ECINEQ

Society for the Study of Economic Inequality

Working Paper Series

Robust dissimilarity comparisons with categorical outcomes

Francesco Andreoli
Claudio Zoli

ECINEQ WP 2019 - 502

ECINEQ 2019 - 502
August 2019

www.ecineq.org

Robust dissimilarity comparisons with categorical outcomes*

Francesco Andreoli†
University of Verona and LISER, Italy

Claudio Zoli
University of Verona, Italy

Abstract

The analysis of many economic phenomena requires partitioning societies into groups, identified for instance, by gender, ethnicity, birthplace, education, age or parental background, and studying the extent at which these groups are distributed with different intensities across relevant outcomes, like jobs, locations, schools, policy treatments. When the groups are similarly distributed, their members could be seen as having equal chances to achieve any of the attainable outcomes. Otherwise, a form of dissimilarity prevails. We frame dissimilarity comparisons of multi-group distributions defined over categorical outcomes by showing the equivalence between axioms underpinning information criteria, majorization conditions, agreement between dissimilarity indicators and new empirical tests based on Zonotopes inclusion. Mainstream approaches to two- and multi-group segregation as well uni- and multivariate inequality analysis are shown to be nested within the dissimilarity model.

Keywords: Dissimilarity, segregation, inequality, majorization, Zonotopes, axiomatic.

JEL Classification: D63, J71, J62, D30.

*This paper forms part of the research projects The Measurement of Ordinal and Multidimensional Inequalities (Contract No. ANR-16-CE41-0005-01 - ORDINEQ) of the French National Agency for Research, the research project MOBILIFE (grant RBVR17KFHX) of the University of Verona Basic Research Grant and the NORFACE/DIAL project IMCHILD: The impact of childhood circumstances on individual outcomes over the life-course (grant INTER/NORFACE/16/11333934/IMCHILD) of the Luxembourg National Research Fund (FNR), whose financial support is gratefully acknowledged.

†Corresponding author: Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy and Luxembourg Institute of Socio-Economic Research (LISER), 11 Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg. E-mail: francesco.andreoli@univr.it.

1 Introduction

Many economic phenomena are concerned with the way individuals, partitioned into different social groups on the basis of characteristics they share in common, are distributed across relevant outcome categories. Groups' distributions could specify, for instance, how people of different ethnic origin (the groups) are assigned with different intensity to the neighborhoods of a city (the classes of realization). Distributions in which these ethnic groups are more evenly located across neighborhoods are less *segregated.* Likewise, school/occupational segregation is concerned with the uneven distribution of ethic groups across schools/jobs.

In other cases, the interest lies on the way one of more assignable attributes (such as income, consumption or wealth) are distributed among individuals or families, and the extent at which this distribution differs from a normatively relevant benchmark. *Income inequality*, for instance, arises when the distribution of income shares across income units differs from the distribution of the demographic weights of these units.

All these examples are concerned with the extent of *dissimilarity* between distributions defined over categorical outcomes. A convenient (and equivalent) way of representing distinct sets of distributions is by adopting a matrix notation. We use distribution matrices to depict configurations of data, each matrix representing, by row, the distribution of individuals belonging to a given group across classes of realizations of an attainable outcome. In the example below we consider two distribution matrices, each with three groups and two classes of realizations:

$$\mathbf{A}: \begin{array}{lcc} & \text{Class 1} & \text{Class 2} \\ \text{Group 1} & 0.9 & 0.1 \\ \text{Group 2} & 0.1 & 0.9 \\ \text{Group 3} & 0.8 & 0.2 \end{array} \quad \textit{and} \quad \mathbf{B}: \begin{array}{lcc} & \text{Class 1} & \text{Class 2} \\ \text{Group 1} & 0.6 & 0.4 \\ \text{Group 2} & 0.4 & 0.6 \\ \text{Group 3} & 0.55 & 0.45 \end{array}. \tag{1}$$

We interpret entries of these matrices as frequencies so that, for instance, the share of group 3 in class 2 in **A** is 20%.[1]

This paper sets out the conditions for robustly ranking distribution matrices such as **A** and **B** by the extent of dissimilarity they display. There is widespread agreement in the literature on what constitutes lack of segregation or equality: these are situations in which the groups are *similarly distributed* across the possible outcomes. The relevant notion of similarity dates back to Gini (1914), who argues that two (or more) groups are similarly distributed whenever "the overall populations of the two groups take the same values with the same frequency."[2] While one can count on well-established methodology for the analysis of lack of similarity between *two* distributions, substantial disagreement persists for what concerns the extension of these results in the *multi-group* setting.

We develop an axiomatic foundation for the measurement of multi-group dissimilarity. The focus is on partial orders of dissimilarity among distributions defined over discrete, non-ordered categorical outcomes. Our main result establishes a complete characterization of the dissimilarity partial order that links the mathematical, economical and statistical aspects of dissimilarity, thereby complementing the literature on multidimensional analysis of ordinal attributes (see for instance Atkinson and Bourguignon 1982, Dardanoni 1993, Gravel and Moyes 2012).

To do so, we consider a number of alternative criteria for ranking distributions in terms of their dissimilarity and we prove that the rankings generated by these criteria are in fact equivalent. These equivalences are reminiscent of well known results in the literatures on the measurement of inequality and risk (see Hardy, Littlewood and Polya 1934, Marshall, Olkin and Arnold 2011, Gajdos and Weymark 2012). For example, when comparing income

[1]The multidimensional nature of the problem arises from the number of distributions to be compared (the groups) and not by the domain of these discrete distributions (the classes, which by convenience are limited to two in the example).

[2]Gini (1914, p. 189), translated from Italian, formalizes similarity through the notion of proportionality: "If n is the size of group α, m is the size of group β, n_x the size of group α 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, $\frac{n_x}{m_x} = \frac{n}{m}$."

distributions in terms of their inequality, the following statements about income distribution vectors $\mathbf{x}$ and $\mathbf{y}$ turn out to be equivalent: (i) $\mathbf{x}$ can be obtained from $\mathbf{y}$ by a finite sequence of Pigou-Dalton (rich-to-poor) transfers or permutations, (ii) $\mathbf{y}$ exhibits at least as much inequality as $\mathbf{x}$ for every inequality index that treats individuals symmetrically and that regards a Pigou-Dalton transfer as an elementary inequality-reducing transformation, (iii) the inequality in $\mathbf{x}$ is not larger than the one in $\mathbf{y}$ for all additive decomposable inequality indices with increasing convex aggregators of individual incomes, (iv) $\mathbf{x}$ can be obtained by multiplying $\mathbf{y}$ by a bistochastic matrix, and (v) $\mathbf{x}$ Lorenz dominates $\mathbf{y}$.

