



Robust dissimilarity comparisons with categorical outcomes

Francesco Andreoli^{1,2} · Claudio Zoli¹

Received: 30 April 2021 / Accepted: 24 June 2022 / Published online: 16 August 2022
© The Author(s) 2022

Abstract

The analysis of many phenomena requires partitioning societies into groups and studying the extent at which these groups are distributed with different intensities across relevant non-ordered categorical outcomes. When the groups are similarly distributed, their members have equal chances to achieve any of the attainable outcomes. Otherwise, a form of dissimilarity between groups distributions prevails. We characterize axiomatically the dissimilarity partial order of multi-group distributions defined over categorical outcomes. The main result provides an equivalent representation of this partial order by the ranking of multi-group distributions originating from the inclusion of their zonotope representations. The zonotope inclusion criterion refines (that is, is implied by) majorization conditions that are largely adopted in mainstream approaches to multi-group segregation or univariate and multivariate inequality analysis.

JEL Classification D63 · J71 · J62 · D30

We are indebted to the Editor and two Reviewers and to Valentino Dardanoni, Vincenzo Denicoló, Louis Eeckhoudt, Martin Foster, Alessio Fusco, Arnaud Lefranc, Brice Magdalou, Patrick Moyes, Eugenio Peluso, Gaston Yalonetzky for their useful comments. A preliminary version of this paper was presented at the ECINEQ meeting, LAGV conference, SSCW meeting, EEA-ESEM meeting, and seminars at Canazei Winter School, THEMA, University of Verona, University of Palermo, CORE, Bocconi University, Leicester University, Universitat Autònoma de Barcelona.

✉ Claudio Zoli
claudio.zoli@univr.it

Francesco Andreoli
francesco.andreoli@univr.it

¹ Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy

² Luxembourg Institute of Socio-Economic Research (LISER), 11 Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg

1 Introduction

The analysis of many phenomena requires partitioning societies into groups and studying the extent at which these groups are distributed with different intensities across relevant non-ordered categorical outcomes. For instance, residential segregation occurs in situations in which ethnic groups sort with different intensity across neighborhoods of a city. Likewise, school or occupational segregation is concerned with the uneven distribution of ethnic groups across schools or jobs.

In other cases, the interest is on the way shares of one or many attributes (such as income, wealth or consumption of different goods) are assigned across population units (such as countries, households or individuals) and the extent at which these distributions differ from the distribution of a normatively relevant benchmark, such as the demographic weights of the units. In this case, the focus is on uni- or multi-dimensional inequality and the often invoked anonymity principle would regard the way units are ordered as irrelevant.

All these examples are concerned with the extent of *dissimilarity* between two or more distributions defined over classes of non-ordered realizations. There is widespread agreement in the literature about what constitutes lack of segregation or equality: These are situations in which the groups are *similarly distributed* across the classes of realizations. The relevant notion of similarity we refer to dates back to Gini (1914), who argues that two (or more) groups are similarly distributed whenever “the populations of the two groups take the same values with the same frequency.”¹ Although a well-established methodology exists for analyzing dissimilarity between *two* distributions, the literature disagrees about the way such comparisons can be extended to the *multi-group* setting.

This paper develops the axiomatic foundations for the measurement of multi-group dissimilarity and provides equivalent testable conditions. We embrace a convenient (and equivalent) way of representing empirical discrete distributions through matrix notation. Each *distribution matrix* displays (by row) the distributions of individuals belonging to one of many (at least two) groups across classes of realizations (by column). The example below refers to distributions of three groups across a variable number of classes:

$$\mathbf{A} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0.8 & 0.2 \\ \frac{3}{7} & \frac{2}{7} & 0 & \frac{2}{7} \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0.4 & 0.6 \\ \frac{10}{14} & \frac{3}{14} & \frac{1}{14} \end{pmatrix}. \quad (1)$$

Entries of these matrices can be interpreted as frequencies so that, for instance, the share of group 2 in class 3 in \mathbf{A} is 80%. In the context of school segregation analysis, each of the two matrices above portrays the way students from each of three distinct ethnic groups are distributed across schools in a schooling district. Matrix

¹ Gini (1914, p. 189), translated from Italian, formalizes similarity as 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 , $\frac{n_x}{m_x} = \frac{n}{m}$.”

A depicts a district with four schools whereas matrix **B** depicts another district with three schools.

Many existing criteria can be employed to compare distribution matrices **A** and **B** by the extent of dissimilarity displayed by their rows, such as dissimilarity indices or majorization conditions. The ranking produced by one or few indices, however, can be challenged by the use of alternative, yet plausible, measures whereas majorization conditions are robust to this criticism but could be empirically untractable. In this paper, we consider all dissimilarity orderings that are consistent with some normatively relevant axioms and we focus on the intersection of all such orderings as a robust criterion for dissimilarity analysis. It is well known (see Donaldson and Weymark 1998) that such criterion leads to a *partial order* of distribution matrices which induces unanimity in the way matrices are ordered by all underlying dissimilarity orderings satisfying the desirable axioms. The axioms that we consider characterize the ordering **B** “displays at most as much dissimilarity as” **A** by the possibility of obtaining **B** from **A** through sequences of elementary operations that either preserve dissimilarity between the matrices’ rows (such as permuting the labels of groups and classes, adding and deleting classes which are empty, splitting proportionally classes) or reduce it (such as merging groups frequencies across two classes of the same distribution matrix) and by some consistency properties of the dissimilarity orderings (with respect to the possibility of producing convex mixtures of classes).

The axioms that we study allow to compare only matrices with the same number of groups but extend dissimilarity comparisons to matrices with a different number of classes. Such extension is relevant for empirical applications. Moreover, the possibility of considering comparisons between distribution matrices with a different number of classes will make explicit the normative content of this approach by highlighting the combination of operations that make possible to rank distribution matrices with the same number of classes.

The main result, in Theorem 1, establishes that the partial order of distribution matrices consistent with our axiomatic model is equivalent to the partial order of distribution matrices induced by the inclusion of their *zonotope* representations. A zonotope is a convex geometric set representation of the data, defined in the space of groups frequencies, which is originated by the Minkowski sum (i.e., element by element sum of fractions) of all column vectors of a distribution matrix.² In inequality analysis, zonotopes have been used to derive interesting multivariate extensions of the Lorenz curve (Koshevoy and Mosler 1996; Mosler 2012). Theorem 1 shows that the zonotope inclusion criterion is also relevant for dissimilarity analysis, for at least three reasons.

First, we demonstrate that the zonotope inclusion criterion is consistent with the implications of operations that unambiguously preserve or reduce dissimilarity. Each

² See McMullen (1971) and Ziegler (1995) for a formal analysis of the zonotopes geometric properties. Dahl (1999) examines the zonotope inclusion criterion in the context of information analysis, whereas a related zonotope inclusion criterion (based on Lorenz zonotopes) provides a non-parametric test for robust comparisons of multivariate inequality (Koshevoy 1995; Koshevoy and Mosler 1996; Andreoli and Zoli 2020).

of these operations is found to have clear and intuitive consequences on the shape of the zonotope and hence provides a normative justification for using the zonotope inclusion criterion. Second, the zonotope inclusion criterion is a refinement of relevant majorization conditions and it is hence implied by them. We demonstrate that *uniform* and *matrix majorization* criteria, which are widely adopted criteria in robust multivariate distributional analysis Marshall et al. (2011), are related to the dissimilarity measurement model developed here.³ In the multi-group setting, some matrices that cannot be ranked by matrix majorization can still be robustly ranked by zonotope inclusion, whereas the two criteria coincide only in the two-groups setting. As a counterexample, we use matrices in (1) to show in Appendix A.15 that the zonotope of matrix \mathbf{B} is included in the zonotope of matrix \mathbf{A} , whereas \mathbf{B} is not majorized by \mathbf{A} . Third, the zonotope inclusion criterion can be empirically tested, whereas this is seldom the case for majorization conditions.

Theorem 1 and the corollaries implied by it contribute to the literature along the following lines. First, the results identify the differences between the zonotope inclusion criterion and majorization conditions. In particular, \mathbf{B} being matrix majorized by \mathbf{A} postulates the existence of a sequence of dissimilarity reducing or preserving transformations mapping the classes of \mathbf{A} into those of \mathbf{B} . The zonotope inclusion condition focuses instead separately on each class of \mathbf{B} and requires that any such class could be obtained through dissimilarity reducing or preserving transformations of the classes of \mathbf{A} . There is no guarantee that such operations can be organized into a sequence, making the zonotope inclusion criterion a refinement of matrix majorization (as observed in Dahl 1999).

Second, Theorem 1 provides an axiomatic justification for using zonotope inclusion as a multi-group criterion that is more robust than (i.e., implies) alternative refinements of matrix majorization based on sequential comparisons of two-groups distributions or on specific dissimilarity indices. Although any two-groups projection of a multi-group zonotope originates a zonotope itself, assessing inclusion among all two-groups projections is not sufficient to conclude about inclusion of the multi-group zonotopes. This is because the zonotope inclusion criterion fails to satisfy a *consistency* property of partial orders, requiring that if two distribution matrices differ only in terms of two groups distributions (i.e., two rows), then the two matrix could be ranked by focusing on dissimilarity in the sub-matrices associated with these groups Moulin (2016). Failing this property is desirable in a multi-group context, because the dissimilarity ranking of matrices that differ by two or few groups distributions should not only depend on dissimilarity between these distributions, but also on the extent at which these distributions are dissimilar from the rest. This feature makes the dissimilarity criterion robust against potential aggregation biases (originating, for instance, the Simpson's paradox, see Blyth 1972).

³ Relevant application of uniform and matrix majorization criteria are in linear algebra (Dahl 1999; Hasani and Radjabalipour 2007), in the comparison of statistical experiments (Blackwell 1953; Torgersen 1992), in information theory Grant et al. (1998), in the study of bivariate dependence orderings for categorical variables Giovagnoli et al. (2009) as well as in the analysis of inequality (see, for instance, Chapter 14 in Marshall et al. 2011; Tsui 1995; Gajdos and Weymark 2005; Weymark 2006) and segregation (see Frankel and Volij 2011).

Third, Theorem 1 rationalizes the normative underpinnings of a variety of sparse and apparently unrelated results on the measurement of (multi-group) segregation and multivariate and univariate inequality, which are shown to be embedded within the dissimilarity model.

