and 2.B.2 in Marshall et al. 2011). By Lemma 3, $\mathbf{B}$ can be obtained from $\mathbf{A}$ through a finite sequence of exchange operations, thus *(i)* holds. *Q.E.D.*

Proof of Theorem 3. *(i)* $\Rightarrow$ *(ii)*. Suppose *(i)* is true but there exists a pair of ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ that are obtained from $\mathbf{A}$ and $\mathbf{B}$ through a sequence of insertion/elimination of empty classes, splits of classes and interchanges of groups, for which there exists an ordering $\preccurlyeq$ satisfying E that ranks $\mathbf{A}^* \prec \mathbf{B}^*$ (where $\prec$ denotes the asymmetric part of $\preccurlyeq$). Since this ordering is defined on the space of ordinal comparable distribution matrices, then it can be uniquely extended to $\mathcal{M}_d$ using axioms *IEC*, *SC* and *I*. There exists therefore an ordering $\preccurlyeq$ satisfying *E*, *IEC*, *SC* and *I* for which $\mathbf{A}^* \prec \mathbf{B}^*$ and by transitivity $\mathbf{A} \prec \mathbf{B}$. A contradiction.

(ii) $\Rightarrow$ *(i)*. If *(ii)*, then $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ must hold also for all $\preccurlyeq$ satisfying *E and IEC*, *SC* and *I*. By transitivity, $\mathbf{B}^* \sim \mathbf{B} \preccurlyeq \mathbf{A} \sim \mathbf{A}^*$, which is *(i)*. *Q.E.D.*

Proof of Corollary 1. By Theorem 1, (7) is equivalent to $\mathbf{B}' = \mathbf{A}' \cdot \mathbf{X}$ for $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$, which gives condition *(i)*. Each entry in the first row of $\mathbf{A}'$ is a constant equal to $1/n_A$, so it can be transformed by $\mathbf{X}$ into the corresponding element in $\mathbf{B}'$, equal to $1/n_B$, only by multiplying each single entry by n_A/n_B, thus *(ii)* should also hold. *Q.E.D.*

References

Alonso-Villar, O. and del Rio, C. (2010). Local versus overall segregation measures, *Mathematical Social Sciences* **60**: 30–38.

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

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.

Atkinson, A. and Marlier, E. (2010). Analysing and measuring social inclusion in a global context, *Technical report*, United Nations.

Blackwell, D. (1953). Equivalent comparisons of experiments, *The Annals of Mathematical Statistics* **24**(2): 265–272.

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities, *The American Economic Review* **85**(3): 365–390.

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

Chakravarty, S. R. and Silber, J. (2007). A generalized index of employment segregation, *Mathematical Social Sciences* **53**(2): 185 – 195.

Dahl, G. (1999). Matrix majorization, *Linear Algebra and its Applications* **288**: 53 – 73.

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.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes, *American Sociological Review* **20**(2): 210–217.

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

Ebert, U. and Moyes, P. (2003). Equivalence scales reconsidered, *Econometrica* **71**(1): 319–343.

Echenique, F., Fryer, Roland G., J. and Kaufman, A. (2006). Is school segregation good or bad?, *American Economic Review* **96**(2): 265–269.

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

Flückiger, Y. and Silber, J. (1999). *The Measurement of Segregation in the Labour Force*, Physica-Verlag, Heidelberg, Germany.

Frankel, D. M. and Volij, O. (2011). Measuring school segregation, *Journal of Economic Theory* **146**(1): 1 – 38.

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

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**.

Giovagnoli, A., Marzialetti, J. and Wynn, H. (2009). Bivariate dependence orderings for unordered categorical variables, *in* L. Pronzato and A. Zhigljavsky (eds), *Optimal Design and Related Areas in Optimization and Statistics*, Vol. 28 of *Optimization and its Applications*, Springer, New York., chapter 4.

Grant, S., Kajii, A. and Polak, B. (1998). Intrinsic preference for information, *Journal of Economic Theory* **83**(2): 233 – 259.

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

Hasani, A. and Radjabalipour, M. (2007). On linear preservers of (right) matrix majorization, *Linear Algebra and its Applications* **423**(2-3): 255 – 261.

Hutchens, R. M. (1991). Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations, *Mathematical Social Sciences* **21**(1): 31 – 51.

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

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

Kolm, S.-C. (1977). Multidimensional egalitarianisms, *The Quarterly Journal of Economics* **91**(1): 1–13.

Koshevoy, G. and Mosler, K. (1996). The Lorenz zonoid of a multivariate distribution, *Journal of the American Statistical Association* **91**(434): 873–882.

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

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

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

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

McMullen, P. (1971). On zonotopes, *Transactions of the American Mathematical Society* **159**: 91–109.

Reardon, S. F. (2009). Measures of ordinal segregation, *in* J. S. Yves Flckiger, Sean F. Reardon (ed.), *Occupational and Residential Segregation (Research on Economic Inequality)*, Vol. 17, Emerald Group Publishing Limited, pp. 129–155.

Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation, *Sociological Methodology* **32**: 33–67.

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

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.

Torgersen, E. (1992). *Comparison of Statistical Experiments*, Vol. 36 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press.

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

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

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

Yalonetzky, G. (2012). Measuring group disadvantage with inter-distributional inequality indices: A critical review and some amendments to existing indices, *Economics: The Open-Access, Open-Assessment E-Journal* **6**(2012-9).

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