Our main result proves the equivalence between the dissimilarity measurement analogues of each of these five claims. Claim (i) states that dissimilarity comparisons can be operationalized by specific *transformations* of the data, or sequences of them. We argue that some transformations preserve dissimilarity across distributions, while a specific operation called the *merge* transformation cannot increase (but not necessarily preserve) dissimilarity. This transformation, when applied to a distribution matrix, produces a convex combination of groups' proportions across two outcome realizations, thus reducing the extent of information about the groups membership one can gather from the knowledge of the outcome realization.[3] Claim (ii) restricts the focus on those orderings that rank distribution matrices consistently with the effect of these transformations. The dissimilarity partial order arises from the agreement of these orderings (Donaldson and Weymark 1998). Claim (iii) narrows down agreement to a specific parametric family of indices depicting dissimilarity as the average degree of dispersion in proportions of the groups in correspondence of each outcome realization. Claim (iv) relates more formally dissimilarity to informativeness (Grant et al. 1998) with a *majorization* result (see Marshall et al. 2011): one distribution does not display more dissimilarity than another if and only if the former *matrix majorizes*

[3]The merge transformation can be associated to the notion of non-proportional split of school invoked by Frankel and Volij (2011) in school segregation analysis or to *linear bifurcations* of a probability function (Grant, Kajii and Polak 1998), and has parallels in the analysis of multivariate inequality (Gajdos and Weymark 2005).

the latter (Dahl 1999).[4] Claims (i)-(iv) define robust and equivalent conditions to rank distribution matrices such as **A** and **B** in (1), which cannot be tested empirically. We introduce with claim (v) a new empirical test for the dissimilarity order which is based on the inclusion of the *Zonotope* representation of the distributions matrices (i.e., the data).[5]

We provide an unified theory for the analysis of multi-group dissimilarity that, on the one hand, nests all results relate to comparisons of two distributions, while on the other hand is based on equivalent characterizations that incorporate the multivariate nature of the problem. In this way, we avoid the undesirable consequences of ranking multi-group distributions on the basis of pairwise comparisons of groups. For instance, one can easily verify that any pair of groups distributions of matrix **B** display less dissimilarity than the corresponding pair of groups distributions from matrix **A**, according to standard bivariate analysis.[6] Nonetheless, the ranking is not preserved, but it is rather reversed, as a result of aggregating groups distributions. To see this, consider the distribution matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ below, obtained from **A** and **B** respectively by mixing distributions of groups 1 and 2 with weights 0.875 and 0.125:

$\tilde{\mathbf{A}}$:		Class 1	Class 2	;	$\tilde{\mathbf{B}}$:		Class 1	Class 2	.
	Groups 1 & 2	0.8	0.2			Groups 1 & 2	0.575	0.425	
	Group 3	0.8	0.2			Group 3	0.55	0.45	

[4]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 categorical variables (Giovagnoli, Marzialetti and Wynn 2009), among others.

[5]Zonotopes are multi-group generalizations of Lorenz and segregation curves and their inclusion tests can be implemented via standard linear programming methods. Detailed references can be found on McMullen (1971) and Ziegler (1995). Routines implementing the inclusion test are made available by the authors.

[6]The proof unravels as follows. Denote $\mathbf{A}(i,h)$ the distribution matrix obtained by isolating only groups $i,h \in \{1,2,3\}$, and similarly $\mathbf{B}(i,h)$. Denote a_{ij} a generic element of **A**. Compute a_{i1}/a_{h1} and a_{i2}/a_{h2} and select the class, either 1 or 2, with the largest value. This value can be compared to that obtained from $\mathbf{B}(i,h)$, which turns out to be always smaller than that from $\mathbf{A}(i,h)$ for any pair $i \neq h$. This procedure is equivalent to rank distributions $\mathbf{A}(i,h)$ and $\mathbf{B}(i,h)$ by their segregation curves (Hutchens 1991).

Distributions in matrix $\tilde{\mathbf{A}}$ are similar, implying that now $\tilde{\mathbf{A}}$ unambiguously displays less dissimilarity than $\tilde{\mathbf{B}}$.

The interest of a multi-group characterization lies in its robustness against paradoxical results, guaranteeing that if $\mathbf{A}$ displays at least as much dissimilarity than $\mathbf{B}$, then $\tilde{\mathbf{A}}$ displays at least as much dissimilarity than $\tilde{\mathbf{B}}$ for any way of reducing the original problem from many to few (two) distributions. This paper explicitly considers the multi-group nature of the problem, and nests most results based on two-groups comparisons as special cases. After describing the setting in Section 2, we set out the main result in Theorem 1: it states the equivalence of claims (i)-(v) and is given in Section 3 (all proofs are collected in the Appendix). The theorem has immediate consequences for the conceptualization, measurement, and multi-group extension of dissimilarity analysis in empirical research. We show in Section 4 that a variety of sparse and apparently unrelated results on the measurement of segregation[7] come down to assess how much dissimilarity there is between two or more distributions. Even inequality[8] comparisons involve specific dissimilarity evaluations, although the converse is not necessarily true.

2 Setting

2.1 Notation

We compare *distribution matrices* of size $d \times n$, depicting sets of distributions (indexed by rows) of $d \geq 1$ *groups* across $n \geq 2$ disjoint *classes* (indexed by columns), each corresponding to a specific category of non-ordered realizations. We develop dissimilarity comparisons of distribution matrices with a fixed number d of groups and a variable number of classes.

[7]We refer to Duncan and Duncan (1955), Hutchens (1991), Reardon and Firebaugh (2002) and Reardon (2009) for the analysis of segregation between two groups, and to Flückiger and Silber (1999), Chakravarty and Silber (2007), Alonso-Villar and del Rio (2010), Lasso de la Vega and Volij (2014) and Frankel and Volij (2011) for a survey on multi-group extensions. For an analysis of the economic motivations behind segregation comparisons, see for instance Echenique, Fryer and Kaufman (2006) and Borjas (1995).

[8]Reference papers dealing with multivariate inequality that can be related to our dissimilarity model are Kolm (1977), Koshevoy and Mosler (1996) and Ebert and Moyes (2003).

These matrices are collected in the set

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

where a_{ij} is the proportion of group i observed in class j. In the analysis of school segregation by ethnic origin of the students, for instance, classes would represent schools and a_{ij} would be the proportion of students with ethnicity i that are enrolled in school j. The column vector $\mathbf{a}_j$ collects the proportions of all groups in school j. The distribution matrices in $\mathcal{M}_d$ are hence *row stochastic*, meaning that matrix $\mathbf{A} \in \mathcal{M}_d$ represents a collection of d elements of the unit simplex Δ^{n_A}. We interpret the rows of $\mathbf{A}$ as distributions of groups frequencies.

We also consider *transformation matrices*, representing the linear transformations 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.[9] 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.[10] Finally, boldface letters always indicate column vectors, with $\mathbf{1}_n := (1, \ldots, 1)^t$ and $\mathbf{0}_n := (0, \ldots, 0)^t$, where the superscript t denotes transposition.

2.2 Dissimilarity orders

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

[9]The entries x_{ij} of matrix $\mathbf{X} \in \mathcal{R}_{n,m}$ can be interpreted as the probability that the population in class i in the distribution of origin “migrates” to class j in the distribution of destination.

[10]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.

at most by one group *and* each group occupies separate classes. Thus, the distributions of the groups $\mathbf{d}_1^t \in \Delta^{n_1}, \ldots, \mathbf{d}_d^t \in \Delta^{n_d}$ do not overlap across classes. In compact notation:

$$\mathbf{S} := \begin{pmatrix} \mathbf{s}^t \\ \vdots \\ \mathbf{s}^t \end{pmatrix} \quad \text{and} \quad \mathbf{D} := \begin{pmatrix} \mathbf{d}_1^t & \ldots & \mathbf{0}_{n_d}^t \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{n_1}^t & \ldots & \mathbf{d}_d^t \end{pmatrix}. \tag{2}$$

In the first case, $\mathbf{S}$, all groups are equally represented with the same intensity in each class. Conversely, in the second case, $\mathbf{D}$, it is possible to forecast the group occupying each class.[11]

This paper investigates the possibility of ordering distribution matrices according to the dissimilarity they display. A *dissimilarity ordering* is a complete and transitive binary relation $\preccurlyeq$ on the set $\mathcal{M}_d$ with symmetric part $\sim$, that ranks $\mathbf{B} \preccurlyeq \mathbf{A}$ whenever $\mathbf{B}$ is *at most as dissimilar as* $\mathbf{A}$.[12] Given $\mathbf{A} \in \mathcal{M}_d$, any dissimilarity ordering should rank $\mathbf{S} \preccurlyeq \mathbf{A} \preccurlyeq \mathbf{D}$ for any perfect similarity matrix $\mathbf{S}$ and for any maximal dissimilarity matrix $\mathbf{D}$. There are infinitely many matrices that can be represented as $\mathbf{S}$ and $\mathbf{D}$ in (2). They are all regarded as equivalent representations of perfect similarity or of maximal dissimilarity, the focus being on differences across group distributions and not on the degree of heterogeneity in the distribution of each group across realizations.The condition $d \leq n$ is, nevertheless, necessary for $\mathbf{D}$ to exist. If $\mathbf{A}$ is such that $d > n$, then it can display some dissimilarity, but not maximal dissimilarity.