The paper is organized as follows. Relevant notation and a definition of the zonotope inclusion ordering is provided in Section 2. Axioms and the main result are in Section 3. Section 4 describes the usefulness of our results for related orders. Section 5 concludes. All proofs are collected in a dedicated appendix.

2 Using zonotopes to test dissimilarity

2.1 Notation

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

$$\mathcal{M}_d := \left\{ \mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{n_A}) : \mathbf{a}_j \in [0, 1]^d \forall j, \sum_{j=1}^{n_A} a_{ij} = 1 \forall i, \text{ for } n_A \geq 2 \right\},$$

where a_{ij} is interpreted as the proportion of group i observed in class j and the column vector \mathbf{a}_j collects the proportions of all groups attaining realization j . Matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_3$ in (1) offer two examples. The distribution matrices in \mathcal{M}_d are *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 let $\bar{a}_j := \sum_{i=1}^d a_{ij}$ denote the "size" of class j , obtained by weighting uniformly the groups occupying it. For instance, the size of class 1 of matrix \mathbf{A} in (1) is $\bar{a}_1 = \frac{19}{28}$.

We follow the convention of using boldface letters to indicate column vectors, so that \mathbf{i}_j is a column vector corresponding to column j of an identity matrix \mathbf{I}_n of size $n \times n$, whereas $\mathbf{1}_n = (1, \dots, 1)'$ and $\mathbf{0}_n := (0, \dots, 0)'$ are the column vector with all n entries respectively equal to 1 or 0. The superscript always denotes transposition.

Every distribution matrix lies in-between two extreme cases. The first case is that of *perfect similarity*, occurring when the distributions of the groups coincide and can be represented by the same row vector $\mathbf{s}' \in \Delta^n$. This situation is depicted by matrix \mathbf{S} , whose d rows are all equal to \mathbf{s}' . A *maximal dissimilarity* matrix \mathbf{D} is a disjoint-row-support matrix where each class is occupied at most by one group, but one group may occupy different classes.⁴

If a distribution matrix displays the structure of \mathbf{S} , then it is not possible to forecast the group belonging from knowledge of the realization. Conversely, if a

⁴ The condition $d \leq n$ is necessary for \mathbf{D} to exist. If \mathbf{A} is such that $d > n$, then it can display some dissimilarity, but never maximal dissimilarity.

distribution matrix is as \mathbf{D} , then it is always possible to forecast the group from the knowledge of the class of the realizations. Any distribution matrix displays a structure which lies in-between that of \mathbf{S} and \mathbf{D} .

We also consider *transformation matrices* that, when multiplied by a distribution matrix, produce effects on the extent of dissimilarity between the rows of such matrix. Denote by $\mathcal{R}_{n,m}$ the set of $n \times m$ row stochastic matrices whose rows lie in Δ^m .⁵ Moreover, denote \mathcal{R}_n whenever $m = n$, while $\mathcal{D}_n \subseteq \mathcal{R}_n$ is the set of doubly stochastic matrices whose rows and columns lie in Δ^n . The set collecting all $n \times n$ permutation matrices is denoted by \mathcal{P}_n .

2.2 The zonotope set

Geometric representations of the data are useful to derive empirical tests for ranking uni- and multivariate distributions. Lorenz curves, for instance, are the workhorse of robust income inequality analysis. A Lorenz curve is obtained by arranging income observations in increasing order and then plotting the cumulative sum of these incomes shares against the cumulative sample frequencies. If the Lorenz curve of one income distribution lies above the Lorenz curve of another income distribution, then one can robustly rank the former distribution as less unequal than the latter.

Lorenz curves do not provide sufficient structure for ranking multivariate distributions, insofar each Lorenz curve allows to compare only one distribution at a time with a normatively relevant one, usually the distribution of population shares. Lorenz curves may be useful to compare distributions of two groups across relevant units, such as in school segregation analysis. In this case, Lorenz curves (known as segregation curves) portray the degree of dissimilarity between the distribution of a group’s members across schools and the distribution of another group across the same schools. Multi-group extensions are however problematic even in this domain.

In this section, we consider using the *zonotope* representation of the data to implement multi-group comparisons and we investigate the robust and testable ordering of distribution matrices generated by the zonotope inclusion criterion.

The zonotope set $Z(\mathbf{A}) \subseteq [0, 1]^d$ of a matrix $\mathbf{A} \in \mathcal{M}_d$ is a convex polytope lying on the hypercube $[0, 1]^d$ that is symmetric with respect to the point $\frac{1}{2}\mathbf{1}_d$ (see McMullen 1971). It is defined as follows:

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

Elements in $Z(\mathbf{A})$ are identified by the Minkowski sum of the vectors with coordinates given by \mathbf{A} ’s classes. In Fig. 1a we represent the 2-dimensions zonotope of the distribution matrix $\mathbf{E} \in \mathcal{M}_2$, defined as follows:

⁵ 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.

$$\mathbf{E} = \begin{pmatrix} 0.4 & 0.1 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0 & 0.5 \end{pmatrix}. \quad (2)$$

The dimensionality of the example \mathbf{E} helps visualizing the way $Z(\mathbf{E})$ is constructed. First, vectors with coordinates corresponding to 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 black square represents the fourth class of \mathbf{E} . 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. Panel b) of Fig. 1 represents instead $Z(\mathbf{A})$ taken from (1), which is defined on the three-dimensional space. We report with solid lines only the visible edges of $Z(\mathbf{A})$. The relevant facets of $Z(\mathbf{A})$ originated by the sequential sum of \mathbf{A} 's classes, are coloured in light gray.

The example above highlights two situations of interest. 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 the maximum dissimilarity zonotope and that include the similarity zonotope. The shape of $Z(\mathbf{D})$ and of $Z(\mathbf{S})$ does not depend on the data in matrices \mathbf{S} and \mathbf{D} , thus highlighting the irrelevance of within-group heterogeneity for dissimilarity evaluations. More broadly, for each matrix in \mathcal{M}_d there exists only one zonotope representation, although the same zonotope may correspond to many distribution matrices.

2.3 The zonotope inclusion criterion

Zonotopes can be used to compare matrices by the extent of dissimilarity they exhibit. In this paper, we study the ranking of distribution matrices such as \mathbf{A} and \mathbf{B} that is generated by the inclusion of the zonotope representations of the two matrices, that is $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$. Inclusion can be easily checked when $d = 2$ from inspection of the zonotopes graphs. Figure 2 shows an example where $Z(\tilde{\mathbf{E}}) \subseteq Z(\mathbf{E})$, for a distribution matrix $\tilde{\mathbf{E}}$ obtained after performing an element-to-element summation of classes 2 and 3 in \mathbf{E} , thereby leading to the central class of $\tilde{\mathbf{E}}$ that contains 40% of the population of both groups. This operation is obviously leveling disparities in the distributions of groups 1 and 2, although similarity is not achieved. The inclusion criterion is more difficult to visualize when $d = 3$. For instance, panel b) of Fig. 1 reports relevant facets of the zonotope $Z(\mathbf{A})$ obtained at fixed proportions of group 1 distribution (in light gray). The figure also shows the corresponding facets of the zonotope $Z(\mathbf{B})$ (in dark gray). In this specific example, it is sufficient to check inclusion of the facets of $Z(\mathbf{B})$ into the corresponding facets of $Z(\mathbf{A})$ to conclude about

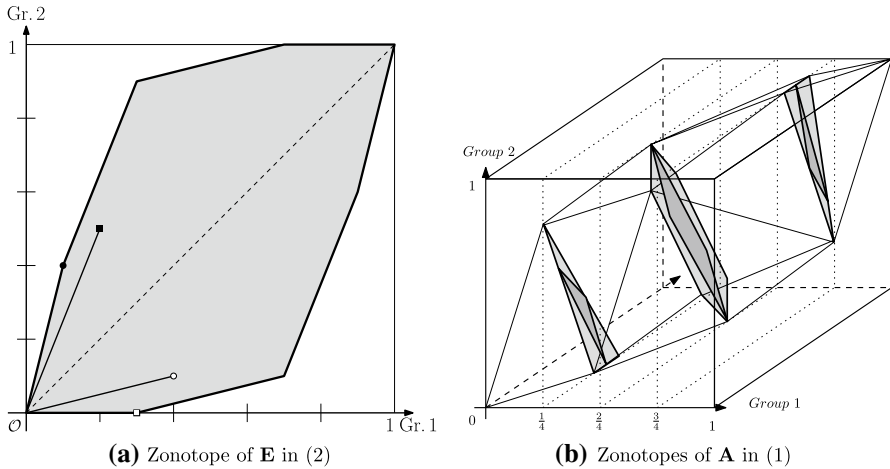


Fig. 1 Zonotopes in $d = 2$ and $d = 3$ dimensions

$Z(\mathbf{B}) \subseteq Z(\mathbf{A})$.⁶ In more general situations when $d > 3$, visualization of zonotope inclusion is not possible and the inclusion criterion should be tested algorithmically.

For general distribution matrices \mathbf{A} and \mathbf{B} , the zonotope inclusion criterion $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ has an intuitive interpretation in terms of disproportionality of groups frequencies, which may lead to conclude about the dissimilarity displayed by \mathbf{B} and \mathbf{A} . To see this, 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 vectors $\mathbf{z} \in Z(\mathbf{A})$ that add up to $p \in [0, 1]$, such that $\frac{1}{d}\mathbf{1}'_d \cdot \mathbf{z} = p$. In other words, p is the average “size” of \mathbf{z} , obtained by weighting equally all groups. Figure 2 depicts an example, based on distribution matrices \mathbf{E} and, $\tilde{\mathbf{E}}$ in which the dashed line segments p' , p'' and p''' correspond to isopopulation lines. In general, $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ 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, thus implying that any disproportional allocation of groups attainable by merging classes of \mathbf{B} can also be obtained by merging the classes of \mathbf{A} , but not the reverse. Proportionality is always attained in $Z(\mathbf{S})$, in which case there is *only one* attainable allocations $\mathbf{z} \in Z(\mathbf{S})$ lying on any isopopulation line p , that is $\mathbf{z} = \mathbf{1}_d p$. Conversely, disproportionality is maximal in $Z(\mathbf{D})$, in which case *every* attainable allocation lying on the isopopulation line p can be obtained from the original data.

The zonotope inclusion criterion always ranks $Z(\mathbf{S}) \subseteq Z(\mathbf{A}) \subseteq Z(\mathbf{D})$ for any $\mathbf{A} \in \mathcal{M}_d$ and for any d . The inclusion criterion, however, entails only a partial order of distribution matrices: two matrices cannot be ordered if their respective zonotope representations intersect. In the following section, we characterize the normative content of the inclusion criterion in terms of unanimous agreement in ranking \mathbf{B} as

⁶ See Appendix A.15 for a formal proof.

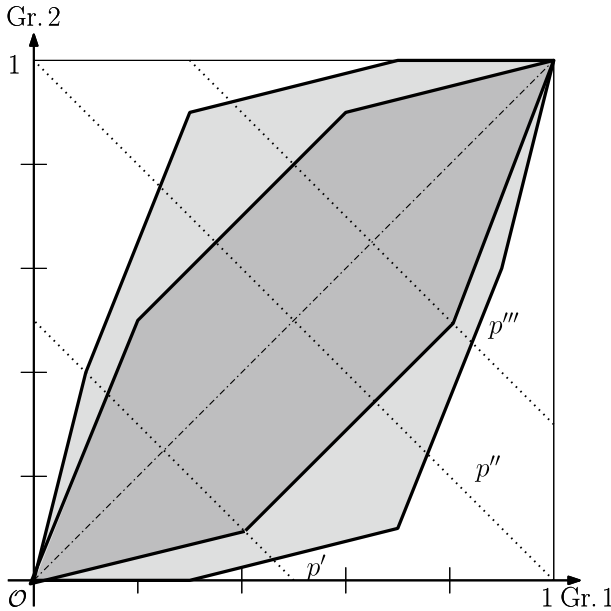


Fig. 2 Zonotope inclusion

displaying less dissimilarity than **A** by all orderings consistent with some basic dissimilarity axioms.

3 Characterizations

3.1 The dissimilarity partial order

We investigate the possibility of ordering distribution matrices according to the dissimilarity they display. A *dissimilarity ordering* is a complete and transitive binary relation \leq on the set \mathcal{M}_d with symmetric part \sim , that ranks $\mathbf{B} \leq \mathbf{A}$ whenever **B** is *at most as dissimilar as A*.⁷ Given $\mathbf{A} \in \mathcal{M}_d$, any dissimilarity ordering should always rank $\mathbf{S} \leq \mathbf{A} \leq \mathbf{D}$ for any matrix that can be represented as **S** and **D**. These matrices are respectively regarded to 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.⁸

⁷ For any $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathcal{M}_d$ the relation \leq is *transitive* if $\mathbf{C} \leq \mathbf{B}$ and $\mathbf{B} \leq \mathbf{A}$ then $\mathbf{C} \leq \mathbf{A}$ and *complete* if either $\mathbf{A} \leq \mathbf{B}$ or $\mathbf{B} \leq \mathbf{A}$ or both, in which case $\mathbf{B} \sim \mathbf{A}$.

⁸ Rao (1982) distinguishes the notion of dissimilarity from that of diversity. The former reflects differences or similarities between populations/groups heterogeneity, the latter reflects instead heterogeneity within the same population/group (see also Nehring and Puppe 2002). To see this, let s' be any row of a perfect similarity matrix **S**. Any two perfect similarity matrices **S** and $\tilde{\mathbf{S}}$ such that s' is a uniform distribu-

One direct implication of this feature of the dissimilarity orderings is that distribution matrices that differ in their number of classes ($n_A \neq n_B$) can be regarded as indifferent from the dissimilarity orderings perspective, as long as their group distributions differ across matrices, but the dissimilarity that they display within each matrix coincides across matrices.⁹ For this reason, we focus on matrices in \mathcal{M}_d , which must have the same number of rows but may differ in the number of classes.

Each dissimilarity ordering induces a complete ranking of distribution matrices. In this paper, we are interested in the robust ranking of distribution matrices generated by the intersection of the dissimilarity orders satisfying some desirable properties. This is a partial order Donaldson and Weymark (1998), that we characterize axiomatically and for which we provide equivalent representations.

3.2 Axioms and preliminary results

Axioms are based on elementary operations that, when applied to distribution matrices, can reduce or preserve dissimilarity among groups. We focus on the dissimilarity orderings that rank distribution matrices consistently with the effects of these operations. In order to ease the understanding of the axioms, we contextualize the consequences of such operations in terms of dissimilarity in the distribution of groups of students with different ethnic background across the schools in a schooling district. This is commonly referred to as a problem of schooling segregation.

The first axiom defines the context, introducing an anonymity property with respect to the labels (and hence the arrangement) of the classes of a distribution matrix.

Axiom 1 *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 of a matrix can be freely permuted without affecting the extent of dissimilarity it displays. In the context of schooling segregation, the axiom posits that the names of the schools are irrelevant to conclude about the dissimilarity in the distributions of students across these schools. This is arguably the case if the schools cannot or should not be ordered according to their performances, their quality or their budget. Another implication of axiom *IPC* is that any distribution matrix that is obtained by permuting the columns of matrix \mathbf{D} has to be regarded to as an equivalent representation of maximal dissimilarity.

Footnote 8 (continued)

tion across classes (high diversity) whereas \tilde{s} is a distribution concentrating the mass in few or one realization (low diversity) are regarded as equivalent by every dissimilarity ordering, i.e. $\mathbf{S} \sim \tilde{\mathbf{S}}$.

⁹ We provide an example along the lines of the previous footnote. There are many matrices of different size displaying the same structure as the perfect similarity matrix \mathbf{S} , for instance $\tilde{\mathbf{S}}$ such that \tilde{s}' is of size $\tilde{n} > n$, where n is the size of s' . Yet, for any of such matrices, $\mathbf{S} \sim \tilde{\mathbf{S}}$.

Next, we consider two transformations that extend comparability to distribution matrices that differ in the number of classes. The first transformation has to do with 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. The presence of these “empty” schools in the district is irrelevant for assessing dissimilarity of groups distributions across the remaining schools of the district.

Axiom 2 IEC (Independence from Empty Classes) For any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, if $\mathbf{B} = (\mathbf{A}, \mathbf{0}_d)$ then $\mathbf{B} \sim \mathbf{A}$.

The IEC axiom emphasizes 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 equivalent to that in \mathbf{B} . When combined with IPC, the axiom IEC allows to regard as indifferent all matrices obtained by adding or delating an empty class in any position.

The second transformation considered increases the number of classes by *splitting proportionally* (the groups frequencies 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. Frankel and Volij (2011) advocate a similar property (called composition invariance) in the study of multi-group school segregation (see also James and Taeuber 1985). The Independence from Split of Classes (ISC) axiom posits that the transformation described above is a source of indifference for every dissimilarity ordering.

Axiom 3 ISC (Independence from Split of Classes) For any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ with $n_B = n_A + 1$, if $\exists j$ such that $\mathbf{b}_j = \beta \mathbf{a}_j$ and $\mathbf{b}_{j+1} = (1 - \beta) \mathbf{a}_j$ with $\beta \in (0, 1)$, while $\mathbf{b}_k = \mathbf{a}_k \forall k < j$ and $\mathbf{b}_{k+1} = \mathbf{a}_k \forall k > j$, then $\mathbf{B} \sim \mathbf{A}$.

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 of classes is implemented by vector summation of two adjacent columns of a distribution matrix, irrespectively of the groups composition of each column. 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, without eliminating them. A merge of these two schools would, instead, establish proportionality in ethnic composition across all schools, leading to perfect similarity. The Dissimilarity Decreasing Merge of Classes (*MC*) axiom states that every merge of classes transformation cannot increase dissimilarity.

Axiom 4 *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 $\mathbf{B} \preceq \mathbf{A}$.

Consider obtaining \mathbf{B} from \mathbf{A} with a merge transformation adding together, distribution by distribution, the group proportions observed in two classes \mathbf{a}_j and \mathbf{a}_{j+1} whenever $\mathbf{a}_{j+1} = \beta \mathbf{a}_j$, $\beta > 0$, such that $\mathbf{b}_j = (1 + \beta)\mathbf{a}_j$ and $\mathbf{b}_k = \mathbf{a}_k \forall k < j$ while $\mathbf{b}_k = \mathbf{a}_{k+1} \forall k > j$. This operation leaves dissimilarity unaffected. The operation is opposite to a split, but supports the same indifference class, gathering all matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ such that $\mathbf{B} \sim \mathbf{A}$ for all orderings consistent with axiom *ISC*, even if not consistent with *MC*.

Axioms *MC*, *IEC*, *ISC* and *IPC* are independent. The transitive closure of all dissimilarity orderings satisfying these axioms defines a *partial order* of distribution matrices (see Donaldson and Weymark 1998), which is represented by the *matrix majorization* criterion. We refer to this partial order as $\mathbf{B} \preceq^R \mathbf{A}$, indicating that matrix \mathbf{B} is matrix majorized by \mathbf{A} whenever there exists $\mathbf{X} \in \mathcal{R}_{n_A, n_B}$ such that $\mathbf{B} = \mathbf{A}\mathbf{X}$ (Marshall et al. 2011; Dahl 1999).

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

- (i) $\mathbf{B} \preceq \mathbf{A}$ for all orderings \preceq satisfying axioms *MC*, *ISC*, *IEC* and *IPC*.
- (ii) $\mathbf{B} \preceq^R \mathbf{A}$.

The notion of matrix majorization has been investigated in a variety of contexts (see p. 625 in Marshall et al. (2011) and literature therein), the most relevant being that of comparisons of informativeness of statistical experiments with a finite number of outcomes.¹⁰ The characterization in Proposition 1, which is alternative to

¹⁰ Blackwell (1953) has formalized a precise condition for “ \mathbf{A} is at least as informative as \mathbf{B} ” which coincides with $\mathbf{B} \preceq^R \mathbf{A}$, insofar the observation of an experiment outcome (a class) in \mathbf{A} is more informative about the underlying signal (the group) than it is in \mathbf{B} . If $\mathbf{A} = \mathbf{D}$, then the experiment outcome identifies the signal and \mathbf{A} would be at least as much informative than any matrix \mathbf{B} , whereas if $\mathbf{B} = \mathbf{S}$, then it

Grant et al. (1998), Frankel and Volij (2011) and Lasso de la Vega and Volij (2014), 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.