In what follows, we characterize the dissimilarity partial order induced by the intersection of the dissimilarity orderings (Donaldson and Weymark 1998) satisfying desirable properties.

[11]The condition of lack of overlapping between distributions $\mathbf{d}_1^t, \ldots, \mathbf{d}_d^t$ represents the case in which group identity and realizations display the highest degree of *connectivity*, a condition regarded to in Gini (1914) and subsequent literature (see Bertino, Drago, Landenna, Leti and Marasini 1987) as the maximal dissimilarity scenario.

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

3 Characterization of the dissimilarity order

3.1 Axioms

We introduce axioms defining the change in dissimilarity that should be registered by every dissimilarity ordering when data are transformed according to some specific operations. These operations apply directly to distribution matrices. The first axiom characterizes the context, introducing an anonymity property with respect to the labels (and hence the arrangement) of the classes of a distribution matrix.

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* $\mathbf{B} \sim \mathbf{A}$.

Axiom *IPC* restricts the focus to evaluations where the classes cannot be meaningfully ordered. To see the implications of the axiom, consider the problem of measuring schooling segregation. It can be conceived as a problem of dissimilarity in the distributions of groups of students with different ethnic background across the schools of a school district. The *IPC* axiom posits that the name of the schools is irrelevant to conclude about the dissimilarity in the distributions of students across these schools. This is arguably the case if the schools should not be ordered according to additional information, for instance on their performances, their quality or their budget. Another implication of axiom *IPC* is that any distribution matrix that is obtained by a permutation of matrix $\mathbf{D}$ columns has to be regarded to as an equivalent representation of maximal dissimilarity.

Dissimilarity comparisons might also be independent of the label assigned to the groups, so that the focus shifts from labels of the groups to their distributions. This is formalized by 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* $\mathbf{B} \sim \mathbf{A}$.

We consider now two transformations that extend comparability to distribution matrices that differ in the number of the classes. The first transformation consider the *insertion or elimination of empty classes*, i.e., classes that are not occupied by groups. The operation consists in adding/eliminating column vectors of size d with only zero entries to/from the original distribution matrix. In the schooling segregation example, the operation corresponds to adding/eliminating schools with no students to/from the same school district. Admittedly, the presence of these schools in the district is irrelevant for assessing schooling segregation therein. We retain with the Independence of Empty Classes (*IEC*) axiom that this transformation is a source of indifference for every dissimilarity ordering.

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

The *IEC* axiom places the emphasis on dissimilarity originated from non-empty columns of a distribution matrix. If $\mathbf{A}$ and $\mathbf{B}$ differ only because of $|n_A - n_B|$ empty classes in one of the two matrices, then the dissimilarity in $\mathbf{A}$ should be regarded to as an equivalent representation of that in $\mathbf{B}$.

The second transformation considered increases the number of classes by *splitting proportionally* (the groups densities in) a class into two new classes. This transformation requires to replicate one column of a distribution matrix and then to scale the entries of the original and of the replicated columns by the splitting coefficients $\beta \in (0, 1)$ and $1 - \beta$, respectively. This operation guarantees that the resulting distribution matrix is row stochastic and that the degree of proportionality of the groups frequencies in the new columns coincides with that in the original column. In the schooling segregation example, splitting a school would require to randomly allocate its students population (i.e., irrespectively of their group assignment) into two smaller institutes, so that ethnic proportions in the two new institutes are not altered. There is agreement in the literature that the split transformation should not affect segregation.[13] According to the Split of Classes (SC) axiom we assume that the transformation described above is a source of indifference for every dissimilarity ordering.

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

The SC axiom highlights that dissimilarity arises from the disproportionality of the groups composition in some classes. A split transformation increases the number of classes and modifies the shape of a distribution matrix, but it does not alter the proportionality of the groups. For this reason, it is regarded to as dissimilarity preserving.

The *merge of classes* transformation complements the split operation. A merge consists in adding together, distribution by distribution, the group proportions observed in two classes. A merge of classes is implemented by vector summation of two adjacent columns of a distribution matrix. The operation has an immediate interpretation in the schooling segregation example: it consists in merging all students from two neighboring schools into a single, larger school. Each ethnic group in the school of destination is increased by an amount equal to the proportion of the corresponding group in the school of departure, which is then emptied. If one or both schools are empty, segregation does not increase nor decreases. Consider, instead, the case of two ethnic groups that are similarly distributed across almost all schools in a district, apart from two schools, such that a group is over-represented compared to the other in one school, and under-represented in the other school. Merging each of these two schools with other schools in the district would reduce the compositional differences. Merging these two schools together would establish proportionality in ethnic composition across all schools, leading to perfect similarity. The Dissimilarity

[13] Frankel and Volij (2011) advocate a similar property (composition invariance) in the study of multi-group school segregation (see also James and Taeuber 1985).

Decreasing Merge of Classes (MC) axiom states that every merge of classes transformation cannot increase dissimilarity.

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

Axioms *MC*, *IEC*, *SC* and *IPC* are independent. Altogether, they characterize dissimilarity as disproportionality in the composition of the groups shares within each class. In fact, splitting classes and inserting empty classes cannot improve on proportionality and by merging two classes, on the contrary, disproportionality in groups composition cannot be increased. These intuitions pave the way for characterizing the dissimilarity partial order coherent with these axioms.

3.2 Characterization

If the distribution matrix $\mathbf{B}$ can be obtained from $\mathbf{A}$ through a sequence of transformations implied by the axioms *IPC*, *IEC*, *SC* and *MC*, then every ordering consistent with these axioms should not rank $\mathbf{A}$ as more dissimilar than $\mathbf{B}$. We represent consensus among these orderings in three distinct ways, each appealing to a specific perspective about dissimilarity. First, we look at functional representations of dissimilarity orderings, i.e. a dissimilarity index, which can be employed to quantify and compare patterns of dissimilarity across distribution matrices. The second way of representing consensus among orderings draws on the statistical foundations of majorization, describing rising dissimilarity as an improvement on the extent at which the class of realization is informative about the identity of the group achieving it. The third representation of consensus involves a statistical test.