While appealing, matrix majorization entails the possibility of ranking distribution matrices only if there exists a *sequence* of relevant transformations of the data that can be represented by a row stochastic matrix. This requirement is very stringent in some cases. Consider for instance matrices **A** and **B** in (1). It is shown in Appendix A.15 that every column of **B** can be obtained by splitting and merging the columns of **A**. Yet, these operations cannot be arranged to form a sequence in this specific example. As a consequence, these operations cannot be represented in the form of a row stochastic matrix, thereby yielding that **A** does not matrix majorize **B**. In fact, as shown in the appendix, there is a *unique* admissible transformation matrix with non-negative entries, denoted

$$\mathbf{X} = \begin{pmatrix} 1 & 0.5 & 0 \\ 1 & 0 & 0.25 \\ 0 & 0.5 & 0.75 \\ 0 & 0 & 0 \end{pmatrix}, \tag{3}$$

yielding $\mathbf{B} = \mathbf{AX}$. The transformation matrix **X** is, clearly, not row stochastic (for a related example, see Koshevoy 1995).

We address the concerns raised by the example above by introducing a new class of dissimilarity axioms. These axioms relax the possibility of ranking distribution matrices exclusively by mean of (sequences of) merge transformations, invoking instead a form of consistency of the dissimilarity orderings with respect to convex combinations of (columns of) distribution matrices. The first axiom, denoted *Strong-MixC*, states that if there are matrices $\mathbf{B}^1, \dots, \mathbf{B}^m$ that are ranked as not more dissimilar than **A**, then the convex mix of such matrices yields a matrix **B** that cannot display more dissimilarity than **A**. Before stating the axiom, recall that any element of the convex hull (denoted *conv*) of matrices $\mathbf{B}^j = (\mathbf{b}_1^j, \dots, \mathbf{b}_n^j) \in \mathcal{M}_d$ is a matrix $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n) \in \mathcal{M}_d$ such that $\mathbf{b}_k = \sum_{j=1}^m w_j \mathbf{b}_k^j \forall k$, for any set of weights $w_j \in [0, 1]$ satisfying $\sum_{j=1}^m w_j = 1$. The axiom name follows from the fact that every mix of matrices can be interpreted as a specific mix of classes that assigns uniform weights to the classes of the same matrix.

Axiom 5 *Strong-MixC (Dissimilarity Consistency with Uniform Mixing of Classes)*
 Consider a $d \times n$ matrix $\mathbf{A} \in \mathcal{M}_d$ and a sequence $j = 1, \dots, m, m \geq 2$ of $d \times n$ matrices $\mathbf{B}^j \in \mathcal{M}_d$ such that $\mathbf{B}^j \preceq \mathbf{A} \forall j$. If $\mathbf{B} \in \text{conv}\{\mathbf{B}^1, \dots, \mathbf{B}^m\}$, then $\mathbf{B} \preceq \mathbf{A}$.

Footnote 10 (continued)

is impossible to disentangle the underlying signal from observation of the experiment outcome and **B** would be less informative than any matrix **A**.

The axiom Strong-MixC postulates a “betweenness” property (see Dekel 1986) for dissimilarity orderings, meaning that all orderings satisfying it regard a new distribution \mathbf{B} whose classes are obtained as convex combinations of the respective classes of $\mathbf{B}^1, \dots, \mathbf{B}^m$ as not more dissimilar than \mathbf{A} . The fact that such convex combination assigns the same weight w_j to each class of a matrix \mathbf{B}^j , with $\sum_j w_j = 1$, guarantees that $\mathbf{B} \in \mathcal{M}_d$. Matrix \mathbf{B} may be regarded as more dissimilar than some of the matrices \mathbf{B}^j , but it cannot display more dissimilarity than \mathbf{A} given that $\mathbf{B}^j \preceq \mathbf{A} \forall j$.

The normative appeal of axiom Strong-MixC rests in its relation with operations of mixing of columns of different matrices. Such operations are regarded to as unambiguously dissimilarity not increasing, given their relation with merge of classes operations. We contextualize this point in terms of school segregation analysis. The axiom Strong-Mix implies that every school or merge of schools from schooling district \mathbf{B} (i.e., columns of the distribution matrix) can always be obtained through a convex combination of schools or merges of schools issued from the schooling districts $\mathbf{B}^1, \dots, \mathbf{B}^m$. Formally, let \mathcal{V}_n^{01} be the set of $n \times 1$ vectors whose elements are either 0 or 1. Non-zero entries of vector $\boldsymbol{\gamma} \in \mathcal{V}_n^{01}$ identify classes (or one class) of a distribution matrix, so that $\mathbf{B}\boldsymbol{\gamma}$ yields a new school that is obtained by merging some of \mathbf{B} 's schools (or by keeping one of its schools). Any mixing operation underlying axiom Strong-MixC always grants that:

$$\forall \boldsymbol{\gamma} \in \mathcal{V}_n^{01} \exists \boldsymbol{\gamma}^j \in \mathcal{V}_n^{01}, j = 1, \dots, m : \mathbf{B}\boldsymbol{\gamma} \in \text{conv}\{\mathbf{B}^1\boldsymbol{\gamma}^1, \dots, \mathbf{B}^m\boldsymbol{\gamma}^m\}, \quad (4)$$

where *conv* is the convex hull of such vectors.

Arguably, every merge of schools smooths the extent of ethnic disparities within a schooling district and can contribute to reduce segregation. Condition (4) implies that there are less opportunities to reduce segregation by merging some of the schools in schooling district \mathbf{B} compared to the extent of opportunities for reducing segregation that are available by merging schools in districts $\mathbf{B}^1, \dots, \mathbf{B}^m$. In fact, the former district/matrix can always be obtained as a combination of the latter matrices, whereas the reverse may not be true. As an example, consider the (rather extreme) case in which \mathbf{B} is the similarity matrix: in this case there are no opportunities at all to reduce segregation by merging \mathbf{B} 's classes, being segregation already at its minimum.

Condition (4) is always granted by axiom Strong-MixC, which imposes additional structure.¹¹ We consider a new axiom, denoted *MixC*, that regards every matrix $\mathbf{B} \in \mathcal{M}_d$ that satisfies condition (4) as displaying not more dissimilarity than \mathbf{A} .

Axiom 6 *MixC* (Dissimilarity Consistency with Mixing of Classes) Consider a $d \times n$ matrix $\mathbf{A} \in \mathcal{M}_d$ and a sequence $j = 1, \dots, m, m \geq 2$ of $d \times n$ matrices $\mathbf{B}^j \in \mathcal{M}_d$ such that $\mathbf{B}^j \preceq \mathbf{A} \forall j$. If $\mathbf{B} \in \mathcal{M}_d$ satisfies condition (4), then $\mathbf{B} \preceq \mathbf{A}$.

¹¹ If \mathbf{B} is obtained as in Strong-MixC, then $\mathbf{B}\boldsymbol{\gamma} = \sum_{j=1}^m w_j \mathbf{B}^j \boldsymbol{\gamma}, \forall \boldsymbol{\gamma} \in \mathcal{V}_n^{01}$, with $\sum_j w_j = 1$, which satisfies (4) because in this case $\boldsymbol{\gamma}^j = \boldsymbol{\gamma} \forall j$.

The axiom MixC values the fact of having less opportunities for reducing segregation in school district \mathbf{B} compared to districts $\mathbf{B}^1, \dots, \mathbf{B}^m$, in the sense that if the latter districts display less school segregation than \mathbf{A} , then \mathbf{B} should also display less segregation than \mathbf{A} . Axiom MixC extends the orderings $\mathbf{B}^j \preceq \mathbf{A}$ for every $j = 1, \dots, m$ to $\mathbf{B} \preceq \mathbf{A}$. The axiom may allow to compare cases such as those in the example related to the transformation matrix in (3): even if there is no sequence of merge of classes operations mapping \mathbf{A} into \mathbf{B} , it may be sufficient to conclude that every class of \mathbf{B} is obtained by merging classes of \mathbf{A} to verify condition (4) and thus rank $\mathbf{B} \preceq \mathbf{A}$, as we show in Appendix A.15.

The axiom MixC does not explicitly mention the way \mathbf{B} is constructed. One way to obtain it (which we rely upon in the proofs) is by considering weights $w_j^k \in [0, 1]$, with $\sum_{j=1}^m w_j^k = 1$, that are specific to each class k , such that $\mathbf{b}_k = \sum_{j=1}^m w_j^k \mathbf{b}_k^j \forall k$. These weights are more general than those considered by axiom Strong-MixC. When configuration \mathbf{B} is obtained in such a way and (4) is satisfied, the axiom MixC posits that mixing students across school of districts $\mathbf{B}^1, \dots, \mathbf{B}^m$ that are not more segregated than \mathbf{A} cannot generate a new schooling district \mathbf{B} that is more segregated than \mathbf{A} .

The fact that axiom Strong-MixC is always consistent with condition (4) proves the next statement.

Remark 1 If \preceq is consistent with MixC then it is consistent with Strong-MixC.

As a consequence, the MixC axiom implies a ranking of distribution matrices that is less partial compared to that characterized by axiom Strong-MixC, in the sense that all dissimilarity orderings consistent with Strong-MixC are capable of ranking unanimously only a subset of all matrices that can be ordered by the orderings consistent with MixC.

We now investigate if the partial orders of distribution matrices supported by matrix majorization and by the zonotope inclusion criterion are consistent with these new axioms. The next Remark shows that matrix majorization \preceq^R is consistent with Strong-MixC.

Remark 2 For $\mathbf{A}, \mathbf{B}^j \in \mathcal{M}_d, j = 1, \dots, m$, let $\mathbf{B} \in \text{conv}\{\mathbf{B}^1, \dots, \mathbf{B}^m\}$: $\mathbf{B}^j \preceq^R \mathbf{A} \forall j \Rightarrow \mathbf{B} \preceq^R \mathbf{A}$.

When axiom Strong-MixC is paired with dissimilarity preserving axioms, it allows to characterize matrix majorization without resorting on axiom MC, which is hence implied by all the axioms considered.

Remark 3 If \preceq satisfies IPC, ISC, IEC and Strong-MixC then it satisfies MC.

Axiom MC becomes redundant when characterizing dissimilarity partial orders if the Strong-MixC axiom is combined with all the dissimilarity preserving axioms. The axiom Strong-MixC yields a new characterization of \preceq^R which is alternative to the one presented in Proposition 1.

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

- (i) $\mathbf{B} \preceq \mathbf{A}$ for all orderings \preceq satisfying axioms Strong-MixC, ISC, IEC and IPC.
- (ii) $\mathbf{B} \preceq^R \mathbf{A}$.

Proposition 2 highlights that, even if axiom MC is implied by Strong-MixC, ISC, IEC and IPC (see Remark 3), these axioms still lead to the partial order of matrix majorization as in Proposition 1 and not to another partial order that is capable of ranking more matrices. Weakening the Strong-MixC axiom towards MixC may help characterizing such partial order. In the two-groups case ($d = 2$), Proposition 2 can be reformulated by weakening Strong-MixC in favor of axiom MixC, since matrix majorization \preceq^R is consistent with axiom MixC when $d = 2$.

Remark 4 For $\mathbf{A}, \mathbf{B}^j, \mathbf{B} \in \mathcal{M}_2$ such that \mathbf{B} and $\mathbf{B}^j, j = 1, \dots, m$, satisfy condition (4): $\mathbf{B}^j \preceq^R \mathbf{A} \forall j \Rightarrow \mathbf{B} \preceq^R \mathbf{A}$.

In the multi-group context ($d \geq 3$), however, matrix majorization is not consistent with axiom MixC. A counterexample is given in Appendix A.15, where we use the matrices in (1) to identify matrices $\mathbf{B}^j \in \mathcal{M}_d, j = 1, 2, 3$ such that $\mathbf{B}^j \preceq^R \mathbf{A}$ and we then show that \mathbf{B} satisfies condition (4) but cannot be obtained as in axiom Strong-MixC and hence $\mathbf{B} \not\preceq^R \mathbf{A}$. As a consequence, matrix majorization is only sufficient but not necessary to conclude about unanimity in the ranking by all dissimilarity orderings \preceq consistent with axioms MixC, IPC, IEC and ISC. Conversely, the zonotope inclusion criterion is always consistent with axiom MixC.

Remark 5 For $\mathbf{A}, \mathbf{B}^j, \mathbf{B} \in \mathcal{M}_d$ such that \mathbf{B} and $\mathbf{B}^j, j = 1, \dots, m$, satisfy condition (4): $Z(\mathbf{B}^j) \subseteq Z(\mathbf{A}) \forall j \Rightarrow Z(\mathbf{B}) \subseteq Z(\mathbf{A})$.

The zonotope inclusion criterion is consistent with MixC and, from Remark 1, it is also consistent with Strong-MixC. However, the Strong-MixC axiom is not necessary to characterize zonotope inclusion: the operations underlying Strong-MixC do not allow, alone, to break down the zonotope inclusion ordering of any pair \mathbf{A} and \mathbf{B} into the existence of simpler mixing transformation mapping one matrix into the other. The main result of the paper shows that the MixC axiom provides the needed structure to establish a characterization of the zonotope inclusion order. Dissimilarity orderings consistent with Strong-MixC but not with MixC can be represented by matrix majorization (Proposition 2), but this guarantees only a sufficient condition for zonotope inclusion.

3.3 Main result and discussion

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

- (i) $\mathbf{B} \preceq \mathbf{A}$ for all orderings \preceq satisfying axioms MixC, ISC, IEC, IPC.
- (ii) $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$.

The theorem provides a novel complete characterization of the zonotope inclusion criterion in terms of dissimilarity. The zonotope inclusion criterion originates a partial order of distribution matrices. If the inclusion test fails, then consensus on the ranking of distribution matrices by all dissimilarity orderings consistent with MixC, IPC, IEC and ISC cannot be reached. Nonetheless, this partial order is “less partial” than matrix majorization (that is, zonotope inclusion is a refinement of \leq^R) and is thus capable of ordering a larger set of cases compared to it.

Remark 6 Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$: $\mathbf{B} \leq^R \mathbf{A} \Rightarrow Z(\mathbf{B}) \subseteq Z(\mathbf{A})$.

The remark shows that there are matrices that cannot be ranked by dissimilarity orderings consistent with Strong-MixC or, equivalently and alternatively, with axiom MC but that can be ordered only by resorting to axiom MixC, which provides the additional structure that is needed to refine \leq^R . The rationale for this refinement is clarified in the proof of Theorem 1. There, we make use of the operations underlying axioms Strong-MixC, IEC, IPC and ISC to characterize the distribution matrices that form the *basis* of the set of all matrices that are matrix majorized by any given \mathbf{A} . We also show that some of the classes in each of the basis matrices identify vertices of $Z(\mathbf{A})$ (and that $Z(\mathbf{A})$ is the convex hull of its vertices). While the convex hull of the basis matrices obtained by using the weights implied by axiom Strong-MixC is sufficient to characterize the full set of matrices \mathbf{B} such that $\mathbf{B} \leq^R \mathbf{A}$, the same operation is not sufficiently flexible to characterize the entire set of matrices \mathbf{B} such that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$. Instead, when condition (4) is applied to the permutations of \mathbf{A} (all regarded as dissimilar as \mathbf{A} itself), it identifies the vertices of $Z(\mathbf{A})$, so that the convex mix of those gives $Z(\mathbf{A})$. Some of the matrices identified in this way cannot be obtained using the weights implied by Strong-MixC, which demonstrates Remark 6.

The reverse implication of Remark 6 is not true in general. The matrices \mathbf{A} and \mathbf{B} in (1) provide a counterexample in which $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ but the two matrices cannot be ranked by matrix majorization. Matrix majorization and zonotope inclusion orderings may coincide only under specific circumstances, such as those identified in Theorem 2 in Koshevoy (1995) or in cases where dissimilarity comparisons are limited to distributions where $d = 2$ (Dahl 1999; Lasso de la Vega and Volij 2014). We provide in the appendix a new geometric proof of the latter statement.

Remark 7 Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$: $Z(\mathbf{B}) \subseteq Z(\mathbf{A}) \Rightarrow \mathbf{B} \leq^R \mathbf{A}$.

The set of matrices that can be ordered in terms of dissimilarity can be further extended by considering the transitive closure originated by all binary relations \leq satisfying axioms in Theorem 1 and a new axiom, the Independence of Permutation of Groups (IPG), which introduces a property of invariance of the dissimilarity orderings with respect to the labeling of the groups.

Axiom 7 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 $\mathbf{B} \sim \mathbf{A}$.

In the context of segregation analysis, the axiom provides a natural multi-group extension of the symmetry of types property Hutchens (2015). Together with ISC, IEC, IPC and MixC, the axiom extends dissimilarity comparisons based on zonotope inclusion orderings that are not concerned with the labeling of the groups. A proof of the following corollary rests on the properties of the zonotope.

Corollary 1 For any $\mathbf{A} \in \mathcal{M}_d$, $\mathbf{B} \preceq \mathbf{A}$ for all axioms satisfying MixC, ISC, IEC, IPC and IPG if and only if $\exists \mathbf{\Pi}_d \in \mathcal{P}_d$ such that $Z(\mathbf{B}) \subseteq Z(\mathbf{\Pi}_d \mathbf{A})$.

3.4 Dissimilarity indices

The partial order of dissimilarity in Theorem 1 can be represented in terms of agreement of *dissimilarity indices* satisfying desirable properties. A dissimilarity index is a multivariate function $D : \mathcal{M}_d \rightarrow \mathbb{R}_+$ mapping a distribution matrix into a number, which can be interpreted as the level of dissimilarity among the d distributions represented in that matrix. These indices measure dissimilarity as the average of within-class *dispersion* of group frequencies.

Consider first the dissimilarity orderings consistent with Strong-MixC. In this case, dispersion of groups frequencies within each class can be quantified by a function h in the class \mathcal{H} of real valued convex functions defined on Δ^d . Dispersion in class j contributes to the overall dissimilarity proportionally to the size of the class j , \bar{a}_j . The dissimilarity index D_h with $h \in \mathcal{H}$ aggregates these evaluations as follows:¹²

$$D_h(\mathbf{A}) := \frac{1}{d} \sum_{j=1}^{n_A} \bar{a}_j \cdot h(a_{1j}/\bar{a}_j, \dots, a_{dj}/\bar{a}_j), \tag{5}$$

where a_{ij}/\bar{a}_j is the proportion of group i relative to the size of class j when groups are uniformly weighted. Dissimilarity is minimized when $a_{ij}/\bar{a}_j = 1/d$ for each of the d groups in all classes. Hence, by normalizing h so that $h\left(\frac{1}{d} \mathbf{1}'_d\right) = 0$, the index takes value 0 when perfect similarity is reached. Dissimilarity is instead maximal when for every j there exists a i such that $a_{ij} = \bar{a}_j$. The following proposition sets out a dominance condition in terms of dissimilarity indices.

Proposition 3 For any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, $\mathbf{B} \preceq \mathbf{A}$ for all orderings \preceq satisfying axioms Strong-MixC, ISC, IEC, IPC if and only if $D_h(\mathbf{B}) \leq D_h(\mathbf{A})$ for all $h \in \mathcal{H}$.

¹² For notational convenience empty classes receive weight $\bar{a} = 0$ and therefore do not contribute to the overall dissimilarity.

Using Proposition 1, we conclude that matrix majorization entails a necessary and sufficient condition for assessing agreement in dissimilarity evaluations for all indices described in Proposition 3. Matrix majorization is refined by the zonotope inclusion criterion. We provide an equivalent representation of the latter criterion in terms of dissimilarity measures based on the so-called *price dominance* criterion (Kolm 1977; Koshevoy and Mosler 1996; Andreoli and Zoli 2020).¹³ Consider a set of “prices” (or normative weights) $\mathbf{p} = (p_1, \dots, p_d)'$, which take on real values and therefore could also be negative, allowing to draw evaluations of the relative groups composition of each class of a distribution matrix by the implied “budget” (or weighted average) $\mathbf{p}'\mathbf{a}_j/\bar{a}_j$. In a case of perfect similarity, $\mathbf{a}_j/\bar{a}_j = \frac{1}{d}\mathbf{1}_d$ for all classes j . Therefore, perfect equality within each group i for all average realizations a_{ij}/\bar{a}_j across all classes j indicates lack of dissimilarity. The same consideration applies if all realizations in each class are weighted in $\mathbf{p}'\mathbf{a}_j/\bar{a}_j$ irrespective of the weighting vector \mathbf{p} . Each class contributes additively to dissimilarity, with evaluations indexed by a convex functional $\phi : \mathbb{R} \rightarrow \mathbb{R}$ such that $\phi(\mathbf{p}'\mathbf{a}_j/\bar{a}_j)$, which is introduced to quantify the inequality across the realizations of all classes. If we quantify this aggregate inequality by $\frac{1}{d} \sum_{j=1}^{n_A} \bar{a}_j \phi(\mathbf{p}' \cdot \mathbf{a}_j/\bar{a}_j)$, then its minimum level can be reached at $m := \phi(\sum_{i=1}^d p_i/d)$. The minimum bound for the dissimilarity index

$$D_{\mathbf{p},\phi}(\mathbf{A}) := \frac{1}{d} \sum_{j=1}^{n_A} \bar{a}_j \phi(\mathbf{p}' \cdot \mathbf{a}_j/\bar{a}_j) - m$$

is therefore 0.¹⁴

Proposition 4 *For any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, $\mathbf{B} \preceq \mathbf{A}$ for all orderings \preceq satisfying axioms MixC, ISC, IEC, IPC if and only if $D_{\mathbf{p},\phi}(\mathbf{B}) \leq D_{\mathbf{p},\phi}(\mathbf{A})$ for all $\mathbf{p} \in \mathbb{R}^d$ and for every ϕ convex.*

Proposition 4 provides the class of dissimilarity indices that are related to zonotope inclusion and defines a dominance condition that is weaker than that implied by Proposition 3.