Dissimilarity indices. This paper studies the intersection of dissimilarity orderings. We consider here those orderings that can be represented by dissimilarity indices, which have an appeal for empirical research. A dissimilarity index is a function $D : \mathcal{M}_d \rightarrow \mathbb{R}$ mapping a distribution matrix into a number, regarded to as the degree of dissimilarity therein. We consider a family of indices measuring dissimilarity as the average of within-class *dispersion* of group frequencies. The degree of dispersion in each class is quantified by a function h in the class $\mathcal{H}$ of real valued convex functions defined on Δ^d. For a given h, a larger dispersion in groups frequencies within a class always indicates an increase in dissimilarity across groups. The convexity of the function h captures this effect. The evaluation of the dispersion within class j contributes to the overall dissimilarity proportionally to the "size" of that class, denoted $\overline{a}_j := \mathbf{1}_d^t \cdot \mathbf{a}_j$. The dissimilarity index D_h with $h \in \mathcal{H}$ aggregates these evaluations, and is defined as[14]

$$D_h(\mathbf{A}) := \frac{1}{d} \sum_{j=1}^{n_A} \overline{a}_j \cdot h\left(a_{1j}/\overline{a}_j, \ldots, a_{dj}/\overline{a}_j\right), \tag{3}$$

where $a_{ij}/\overline{a}_j$ can be interpreted as the proportion of group i relative to the size of class j. Dissimilarity is minimized when $a_{ij}/\overline{a}_j = 1/d$ for each of the d groups in all classes. Hence, by setting $h\left(\frac{1}{d}\mathbf{1}_d^t\right) = 0$ the index can be normalized to 0 when perfect similarity is reached.

A robust dissimilarity evaluation would require to verify that $D_h(\mathbf{B}) \leq D_h(\mathbf{A})$ holds for every $h \in \mathcal{H}$, thus expressing unanimous consensus over changes in dissimilarity. The condition also implies agreement on the fact that the composition of the groups in $\mathbf{A}$ classes is at least as informative about the group identity of any randomly selected occupant than is the composition of the groups in $\mathbf{B}$ classes.

Majorization. In the context of comparisons of statistical experiments with finite number of outcomes, Blackwell (1953) has formalized a precise condition for "$\mathbf{A}$ is at least as informative as $\mathbf{B}$", which consists in checking that $\mathbf{B}$ is *matrix majorized* by $\mathbf{A}$. Matrix majorization is denoted $\mathbf{B} \preccurlyeq^R \mathbf{A}$ (see also Dahl 1999), meaning that there exists a row stochastic matrix $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$, representing a set of linear transformations of the original

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

data, such that $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$. The notion of matrix majorization has been investigated in a variety of contexts (see p. 625 in Marshall et al. (2011) and literature therein). In the spirit of the fundamental theorem for inequality measurement by Hardy et al. (1934), the main result of this section consists in a set of equivalences between data transformations, agreement among dissimilarity orderings and matrix majorization. None of these criteria, however, can be empirically tested.

Dissimilarity tests. When the interest is not on measuring dissimilarity but, rather, on testing the ranking of distribution matrices in a way consistent with consensus of all inequality indices, it is meaningful to advocate for geometric representations of the data. When the focus is on income distributions, for instance, Lorenz curves allow to conclude on robust inequality rankings of the distributions. The problem of testing for dissimilarity is, nonetheless, more complex than that of testing for income inequality, insofar it involves testing over multiple distributions without resorting to a benchmark distribution (such as that expressing equality) as a reference. Comparing Lorenz curves is therefore not sufficient to conclude on dissimilarity ranking of distribution matrices.

We propose an empirically implementable criterion based on *Zonotopes inclusion* that identifies the dissimilarity partial order of distribution matrices. In the multi-group setting, the Zonotope $Z(\mathbf{A}) \subseteq [0,1]^d$ of any matrix $\mathbf{A} \in \mathcal{M}_d$ is defined as:

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

Every element of the Zonotope is obtained from the Minkowski sum of the vectors with coordinates given by $\mathbf{A}$'s classes. The Zonotope's graph is therefore a convex polytope symmetric with respect to the point $\frac{1}{2}\mathbf{1}_d$ (see McMullen 1971). The maximum Dissimilarity Zonotope is the d-dimensional hypercube and corresponds to $Z(\mathbf{D})$. Its diagonal is the Similarity Zonotope, which corresponds to $Z(\mathbf{S})$. All distribution matrices displaying some dissimilarity originate Zonotopes that lie in $Z(\mathbf{D})$ and that share the reference diagonal $Z(\mathbf{S})$, for any matrix $\mathbf{S}$ and $\mathbf{D}$ defined as in (2). For each matrix in $\mathcal{M}_d$ there exists only one Zonotope representation. The Zonotope is also unique up to splits of classes, insertion/elimination of empty classes and permutation of classes, as a consequence of the invariance properties of the Minkowski sum. The Zonotope is not invariant to merge operations.[15]

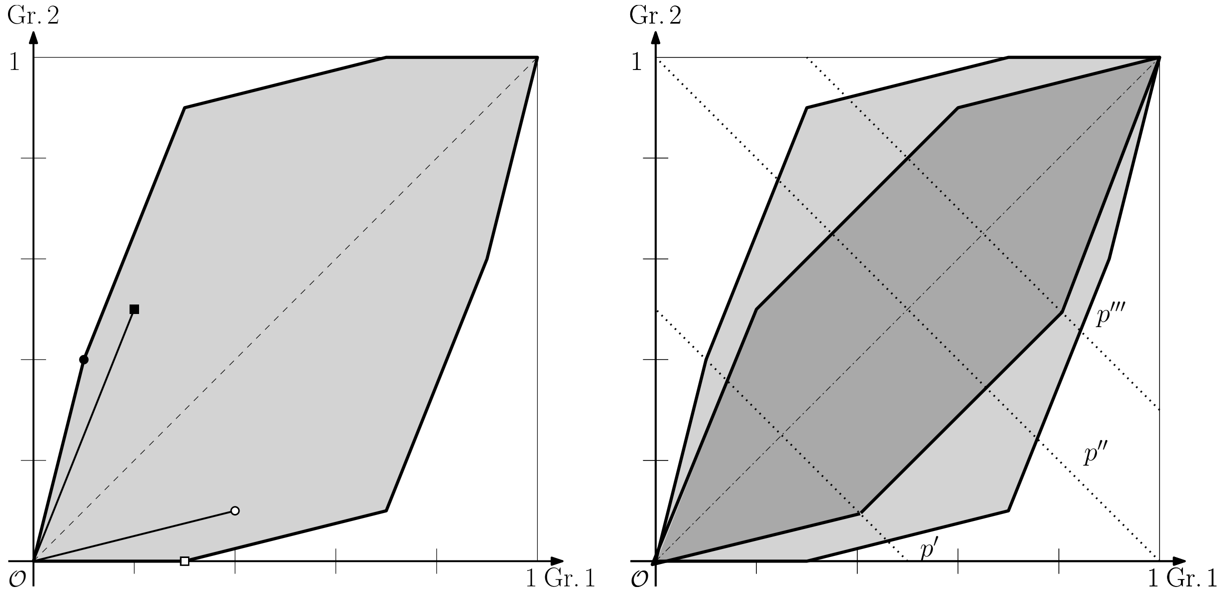
In Figure 1(a) we provide a graphical example of the 2-dimensions Zonotope of the distribution matrix $\mathbf{E} \in \mathcal{M}_2$:

$$\mathbf{E} = \begin{pmatrix} 0.4 & 0.1 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0 & 0.5 \end{pmatrix}. \tag{4}$$

The dimensionality of the example matrix helps visualizing the way $Z(\mathbf{E})$ is constructed. First, the vectors representing the classes of $\mathbf{E}$ are plotted in the unit square and connected to the origin with line segments. In the figure, these vectors are marked with different symbols. For instance, the vector marked with a black square represents the fourth class of $\mathbf{E}$ and has coordinates 0.2 (the proportion of groups 1) and 0.5 (the proportion of group 2). Then, the resulting segments are tied together in any possible arrangement. In the figure, adding together the vector corresponding to classes one and three gives the vector with coordinates $(0.7, 0.1)$, while adding this vector to the one representing the fourth class gives $(0.9, 0.6)$. The resulting Zonotope of $\mathbf{E}$ is the grey area in the figure that contains all possible arrangements of these segments, or portions of them.