4 Related orders

This section highlights the relevance of the dissimilarity model for the analysis of multi-group segregation and multivariate inequality. First, we argue that the zonotope inclusion criterion can be meaningfully used in the analysis of segregation with many (more than two) groups. We also demonstrate that widely used multi-group segregation indices characterized in the literature are consistent with dissimilarity preserving or reducing axioms. Second, we analyze the implications of the

¹³ Lemma 3 in Appendix formalizes this equivalence within the setting of this paper.

¹⁴ Notice that the theoretical bound does not depend on the sizes of \mathbf{A} but only on its structure: if $\mathbf{A} = \mathbf{S}$, then $D_{\mathbf{p},\phi}(\mathbf{S}) = 0$ for any $\mathbf{p} \in \mathbb{R}^d$ and ϕ convex.

dissimilarity model for multivariate orderings of dispersion, where the focus is on dissimilarity between the distributions of certain attributes and a normatively relevant benchmark distribution. In this case, the configuration that displays less dissimilarity among the underlying distributions and with respect to the benchmark distribution, also exhibits less multivariate inequality/dispersion. We prove that the zonotope inclusion criterion weakens some of the most widely adopted robust criteria in the multivariate inequality literature. Third, we emphasize the relevance of the dissimilarity axioms for the analysis of univariate inequality.

4.1 Segregation

Segregation arises when individuals with different characteristics (such as their ethnic origin or gender) are distributed unevenly across the neighborhoods of a city, or across the schools of a schooling district, or across jobs within a firm. In segregation analysis, the realizations of interest are categorical and not-ordered. Mainstream approaches to segregation focus on the two-groups case and postulate consistency with the partial order generated by non-intersecting segregation curves Duncan and Duncan (1955) as a baseline.

A 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 cumulative shares of groups 1 and group 2 across categories.

The ranking of two-groups distributions generated by non-intersecting segregation curves can be characterized through elementary segregation-reducing operations (see Hutchens 1991) or, alternatively, by matrix majorization (Lasso de la Vega and Volij 2014; Hutchens 2015). When segregation curves cross, distributions can be ranked by segregation indices consistent with the segregation curve ordering (Reardon and Firebaugh 2002; Reardon 2009). Alternatively, segregation curves have been used to assess the dissimilarity between each group distribution and the population distribution as in Alonso-Villar and del Rio (2010).

We are not, however, aware of any ordering generated by a multi-group expansions of the segregation curves ordering. Frankel and Volij (2011) have provided normative justifications for using matrix majorization as a robust segregation criterion for ranking multi-group distributions (see also Flückinger and Silber 1999; Chakravarty and Silber 2007). However, matrix majorization is a demanding condition that is not testable in the multi-group setting. Results in this paper deliver three contributions to this literature.

First, Proposition 1 clarifies that the operations of merge (or, equivalently, Strong-MixC), split, permutation and insertion/elimination of empty classes characterize the ranking produced by non-intersecting segregation curves when $d = 2$. The same axioms characterize matrix majorization when $d \geq 2$, thus showing that every segregation comparison involves a dissimilarity comparison. The dissimilarity axiom

MixC allows to weaken this criterion to a less partial ordering compared to matrix majorization, which can also be interpreted in terms of segregation.

Second, we promote the zonotopes inclusion criterion as a relevant multi-group extension of the segregation curve dominance criterion. The zonotope inclusion criterion is new in the segregation literature, it is testable and it allows to deal with the multi-dimensional nature of the data. In the case $d = 2$, segregation curve dominance is always consistent with zonotope inclusion, insofar the segregation curve can be understood as the lower boundary of a zonotope.¹⁵ Moreover, segregation curve dominance is also equivalent to matrix majorization.

In the multi-group setting ($d \geq 3$), however, a dominance criterion based on comparisons of segregation curves across all pairs of groups provides only a necessary condition for zonotope inclusion. Furthermore, in this context the zonotope inclusion criterion is weaker than matrix majorization, thus providing a natural refinement to it. Theorem 1 characterizes it in terms of MixC axiom, thus proving the link between multi-group segregation and dissimilarity.

Third, Proposition 3 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}/\bar{a}_j and a_{2j}/\bar{a}_j in each class. By setting

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

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 as $A_\omega(\mathbf{A}) := 1 - \sum_{j=1}^{n_A} \prod_{i=1}^d (a_{ij})^{\omega_i}$ for $\omega_i \geq 0$ such that $\sum_{i=1}^d \omega_i = 1$. By setting

$$h(a_{1j}/\bar{a}_j, \dots, a_{dj}/\bar{a}_j) := 1 - d \prod_{i=1}^d (a_{ij}/\bar{a}_j)^{\omega_i}$$

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) and Moulin (2016) is $M(\mathbf{A}) := \log_2(d) - \sum_{j=1}^{n_A} \left(\frac{\bar{a}_j}{d}\right) \sum_{i=1}^d \frac{a_{ij}}{\bar{a}_j} \cdot \log_2\left(\frac{\bar{a}_j}{a_{ij}}\right) + \frac{a_{ij}}{\bar{a}_j} \cdot \log_2\left(\frac{\bar{a}_j}{a_{ij}}\right)$ set equal to 0 if $a_{ij} = 0$. By setting

¹⁵ For any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$, $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ if and only if the segregation curve of \mathbf{B} lies nowhere below the segregation curve of \mathbf{A} .

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

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

4.2 Multivariate majorization and the Lorenz zonotope

In this section, the focus is on the multivariate majorization criteria that are adopted in robust inequality analysis. We argue that every inequality comparison involves the assessment of the dissimilarity between some relevant distributions and a benchmark distribution. A canonical example is that in which the distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ represent multivariate distributions of commodities. A matrix represents the way in which shares of each commodity (by row) are allocated to certain classes, which can represent the demographic units (e.g. families or individuals) that consume these commodities. Units are not ordered in any meaningful way.

Multidimensional inequality arises from the dissimilarity between the d distributions under analysis and the distribution of the demographic weight of the n units. It is common to assume that every unit receives a uniform weight equal to $1/n$. Under these circumstances, the next corollary, which follows from Proposition 1, formalizes the relation between multivariate inequality analysis and dissimilarity.

Corollary 2 *Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$. Then*

$$\tilde{\mathbf{B}} := \begin{pmatrix} \frac{1}{n_B} \mathbf{1}'_{n_B} \\ \mathbf{B} \end{pmatrix} \preceq \tilde{\mathbf{A}} := \begin{pmatrix} \frac{1}{n_A} \mathbf{1}'_{n_A} \\ \mathbf{A} \end{pmatrix} \tag{6}$$

for every dissimilarity ordering \preceq satisfying axioms MC (or Strong-MixC), ISC, IEC and IPC if and only if there exists $\mathbf{X} \in \mathcal{R}_{n_A, n_B}$ such that (i) $\mathbf{B} = \mathbf{A}\mathbf{X}$ and (ii) $\frac{n_A}{n_B} \mathbf{1}'_{n_B} = \mathbf{1}'_{n_A} \mathbf{X}$.

When $n_A = n_B = n$, matrix \mathbf{X} in the corollary is *doubly stochastic* ($\mathbf{X} \in \mathcal{D}_n$). 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). All social welfare functions that are increasing and Schur-concave (i.e. display some degree of inequality aversion) would rank the two multivariate distributions accordingly.

Uniform majorization is a demanding criterion, alike matrix majorization, insofar it posits that inequality can be reduced only when every row of a distribution matrix \mathbf{B} is obtained from the corresponding row of another distribution matrix \mathbf{A} through a common set of transformations implied by the matrix $\mathbf{X} \in \mathcal{D}_n$. The resulting ordering of distribution matrices is therefore partial. Koshevoy (1995) and Koshevoy and Mosler (1996) have studied a less partial order of multivariate distributions that is based on the *Lorenz zonotopes* inclusion criterion. A Lorenz zonotope, denoted $LZ(\mathbf{A})$ with $\mathbf{A} \in \mathcal{M}_d$, is a $d + 1$ dimensional zonotope of a distribution matrix \mathbf{A} augmented by the population distribution vector, that is $LZ(\mathbf{A}) := Z(\tilde{\mathbf{A}})$

with $LZ(\mathbf{A}) \in \mathbb{R}_+^{d+1}$ and $\tilde{\mathbf{A}}$ defined as in (6). The Lorenz zonotope inclusion criterion induces a partial order of distribution matrices that provides a testable refinement of uniform majorization.

The following chain of implications clarifies the relation between multivariate inequality and dissimilarity. The proof follows from previous results.

Remark 8 Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ such that $d \geq 2$ and $n_A = n_B = n$. Then:

$$\mathbf{B} = \mathbf{A}\mathbf{X}, \mathbf{X} \in \mathcal{D}_n \Rightarrow \mathbf{B} \leq^R \mathbf{A} \Rightarrow LZ(\mathbf{B}) \subseteq LZ(\mathbf{A}) \Rightarrow Z(\mathbf{B}) \subseteq Z(\mathbf{A})$$

The first implication, showing that uniform majorization implies matrix majorization, has been discussed in the previous section. The Lorenz zonotope inclusion criterion defines an inequality partial order of distribution matrices that is less partial than (i.e., is implied by) matrix majorization. It follows that the ranking of distribution matrices given by $LZ(\mathbf{B}) \subseteq LZ(\mathbf{A})$ is always consistent with the implications of a merge transformation (or, alternatively, a convex combination underlying Strong-MixC axiom) on matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$. Any such transformation bears two implications for multidimensional inequality.