The main result. We show that the inclusion condition $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ always indicates that $\mathbf{B}$ displays at most as much dissimilarity as $\mathbf{A}$. This completes the list of criteria to

[15] The Minkowski sum $\sum_{j=1}^{m} \theta_j \mathbf{a}_j$ for a set of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is invariant to operations involving adding or eliminating empty vectors or permuting vectors labels. Every split of vectors can be obtained by setting $\theta_j \in (0, 1)$, so that the operation does not impose constraints on the parameters of the Minkowski sum. Conversely, any merge operation of vectors $\mathbf{a}_j$ and $\mathbf{a}_k$ implies a constraint $\theta_j = \theta_k$, which reduces the extent of the set of points in the $[0, 1]^d$ space that can be characterized as the Minkowski sum of $\mathbf{A}$'s columns.



(a) Zonotope of **E** in (4) (b) Zonotopes inclusion

Figure 1: Zonotopes of matrices **E** (light grey area) and **E**′ (dark grey area).

assess dissimilarity with a testable dissimilarity criterion. Our main result states that all these criteria are equivalent.

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 insertions/elimination 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 equivalences between statements *(i)*, *(ii)* and *(iii)* are novel. They provide the basis for characterizing the class of measures D_h that are consistent with the dissimilarity criterion. Claim *(iii)* clarifies that dissimilarity comparisons can be modeled as an average, taken across classes of a distribution matrix, of judgements (expressed by the function h)

about the dispersion in the composition of the groups proportions in each of the classes (expressed by the ratio $a_{ij}/\overline{a}_j$).[16]

The equivalence between claims *(iii)* and *(iv)* formalizes, in the context of dissimilarity analysis, the criterion of informativeness in statistical experiments first documented in Blackwell (1953). The interesting equivalence involves statements *(i)* and *(iv)* and provides a characterization of matrix majorization alternative to Grant et al. (1998) and Frankel and Volij (2011). It also shows that every informativeness comparison of matrices in $\mathcal{M}_d$ verifies the existence of dissimilarity preserving and reducing transformations mapping the most informative distribution matrix into the least informative one.

The most relevant and new equivalence in Theorem 1 involves statements *(iv)* and *(v)*. It establishes that Zonotope inclusion (a testable condition) is sufficient and necessary to implement dissimilarity comparisons according to the different perspectives highlighted in claims *(i)*, *(ii)*, *(iii)* and *(iv)* (all not testable). This result extends to the multi-group setting what has been already shown in Dahl (1999) for the case $d = 2$ in the context of the analysis of informativeness in experiments. Furthermore, the equivalence of *(i)* and *(v)* demonstrates that Zonotopes inclusion always grants the existence of a sequence of dissimilarity preserving and reducing transformations that, when applied to matrix $\mathbf{A}$, allow to obtain matrix $\mathbf{B}$ in a finite number of steps, although the sequence itself cannot be identified.

Claim *(v)* offers a novel geometric interpretation of dissimilarity. Note that every element $\mathbf{z} \in Z(\mathbf{A})$ can be obtained by merging together proportions of the classes of $\mathbf{A}$ (i.e., it results from the Minkowski sum of vectors). Define an *isopopulation line* (when $d = 2$) or *(hyper)plane* (when $d \geq 3$) as the set of all combinations of proportions of the groups collected in $\mathbf{z}$ that add up to $p \in [0, 1]$, such that $\frac{1}{d}\mathbf{1}_d^t \cdot \mathbf{z} = p$. In this case, p is the average proportion across groups depicting the "size" of $\mathbf{z}$, obtained by weighting equally

[16] Bohnenblust, Shapley and Sherman (1949) have shown that a condition similar to *(iii)* can be interpreted in terms of loss functions and is tightly connected to comparisons of information distribution in games.

all groups. Claim *(v)* is verified if the set of all proportions of the groups adding up to p in $\mathbf{B}$ is included in (i.e., is less dispersed than) the corresponding set of all proportions of the groups adding up to p in $\mathbf{A}$. The criterion is robust, given that the inclusion should be verified *for all* p's.

An example with two groups clarifies this point. In Figure 1(b) we depict the Zonotope of $\mathbf{E}$ in (4) along with the Zonotope of the distribution matrix $\mathbf{E}'$, obtained from $\mathbf{E}$ by merging classes two and three as follows:

$$\mathbf{E}' = \begin{pmatrix} 0.4 & 0 & 0.4 & 0.2 \\ 0.1 & 0 & 0.4 & 0.5 \end{pmatrix}. \tag{5}$$

As expected, we observe that $Z(\mathbf{E}') \subseteq Z(\mathbf{E})$. The three dashed line segments crossing the Zonotopes in Figure 1(b) correspond to the isopopulation lines at proportions p', p'' and p''' in $\mathbf{E}$ and in $\mathbf{E}'$. Each of these isopopulation lines identifies cross-sections of the two Zonotopes, which are displayed as line segments delimited by the Zonotopes boundaries. The size of these segments reflects the degree of dissimilarity in the distribution of the proportions of the groups adding up to p', p'' and p'''. Consistently with the main theorem, all points lying on these segments can be obtained by different arrangements of splits, merges and permutations of the classes of $\mathbf{E}$ and $\mathbf{E}'$, respectively. The inclusion $Z(\mathbf{E}') \subseteq Z(\mathbf{E})$ guarantees that the dispersion in the proportions of the groups is smaller in $\mathbf{E}'$ than in $\mathbf{E}$ for *every* proportion $p \in [0, 1]$.

3.3 Remarks

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 $\mathbf{A} = \mathbf{B} \cdot \mathbf{X}'$ (see Theorem 1, statement *(iv)*). As a consequence, dissimilarity indifference holds whenever $Z(\mathbf{B}) = Z(\mathbf{A})$.

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 $d' < d$ 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.[17] The 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 reverse implication is not true. The Introduction offers a counterexample, showing that the ordering of distribution matrices induced by pairwise groups comparisons is not robust to the mixing of these groups. 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 $\widehat{\mathbf{X}} \cdot \mathbf{A}$. It follows that when $d' = 2$, a widely explored case in segregation analysis, then 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 transformations. Nonetheless, the Zonotopes inclusion criterion satisfied a "weak" form of *subgroup consistency* (Foster and Shorrocks 1988), insofar if any pairwise comparisons of distributions agrees that the two groups in $\mathbf{A}$ display at least as much dissimilarity as the same groups in $\mathbf{B}$, then one cannot conclude that $\mathbf{B}$ is more dissimilar than $\mathbf{A}$, albeit there is no guarantee that the two matrices can be eventually ranked.

Finally, the axiom *IPG* extends the dissimilarity indifference set to all comparisons involving a relabeling of the groups. The next remak formalizes the implications of adding *IPG* to the axioms considered in Theorem 1, and provides a natural multi-group extension of Hutchens (2015) results, postulating symmetry of types.

[17] For instance, the matrices in $\widehat{\mathcal{R}}_{2,3}$ generate comparisons involving two groups starting from a population partitioned into three groups. These comparisons involve either all three pairs of groups, or each single group against a mixture of the remaining two groups. This case might be of interest for disentangling the contribution of each dimension partitioning the population into groups on overall dissimilarity.

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 4.

4 Related orders

4.1 Inequality

Consider two n-variate vectors $\mathbf{a}^t, \mathbf{b}^t \in \mathcal{M}_1$, with $\mathbf{a}^t \cdot \mathbf{1}_n = \mathbf{b}^t \cdot \mathbf{1}_n = 1$. These vectors may represent, for instance, distributions of income shares 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 donors to poor recipients, without switching their relative positions in the income ranking (Hardy et al. 1934, Marshall et al. 2011).[18] If this is the case, then the Lorenz curve of $\mathbf{b}^t$ lies nowhere below the Lorenz curve of $\mathbf{a}^t$.

Recall that the Lorenz curve is a joint plot of the cumulative income shares, arranged by increasing income magnitude, and the cumulative weight of the recipients of these income shares observed in the data. Visually (see also Koshevoy and Mosler 1996), the curve coincides with the lower bound of a Zonotope and the associated Lorenz dominance is equivalent to the Zonotope inclusion criterion in statement (v) of Theorem 1. It follows that *every (income) inequality comparison involves the assessment of the dissimilarity between the distributions of income shares and of the weights of the income recipients.* The validity of the claim extends as well to multidimensional inequality comparisons, where matrices $\mathbf{A}$

[18] A *PD* transfer applied to $\mathbf{a}^t$ predicts that, for $a_j > a_k$, inequality is attenuated by operations involving a reduction of a_j by a quantity $\epsilon > 0$ and an equal increase of a_k by the same magnitude such that $a_j - \epsilon \geq a_k + \epsilon$, therefore preserving $\mathbf{a}^t \cdot \mathbf{1}_n$.

and $\mathbf{B}$ with $d \geq 2$ represent, for instance, distributions of d-variate bundles of goods shares across n equally weighted individuals or households. Thus, the "less spread out" relation can be formalized as a dissimilarity relation between one (or many) distribution(s) and a reference distribution.

The role of the dissimilarity model for inequality analysis is formalized in the following corollary. It establishes equivalent conditions regarding the dissimilarity between the distributions of relevant outcomes and the distribution of individual weights, here assumed uniform and equal to $1/n$.

Corollary 1 *Let* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$. *For every dissimilarity ordering* $\preccurlyeq$ *satisfying axioms IEC, IPC, SC and MC it holds that*

$$\mathbf{B}' := \begin{pmatrix} \frac{1}{n_B}\mathbf{1}^t_{n_B} \\ \\ \mathbf{B} \end{pmatrix} \quad \preccurlyeq \quad \mathbf{A}' := \begin{pmatrix} \frac{1}{n_A}\mathbf{1}^t_{n_A} \\ \\ \mathbf{A} \end{pmatrix} \tag{6}$$

if and only if there exists $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$ *such that (i)* $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ *and (ii)* $\frac{n_A}{n_B}\mathbf{1}^t_{n_B} = \mathbf{1}^t_{n_A} \cdot \mathbf{X}$.

Corollary 1 follows from the equivalence of claims (ii) and (iv) in Theorem 1 (a more formal proof is in the Appendix). When $n_A = n_B = n$, matrix $\mathbf{X}$ in the corollary should be *doubly stochastic*.[19] The condition $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ with $\mathbf{X} \in \mathcal{D}_n$ implied by (6), often referred to as *uniform majorization*, is widely adopted in robust univariate and multivariate inequality analysis (see p. 613 in Marshall et al. 2011). It states that one income distribution should be compared to another if the former looks more similar to a uniform distributions (where every unit receives the same income share) compared to the latter. All social welfare functions that are increasing in incomes and Schur-concave (i.e. display some degree of inequality aversion) would rank the two income distributions accordingly.

In the univariate case ($d = 1$), $\mathbf{B}' \preccurlyeq \mathbf{A}'$ in (6) indicates that the *PD* transfers applied to

[19] In fact, if $n_A = n_B$, then matrix $\mathbf{X}$ that is row stochastic should also be column stochastic because of condition (ii) in Corollary 1.

$\mathbf{a}^t$ to obtain $\mathbf{b}^t$ involve precise sequences of split and merge transformations of the classes of $\mathbf{A}'$. Hence:

Corollary 2 *Every PD transformation can be decomposed into a sequence of split of classes and merge of classes operations.*

Split and merge operations can hence be seen as inequality reducing transformations that are more elementary than *PD* transformations. In the Appendix, we provide an algorithm that shows how any T-transform, an equivalent matrix representation of a *PD* transfer (see p. 33 in Marshall et al. 2011), can be exactly decomposed into the product of matrices representing split and merge operations. It follows that any univariate inequality comparison based on uniform majorization can be seen as a dissimilarity comparison but not the reverse, insofar the dissimilarity preserving operations of split and merge characterize matrix majorization of which uniform majorization is a particular case.

The interesting result is that there always exists a sequence of splits and merges that supports uniform majorization even in the multidimensional case ($d \geq 2$), while this is not the case for *PD* transfers (Kolm 1977).

4.2 Lorenz Zonotopes and inequality analysis

Alternative criteria for assessing multivariate inequality are also embedded within the dissimilarity model. Koshevoy and Mosler (1996) have analyzed the distributions of consumption goods shares across individuals. Their proposal is to rank distribution matrices through the *Lorenz Zonotopes* inclusion order. A Lorenz Zonotope $LZ(.)$ in $\mathbb{R}^{d+1}_+$ is the plot, for each population share, of the associated set of possible bundles of goods that this share of population may achieve. When $d = 1$, the order induced by LZ is consistent with the Lorenz curve order. Following the notation in Corollary 1, $LZ(\mathbf{A}) := Z(\mathbf{A}')$. Theorem 1 allows to conclude that the Lorenz Zonotope relation $LZ(\mathbf{B}) \subseteq LZ(\mathbf{A})$ always indicates that there exists a sequence of merge, split, permutation and insertion/elimination of empty classes transforming $\mathbf{B}'$ into $\mathbf{A}'$. This result characterizes the Lorenz Zonotope inclusion order in terms of elementary transformations for any $d \geq 1$.

4.3 Welfare

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 distribution of population weights. A direct application of Theorem 1, consistent with results in Ebert and Moyes (2003), is that every welfare-consistent measure of inequality can be written as an average of convex transformations of equivalized incomes, scaled by their demographic weights. This is formalized by the inequality index $D_h = \sum_{j=1}^{n} \omega_j h(a_j/\omega_j)$ with h convex, where ω_j is individual j's weight and a_j/ω_j is her equivalent income.[20]

4.4 Inequality of opportunity

An increasingly popular notion of inequality, alternative to inequality of outcomes, is that of *inequality of opportunity* (Roemer 2012, Andreoli, Havnes and Lefranc 2019). According to this theory, outcomes are generated by individual effort (gathering all dimensions upon which people have full control and responsibility), by circumstances (such as the background of origin), and by the interaction of these two. Inequality of opportunity criteria account for the implications of the unequal distribution of circumstances on the distribution of some relevant outcome. In the context of income opportunities, some authors (for a review, see Ramos and Van de gaer 2016) have suggested to use as benchmark the counterfactual fair income distribution (representing the income distribution that would have occurred if the implications of the circumstances on income were eliminated). Inequality of opportunity

[20]The result, based on Lemma 1 in the Appendix, follows from the homogeneity and convexity of $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, that give $g(a_j, \omega_j) = \omega_j g(a_j/\omega_j, 1) = \omega_j h(a_j/\omega_j)$ with h convex.

stems from the dissimilarity in the distribution of the actual income shares across the entire population and the distribution of the counterfactual (fair) income shares in the same population. Theorem 1 hints on the possibility of using Zonotopes inclusion to test inequality of opportunity for income, and provides a consistent measurement framework.

4.5 Segregation

Segregation arises when individuals with different characteristics (such as their race or gender) are distributed unevenly across the neighborhoods of a city, the schools of a school district, or the jobs within a firm. In segregation analysis, the realizations of interest are categorical by nature. Mainstream approaches to segregation focus on the two-groups case and are consistent with the segregation curve order (Duncan and Duncan 1955).