First, a merge transformation reduces dissimilarity between the distribution of each dimension separately and the benchmark distribution, implying a reduction of inequality in each dimension. Second, a merge transformation reduces the dissimilarity across dimensions, implying an increase in correlation between dimensions. This aspect is controversial, since the Lorenz zonotopes inclusion criterion may fail to rank distribution matrices that are instead unanimously ordered by social welfare functions satisfying *aversion to correlation increasing transfers* (Epstein and Tanny 1980; Atkinson and Bourguignon 1982; Decancq 2012), a desirable property in multidimensional inequality analysis stating that any transformation that rises the degree of association in realizations is bound to decrease social welfare Andreoli and Zoli (2020). The Extended Lorenz zonotope inclusion criterion proposed in Mosler (2012) addresses these concerns.

Although the Lorenz zonotope inclusion criterion may be problematic for multidimensional welfare analysis, it is still a relevant criterion for assessing dissimilarity between distributions. The criterion $LZ(\mathbf{B}) \subseteq LZ(\mathbf{A})$ can be weakened by looking at inclusions of the projections of the Lorenz Zonotope in the space of outcomes, that is $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$. The latter is useful to analyze inequalities that arise from differences between distributions, regardless of the degree of inequality of each of these distributions. This feature is relevant, for instance, for constructing robust inequality of opportunity comparisons (see, for instance, Roemer and Trannoy 2016; Andreoli et al. 2019).¹⁶

Corollary 2 provides a characterization result that extends robust inequality assessments based on uniform majorization to matrices that possibly differ in size ($n_A \neq n_B$) but with the same number d of dimensions.

¹⁶ There are other robust multivariate inequality criteria that are weaker than uniform majorization (Dahl 1999; Martínez Pería et al. 2005) and that imply the Lorenz zonotope inclusion criterion. These criteria are also nested within the dissimilarity model.

4.3 Income inequality

Corollary 2 also holds in the case $d = 1$. This case is of particular interest for social welfare analysis, as it rationalizes empirical comparisons of *income inequality*. In this section, we argue that every income inequality comparison involves a dissimilarity comparison, but not the reverse.

Empirical comparisons of income inequality consist in assessing the way total income in a sample of n income recipient units (such as households or individuals) is split across these units. We can hence represent a distribution of income shares by the n -variate vectors $\mathbf{a}', \mathbf{b}' \in \mathcal{M}_1$, with $\mathbf{a}' \cdot \mathbf{1}_n = \mathbf{b}' \cdot \mathbf{1}_n = 1$. Each entry of the vectors corresponds to an income share allocated to a given unit. Anonymity is often invoked by the literature addressing income inequality measurement, thereby implying that any permutation of the units does not affect the extent of inequality displayed by \mathbf{a}' of \mathbf{b}' .

As per condition (6) in Corollary 2, every empirical income inequality comparison involves a dissimilarity comparison between the distribution of income shares owned by each of the n units and the units' weights. Furthermore, the ranking of distribution matrices induced by the *LZ* inclusion is consistent with Lorenz curves partial order. The chain of implications in Remark 8 hence runs in both directions when $d = 1$: Lorenz zonotopes inclusion implies unanimity for all social welfare functions that are increasing and concave in income which is equivalent to uniform majorization. All those conditions imply that income inequality analysis always subsumes a dissimilarity comparison.

A similar result holds even when the benchmark distribution of population weights is not uniform. 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 this case, the interest is in ranking matrices such as $\hat{\mathbf{A}} := (\boldsymbol{\omega}, \mathbf{a}')$, where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)'$ and ω_j can be understood as individual j 's weight. Using Corollary 2, every welfare-consistent measure of inequality can be written as an average of convex transformations of equalized incomes, scaled by their demographic weights. This is formalized by the inequality index $D_h(\hat{\mathbf{A}}) = \sum_{j=1}^n \omega_j h(a_j/\omega_j)$ with $h \in \mathcal{H}$ convex and a_j/ω_j is j 's equivalent income.¹⁷ From Proposition 3 (in combination with Remark 7), the zonotope inclusion criterion $Z((\boldsymbol{\omega}, \mathbf{b}') \subseteq Z((\boldsymbol{\omega}, \mathbf{a}'))$ provides a sufficient test for welfare dominance.

A well known result in inequality measurement is that an income distribution \mathbf{b}' displays less inequality than another distribution \mathbf{a}' 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 In the univariate case ($d = 1$), Corollary 2 implies that every sequence of *PD* transfers

¹⁷ The result (see Lemma 1 in the Appendix) follows from the homogeneity and convexity of $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, yielding $g(\omega_j, a_j) = \omega_j g(1, a_j/\omega_j) = \omega_j h(a_j/\omega_j)$ with h convex. Here we assume that income and population weights have unit size. If this is not the case, then a_j/ω_j is proportional to j 's equivalent income.

¹⁸ Consider a distribution \mathbf{a}' . A *PD* transfer consists in a movement of a mass $\epsilon > 0$ from class j to class k such that $a_j > a_k$, yielding $b_j = a_j - \epsilon$, $b_k = a_k + \epsilon$ and $b_\ell = a_\ell \forall \ell \neq j, k$ such that $b_j \geq b_k$. As a consequence of this transformation, $\mathbf{b}' = \mathbf{a}'\mathbf{X}$, $\mathbf{X} \in \mathcal{D}_n$ (Lorenz dominance) and $\sum_j a_j = \sum_j b_j$.

of incomes can be rationalized by a specific sequence of more fundamental dissimilarity preserving and reducing operations that are concerned with the way income shares and weights are shifted across units:

Corollary 3 *Every PD transfer operation 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* transfers. The proof of Corollary 3 rests on the fact that 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 and reducing operations of respectively split and merge characterize matrix majorization of which uniform majorization is a particular case.

The interesting and new result provided by Corollary 2 is that there always exists a sequence of split and merge operations that supports uniform majorization even in the multidimensional case ($d \geq 2$), although the same sequence cannot be generally rearranged to represent *PD* transfers Kolm (1977).

5 Concluding remarks

A large and sparse literature on segregation and inequality measurement has proposed criteria for ranking multi-group distributions according to the dissimilarity they exhibit. This paper establishes the axiomatic foundations of the dissimilarity criterion. We do so by developing a parsimonious axiomatic model which is based on dissimilarity preserving operations and a dissimilarity reducing operation, the merge, which consists in aggregating, distribution by distribution, the proportion of people observed in two separate classes. We study the partial order of distribution matrices originated by the transitive closure of all binary dissimilarity relations consistent with the operations and with a mixing axiom. This last axiom is crucial to justify an equivalent characterization of the “displays at most as much dissimilarity as” partial order. Our main theorem identifies a novel non-parametric criterion, based on inclusion of the zonotope set representations of the data, which is equivalent to the dissimilarity partial order thus identified.

The zonotope inclusion criterion is relevant in many contexts. One application could be in evaluating the impact of certain educational policies for the patterns of dissimilarity between multi-ethnic distributions of students across schools in a district. This problem is commonly referred to as schooling segregation. We can use zonotopes to conclude about robust changes in segregation when comparing the actual situation and a counterfactual distribution that would have emerged in the absence of the policy. While the education policy itself may have little to

do with splitting, merging or mixing schools, the zonotope inclusion signals that one can always move from the counterfactual to the actual allocation of students through operations that are unanimously understood as segregation-reducing.

In some cases, zonotopes inclusion is rejected by the data. The dissimilarity indices analyzed in the paper allow to produce conclusive evaluations about the changes in dissimilarity. The implied ranking is always 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.

Appendix A: Proofs

Appendix A.1 Useful additional 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 $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d, \mathbf{B} \preceq^R \mathbf{A}$ if and only if*

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

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

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 by $\mathcal{R}_{n_A, n_B}^{IEC} \subset \mathcal{R}_{n_A, n_B}$ the set of all matrices reproducing an insertion of empty classes when post-multiplied to a distribution matrix \mathbf{A} . Hence $\mathbf{Y} \in \mathcal{R}_{n_A, n_B}^{IEC}$ 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 $\mathbf{b}_{j+1} = \mathbf{a}_{j+1} + (1 - \lambda)\mathbf{a}_j = (1 - \lambda)\mathbf{a}_j$ for $\lambda \in [0, 1]$. Let $k \neq j$, the set of all transformation matrices $\mathbf{Z}_{[j]}$ reproducing a split of classes is denoted by:

$$\mathcal{R}_A^{ISC} := \left\{ \mathbf{Z}_{[j]} \in \mathcal{R}_n : \begin{array}{l} z_{jj} := \lambda, \quad z_{j,j+1} := (1 - \lambda), \quad \lambda \in [0, 1], j \in \mathcal{J}_A, j + 1 \in \mathcal{J}_A^0, \\ z_{kk} := 1, \quad z_{ih} := 0 \text{ in all other cases} \end{array} \right\}.$$

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 $j + 1$. Such a matrix belongs to the set:

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

The *third preliminary result* provides an equivalent (algebraic and finite) condition for testing zonotope inclusion. Let denote first the sets $\mathcal{V}_n := \{ \mathbf{v} : v_j \in [0, 1], j = 1, \dots, n \}$ and $\mathcal{V}_n^{01} = \{ \mathbf{v} : v_j \in \{0, 1\}, j = 1, \dots, n \}$.