The segregation curve is obtained by ordering the classes of $\mathbf{A}$ by increasing magnitude of the ratio a_{2j}/a_{1j} evaluated for each class j. It gives the proportions of group 1 and of group 2 that are observed in the classes where group 2 is relatively over-represented. The graph of the segregation curve coincides with the lower boundary of the Zonotope representing the data about groups 1 and group 2 distributions across categories. The segregation curve of matrix $\mathbf{E}$ in (4), for instance, is the lower boundary of the Zonotope represented in Figure 1(a). It is obtained by placing first class three, which has the lowest concentration ratio (equal to 0), while class two is placed last, as it displays the largest concentration ratio (equal to 4).

Theorem 1 bears three contributions to the field. First, the theorem clarifies that the operations of merge, split, permutation and insertion/elimination of empty classes characterize the ranking produced by non-intersecting segregation curves (Hutchens 1991). Second, the theorem establishes that the segregation curve dominance criterion can be naturally extended to the multi-group setting by looking at the d-variate Zonotopes inclusion order, which generalizes connections between the segregation curve and informativeness discussed in the bivariate setting (Dahl 1999, Frankel and Volij 2011, Lasso de la Vega and Volij 2014). Furthermore, the two tests are equivalent in the case of two-groups distributions: 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 $\mathbf{A}$. Third, Theorem 1 identifies and characterizes the class of multi-group segregation indices that are coherent with the family D_h. Below are some examples of well-known segregation indices belonging to this class.

The Duncan and Duncan's dissimilarity index for a matrix $\mathbf{A} \in \mathcal{M}_2$ 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. By setting

$$h(a_{1j}/\overline{a}_j, a_{2j}/\overline{a}_j) := |a_{1j}/\overline{a}_j - a_{2j}/\overline{a}_j|$$

it follows that $D_h(\mathbf{A}) = D(\mathbf{A})$.

In the multi-group context ($\mathbf{A} \in \mathcal{M}_d$), segregation can be measured by the Atkinson-type segregation index, defined $A_\omega(\mathbf{A}) := 1 - \sum_{j=1}^{n_A} \prod_{i=1}^{d} (a_{ij})^{\omega_i}$ for $\omega_i \geq 0$ such that $\sum_{i=1}^{d} \omega_i = 1$. By setting

$$h\left(a_{1j}/\overline{a}_j, \ldots, a_{dj}/\overline{a}_j\right) := 1 - d \prod_{i=1}^{d} \left(a_{ij}/\overline{a}_j\right)^{\omega_i}.$$

it follows that $D_h(\mathbf{A}) = A_\omega(\mathbf{A})$. Convexity of h stems from the features of the weighting scheme.

The mutual information index characterized in Frankel and Volij (2011) 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 $a_{ij} = 0$. By setting

$$h\left(a_{1j}/\overline{a}_j, \ldots, a_{dj}/\overline{a}_j\right) := \sum_{i=1}^{d} \cdot \log_2(d) - \frac{a_{ij}}{\overline{a}_j} \cdot \log_2\left(\frac{\overline{a}_j}{a_{ij}}\right)$$

it follows that $D_h(\mathbf{A}) = M(\mathbf{A})$. Convexity of h stems from the log operator.

5 Concluding remarks

A large and sparse literature on segregation and inequality measurement has proposed criteria for ranking multi-groups distributions according to the dissimilarity they exhibit. This paper establishes the foundation of the dissimilarity model, which provides an organized and integrated measurement framework for a variety of socio-economic phenomena. For empirical purposes, the interesting result is that the existence of dissimilarity preserving and/or reducing transformations mapping one configuration into another can be tested upon inclusion of the distribution matrices Zonotopes representation, a multidimensional generalizations of the segregation curve. This last aspect allows to implement dissimilarity analysis for policy evaluation purposes.

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

In some cases, Zonotopes inclusion is rejected by the data. The dissimilarity indices characterized in Theorem 1 allow to produce conclusive evaluations about the changes in dissimilarity, in a way consistent with the implications of the "elementary" transformations. Evaluations based on one or few dissimilarity indicators, however, are not robust and can always be challenged on the perspective offered by alternative measures. The complete characterization of the dissimilarity indicators presented here is left for future research.

A Proofs

A.1 Preliminary results

The first result shows that matrix majorization admits an equivalent representation in terms of unanimous ranking for 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*

$$\sum_{j=1}^{n_B} g(b_{1j}, \ldots, b_{dj}) \ \leq \ \sum_{j=1}^{n_A} g(a_{1j}, \ldots, a_{dj}), \tag{7}$$

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 merge 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 all matrices reproducing an insertion of empty classes when post-multiplied to a distribution matrix $\mathbf{A}$. Hence $\mathbf{Y} \in \mathcal{R}^{IEC}_{n_A,n_B}$ is an identity matrix of size n_A 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 the other 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 a distribution matrix $\mathbf{B} = \mathbf{A} \cdot \mathbf{M}_{[j]}$, where the matrix $\mathbf{M}_{[j]}$ performs a merge of class j towards $j+1$. It belongs to the set:

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

A.2 Proofs of Theorem 1

Proof. 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' with split parameters λ and $1-\lambda$ 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 distribution of groups' frequencies across classes, that is $\overline{\mathbf{b}}^t = \mathbf{1}_d^t \cdot \mathbf{B}$ and $\overline{\mathbf{a}}^t = \mathbf{1}_d^t \cdot \mathbf{A}$. Hence condition (7) in Lemma 1 can be written as $\sum_j g(\mathbf{b}_j^t, \overline{b}_j) \leq \sum_j g(\mathbf{a}_j^t, \overline{a}_j)$ with g defined on $\mathbb{R}^{d+1}$. Given that g is convex and homogeneous, then $g(\mathbf{a}_j^t, \overline{a}_j) = \overline{a}_j g(\mathbf{a}_j^t/\overline{a}_j, 1) = \overline{a}_j h(\mathbf{a}_j^t/\overline{a}_j)$ where $h \in \mathcal{H}$, 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 (7). We have therefore obtained the index D_h in (3). By Lemma 1, $D_h(\mathbf{B}) \leq D_h(\mathbf{A})\ \forall h \in \mathcal{H}$ is equivalent to (7) and implies *(iv)*.

(iv) $\Rightarrow$ *(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:

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

Denote $\widetilde{\boldsymbol{\theta}} = \left(\widetilde{\theta}_1, \ldots, \widetilde{\theta}_j, \ldots, \widetilde{\theta}_{n_A} \right)$ where $\widetilde{\theta}_j$ is defines as above, the Zonotope of matrix $\mathbf{B}$ writes:

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

where $\mathcal{I}$ is a subset of the n_A-fold cartesian product of $[0,1]$, following from the fact that if $x_{jk}(h) = 1$ then $x_{jk'}(h) = 0$ 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 $\mathcal{I} \subseteq [0,1]^{n_A}$ then 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 $\mathcal{I} = [0,1]^{n_A}$, $Z(\mathbf{B}) = Z(\mathbf{A})$, otherwise $Z(\mathbf{B})$ could be strictly included in $Z(\mathbf{A})$.

(v) $\Rightarrow$ *(i)*. Let consider $\mathbf{B}, \mathbf{A} \in \mathcal{M}_d$ and use the indices k and j to denote the columns of

B and **A** respectively. While assuming $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$, 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 **A** that gives **B**. Assuming statement (v) and knowing that $\mathbf{A} \cdot \mathbf{1}_{n_A} = \mathbf{B} \cdot \mathbf{1}_{n_B}$, the columns of **B** identify vectors that lie on $Z(\mathbf{A})$ and can hence be written (using the Zonotope definition) as follows:

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

for a generic class k' of **B**. In shorthand notation

$$
\mathbf{B} = \sum_{j=1}^{n_A} (\theta_j(1)\mathbf{a}_j, \dots, \theta_j(n_B)\mathbf{a}_j). \tag{8}
$$

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 (8) can be written as:

$$
\left(\lambda_{j1}\mathbf{a}_j, \lambda_{j2}(1 - \lambda_{j1})\mathbf{a}_j, \dots, \lambda_{j(n_B - 1)} \prod_{1 \leq k < n_B - 1} (1 - \lambda_{jk})\mathbf{a}_j, \prod_{1 \leq k \leq n_B - 1} (1 - \lambda_{jk})\mathbf{a}_j\right), \tag{9}
$$

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:

$$
\begin{aligned}
&\theta(1) = \lambda_1 \;\in\; [0, 1] \\
&\theta(k) = \lambda_k \left(1 - \textstyle\sum_{j=1}^{k-1} \theta(j)\right) \quad \text{with} \;\; \lambda_j \in [0, 1]\ \forall j = 2, \dots, n_B.
\end{aligned} \tag{10}
$$

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 (10) leads to (9) given that $\theta(k) = \lambda_k \cdot \prod_{j=1}^{k-1}(1-\lambda_j)$ with $\lambda_j \in [0,1]\ \forall j$ and $\lambda_k \in [0,1]\ \forall k = 2,\ldots,n_B$.

Consider a sequence of matrices $\mathbf{Z}_{[k]} \in \mathcal{R}_A^{SC}$.[21] 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 (9) can be equivalently written as:

$$\left(\lambda_{j1}\mathbf{a}_j, \ldots, \prod_{1\le k<n_B-1}(1-\lambda_{jk})\mathbf{a}_j, \mathbf{0}_d\right)\cdot \mathbf{Z}_{[n_B-1]} = (\mathbf{a}_j, \mathbf{0}_d, \ldots, \mathbf{0}_d)\cdot \prod_{1\le k\le n_B-1}\mathbf{Z}_{[k]}. \tag{11}$$

Extending the representation in (11) to all addends in (8) 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 (8) reveals that the order of the classes of $\mathbf{A}$ is irrelevant. Thus operations of *permutations of classes* are admitted.[22] 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:

$$\begin{aligned}
\widehat{\mathbf{X}} &:= \mathbf{\Pi}_{n_A}\cdot\mathbf{Y}\cdot \mathrm{diag}\left(\prod_{k=1}^{n_B-1}\mathbf{Z}_{[k]}(1), \ldots, \prod_{k=1}^{n_B-1}\mathbf{Z}_{[k]}(n_A)\right) && (12)\\
&= \mathbf{\Pi}_{n_A}\cdot\mathbf{Y}\cdot\prod_{j=1}^{n_A}\left(\prod_{k=1}^{n_B-1}\widetilde{\mathbf{Z}}_{[k]}(j)\right), && (13)
\end{aligned}$$

where $\mathbf{Z}_{[k]}(j)$ is indexed for j to highlight the relation with the class j in $\mathbf{A}$. Here $\widetilde{\mathbf{Z}}_{[k]}(j) := \mathrm{diag}\left(\mathbf{I},\ \mathbf{Z}_{[k]}(j), \mathbf{I}'\right)$ and $\mathbf{I}$ and $\mathbf{I}'$ are two identity matrices of size $(j-1)n_B$ and $(n_A -$

[21]See the preliminary results in Appendix A.1.

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

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

$j)n_B$ respectively. Line (13) comes from the fact that every block diagonal matrix can be represented as the product of the matrices associated with 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 $(\lambda_{1j}\mathbf{a}_1, \ldots, \lambda_{n_A j}\mathbf{a}_{n_A}, \ldots)$.[23] 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:

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

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.*

A.3 Proofs of Corollary 1

Proof. By Theorem 1, (6) 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

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

each single entry by n_A/n_B, thus *(ii)* should also hold. *Q.E.D.*

A.4 Proofs of Corollary 2

Proof. A formal proof draws on the fact that any (inequality reducing) *PD* transfer of an income share λ among classes j and k can be formalized through a linear transformation of vector $\mathbf{a}^t$ towards $\mathbf{b}^t$ involving a *T-transform* matrix $\mathbf{T}(\lambda, k, j)$, such that $\mathbf{b}^t = \mathbf{a}^t \cdot \mathbf{T}(\lambda, k, j)$, with $\mathbf{T}(\lambda, k, j) := \lambda \mathbf{I}_n + (1 - \lambda)\mathbf{\Pi}_{j,k}$, where $\mathbf{I}_n$ is the identity matrix, $\lambda \in [0, 0.5]$ and $\mathbf{\Pi}_{j,k} \in \mathcal{P}_n$ is a permutation matrix obtained from $\mathbf{I}_n$ by permuting columns j and k. Given a matrix $\mathbf{A} \in \mathcal{M}_d$ with n columns, let $\mathbf{S}(\lambda, k, j) \in \mathcal{R}_{n_A, n_B}$ be a row stochastic matrix that splits column k of $\mathbf{A}$ and merges a share $(1 - \lambda)$ of k with column j. This row stochastic matrix writes:

$$\mathbf{S}(\lambda, k, j) := [\lambda (\mathbf{I}_n, \mathbf{0}_n) + (1 - \lambda) (\mathbf{I}_n, \mathbf{0}_n) \mathbf{\Pi}_{n+1,k}] \cdot \begin{pmatrix} \mathbf{I}_n \\ \mathbf{i}_{j,\cdot} \end{pmatrix},$$

where $\mathbf{i}_{j,\cdot}$ is a row vector corresponding to row j of $\mathbf{I}_n$, $\lambda \in [0, 0.5]$ and $\mathbf{\Pi}_{n+1,k} \in \mathcal{P}_{n+1}$ is $n + 1$ dimensional permutation matrix obtained from $\mathbf{I}_{n+1}$ permuting columns $n + 1$ and k. Any T-transform involves a proportional movement of population masses from two classes, which amounts to repeating twice a sequence of splits and merges $\mathbf{S}(\lambda, k, j)$, so that $\mathbf{T}(\lambda, k, j) := \mathbf{S}(\lambda', k, j) \cdot \mathbf{S}(\lambda'', j, k)$, where the splitting parameters must satisfy $\lambda'' = 1 - \lambda$ and $\lambda' = \frac{1-2\lambda}{1-\lambda}$. This concludes the proof.

Q.E.D.

References

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

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

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

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

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

Bohnenblust, H. F., Shapley, L. S. and Sherman, S. (1949). Reconnaissance in game theory, *RAND Research Memorandum-208* .

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

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.

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. 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.

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

Foster, J. E. and Shorrocks, A. F. (1988). Poverty orderings, *Econometrica* **56**(1): pp. 173–177.

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

Gajdos, T. and Weymark, J. A. (2005). Multidimensional generalized gini indices, *Economic Theory* **26**(3): 471–496.

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

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.

Gravel, N. and Moyes, P. (2012). Ethically robust comparisons of bidimensional distributions with an ordinal attribute, *Journal of Economic Theory* **147**(4): 1384 – 1426. Inequality and Risk.

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. (2015). Symmetric measures of segregation, segregation curves, and blackwells criterion, *Mathematical Social Sciences* **73**: 63 – 68.

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

James, D. R. and Taeuber, K. E. (1985). Measures of segregation, *Sociological Methodology* **15**: 1–32.

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.

Lasso de la Vega, C. and Volij, O. (2014). Segregation, informativeness and lorenz dominance, *Social Choice and Welfare* **43**(3): 547–564.

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.

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

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.

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

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