Lemma 2 *Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ of size $d \times n$ such that $\mathbf{1}'_d \mathbf{B} = \mathbf{1}'_d \mathbf{A} = \frac{d}{n} \mathbf{1}'_n$. i) $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ if and only if ii) $\forall \boldsymbol{\gamma} \in \mathcal{V}_n^{01} \exists \boldsymbol{\theta} \in \mathcal{V}_n : \mathbf{B}\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\theta}$.*

Proof Note that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ if and only if for any $\mathbf{z} \in Z(\mathbf{B})$ then $\mathbf{z} \in Z(\mathbf{A})$, whereas for $Z(\mathbf{B}) \subset Z(\mathbf{A})$ we also have that $\exists \tilde{\mathbf{z}} \in Z(\mathbf{A})$ such that $\tilde{\mathbf{z}} \notin Z(\mathbf{B})$. Using the Minkowski sum properties, this means that $\forall \boldsymbol{\gamma} \in \mathcal{V}_n \exists \boldsymbol{\theta} \in \mathcal{V}_n : \mathbf{z} := \sum_j \gamma_j \mathbf{b}_j = \sum_j \theta_j \mathbf{a}_j$, that is $\mathbf{B}\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\theta}$ for $\boldsymbol{\gamma}, \boldsymbol{\theta} \in \mathcal{V}_n$. Let assume, for ease of notation, that $\boldsymbol{\gamma} \in \mathcal{V}_n$ is ordered so that $0 \leq \gamma_1 \leq \gamma_2 \leq \dots \leq \gamma_n \leq 1$. Denote $\delta_1 = \gamma_1$ and $\delta_k = \gamma_k - \gamma_{k-1}$ for $k = 2, \dots, n$. Note that $\delta_k \in [0, 1] \ \forall k$. Recall that \mathbf{i}_k is the column vector k of the identity matrix of size n and denote further $\mathbf{1}_{(h,h+1,\dots,n)} := \sum_{k=h}^n \mathbf{i}_k$ for any $1 \leq h \leq n$ so that if $h = n - 1$ then $\mathbf{1}_{(n-1,n)} = \mathbf{i}_{n-1} + \mathbf{i}_n$ and so on. We have that $\mathbf{1}_{(h,h+1,\dots,n)} \in \mathcal{V}_n^{01} \ \forall h$. For any $\boldsymbol{\gamma} \in \mathcal{V}_n$ there always exists a class j such that $\boldsymbol{\gamma} = \sum_{h=j}^n \delta_h \mathbf{1}_{(h,h+1,\dots,n)}$, so that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ is equivalent to

$$\mathbf{B} \cdot \sum_{h=j}^n \delta_h \cdot \mathbf{1}_{(h,h+1,\dots,n)} = \mathbf{A}\boldsymbol{\theta}, \tag{7}$$

for $\boldsymbol{\gamma}, \boldsymbol{\theta} \in \mathcal{V}_n$. We use this equivalence in the proof.

i) \Rightarrow ii). Immediate, since $\mathcal{V}_n^{01} \subseteq \mathcal{V}_n$.

ii) \Rightarrow i). Assume that ii) holds, which can be equivalently stated as: for any $1 \leq h \leq n$ there exists $\boldsymbol{\theta}_{(h,h+1,\dots,n)} \in \mathcal{V}_n : \mathbf{B}\mathbf{1}_{(h,h+1,\dots,n)} = \mathbf{A}\boldsymbol{\theta}_{(h,h+1,\dots,n)}$. Consider substituting into $\mathbf{B} \cdot \sum_{h=j}^n \delta_h \cdot \mathbf{1}_{(h,h+1,\dots,n)}$, yielding for any j :

$$\begin{aligned} \sum_{h=j}^n \mathbf{B} \cdot \delta_h \mathbf{1}_{(h,h+1,\dots,n)} &= \sum_{h=j}^n \mathbf{A} \cdot \delta_h \boldsymbol{\theta}_{(h,h+1,\dots,n)} \\ &= \mathbf{A} \cdot \sum_{h=j}^n \delta_h \boldsymbol{\theta}_{(h,h+1,\dots,n)}. \end{aligned}$$

Since $\sum_{h=j}^n \delta_h \boldsymbol{\theta}_{(h,h+1,\dots,n)} \in \mathcal{V}_n$ for any $\delta_h \in [0, 1]$ and for any j , the latter equation implies (7) and then i). □

The *fourth and last preliminary result* shows the equivalence between the zonotope inclusion criterion and price majorization. Although the proof largely draws on Koshevoy (1995), the setting we investigate is logically distinct, insofar the price majorization we use invokes d -dimensional prices to evaluate inclusion

of d -dimensional zonotopes, whereas Koshevoy shows the equivalence with $d + 1$ -dimensional extensions of the zonotope (the Lorenz zonotope). Denote by \mathcal{C}_n the set of *column stochastic* matrices, so that $\mathbf{Y} \in \mathcal{C}_n \Leftrightarrow \mathbf{Y}' \in \mathcal{R}_n$. Define: " \mathbf{B} is price majorized by \mathbf{A} " whenever $\forall \mathbf{p} \in \mathbb{R}^d, \exists \mathbf{Y} \in \mathcal{C}_n$ such that $\mathbf{p}'\mathbf{B} = \mathbf{p}'\mathbf{A}\mathbf{Y}$. Given that we consider matrices in \mathcal{M}_d , then condition $\mathbf{p}'\mathbf{B}\mathbf{1}_n = \mathbf{p}'\mathbf{A}\mathbf{1}_n$ is satisfied by construction, therefore it suffices to consider only transformations $\mathbf{Y} \in \mathcal{C}_n$ to obtain price majorization.

Lemma 3 *Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ of size $d \times n$ such that $\mathbf{1}'_d \mathbf{B} = \mathbf{1}'_d \mathbf{A} = \frac{d}{n} \mathbf{1}'_n$. i) $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ if and only if ii) $\forall \mathbf{p} \in \mathbb{R}^d \exists \mathbf{Y} \in \mathcal{C}_n$ such that $\mathbf{p}'\mathbf{B} = \mathbf{p}'\mathbf{A}\mathbf{Y}$.*

Proof i) \Rightarrow ii). Assume that i) holds, then from Lemma (2) we have that $\forall k \in \{1, \dots, n\} \exists \theta_{jk} \in [0, 1]$ for $j = 1, 2, \dots, n$ such that $\mathbf{b}_k = \sum_j \theta_{jk} \mathbf{a}_j$. In compact notation:

$$\mathbf{B} = \sum_{j=1}^n (\theta_{j1} \mathbf{a}_j, \dots, \theta_{jn} \mathbf{a}_j) = \mathbf{A}\Theta. \tag{8}$$

Matrix Θ may not be row stochastic but it is guaranteed that $\mathbf{B}\mathbf{1}_n = \mathbf{A}\mathbf{1}_n$. Using the fact that $\mathbf{1}'_d \mathbf{B} = \mathbf{1}'_d \mathbf{A}$ we get from (8) that $\mathbf{1}'_d \mathbf{A}\Theta = \mathbf{1}'_d \mathbf{A}$, which gives $\sum_i a_{ik} = \sum_i \sum_j a_{ij} \theta_{jk} = \sum_j \theta_{jk} \sum_i a_{ij}, \forall k$. Given that $\sum_i a_{ik} = \sum_i a_{ij} = d/n$ for any $k \neq j$, by definition of \mathbf{A}, \mathbf{B} , we get $\sum_j \theta_{jk} = 1$, which means that $\Theta \in \mathcal{C}_n$. Hence i) $\Rightarrow \mathbf{B} = \mathbf{A}\Theta, \Theta \in \mathcal{C}_n \Rightarrow \mathbf{p}'\mathbf{B} = \mathbf{p}'\mathbf{A}\Theta$ for all $\mathbf{p} \in \mathbb{R}^d$ and $\Theta \in \mathcal{C}_n$, which is ii).

ii) \Rightarrow i). Assume that ii) holds, which implies that $\mathbf{p}'\mathbf{B}$ Lorenz dominates $\mathbf{p}'\mathbf{A}$ for any $\mathbf{p} \in \mathbb{R}^d$, that is $\sum_{j=1}^k \mathbf{p}'\mathbf{b}_{(j)} \geq \sum_{j=1}^k \mathbf{p}'\mathbf{a}_{(j)} \forall k = 1, \dots, n$ where for each $\mathbf{p} \in \mathbb{R}^d$ the classes of \mathbf{B} have been ordered by increasing magnitude of $\mathbf{p}'\mathbf{b}_{(j)}$, so that $\mathbf{p}'\mathbf{b}_{(j)} \leq \mathbf{p}'\mathbf{b}_{(j+1)} \forall j$ (and similarly for the classes of \mathbf{A}). This is equivalent (see Marshall et al. 2011) to:

$$\max_{j_1, \dots, j_k} \mathbf{p}' \cdot (\mathbf{b}_{j_1} + \dots + \mathbf{b}_{j_k}) \leq \max_{j_1, \dots, j_k} \mathbf{p}' \cdot (\mathbf{a}_{j_1} + \dots + \mathbf{a}_{j_k}) \forall k, \forall \mathbf{p} \in \mathbb{R}^d,$$

where the max operator selects the k -tuple of columns of \mathbf{A} and \mathbf{B} that yield the largest value when multiplied by any vector of prices \mathbf{p} . An equivalent formulation is:

$$\max_{j_1, \dots, j_k} \mathbf{p}'\mathbf{B} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k}) \leq \max_{j_1, \dots, j_k} \mathbf{p}'\mathbf{A} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k}) \forall k, \forall \mathbf{p} \in \mathbb{R}^d,$$

which implies

$$\mathbf{p}'\mathbf{B} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k}) \leq \max_{j_1, \dots, j_k} \mathbf{p}'\mathbf{A} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k}) \forall k, \forall j_1, \dots, j_k, \forall \mathbf{p} \in \mathbb{R}^d.$$

The previous condition identifies a situation where $\mathbf{B} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k})$ lies in the convex hull of $\mathbf{A} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k})$, see Koshevoy (1995). Making use of the definition of convex hull inclusion, we can alternatively write:

$$\begin{aligned} & \mathbf{B} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k}) \\ &= \sum_{\forall j_1, \dots, j_k} \alpha_{j_1, \dots, j_k} \mathbf{A} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k}), \alpha_{j_1, \dots, j_k} \in [0, 1], \sum \alpha_{j_1, \dots, j_k} = 1, \forall k, \forall j_1, \dots, j_k, \end{aligned} \tag{9}$$

or equivalently $\mathbf{B} \cdot (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k}) = \mathbf{A} \cdot \sum_{\forall j_1, \dots, j_k} \alpha_{j_1, \dots, j_k} (\mathbf{i}_{j_1} + \dots + \mathbf{i}_{j_k}), \forall k, \forall j_1, \dots, j_k$. Noticing that $\bigcup_k \{(\mathbf{i}_{i_1} + \dots + \mathbf{i}_{i_k}) : i_1, \dots, i_k \in \{1, \dots, n\}\} = \mathcal{V}_n^{01}$ and that $\bigcup_k \{\sum_{\forall i_1, \dots, i_k} \alpha_{j_1, \dots, j_k} (\mathbf{i}_{i_1} + \dots + \mathbf{i}_{i_k}) : i_1, \dots, i_k \in \{1, \dots, n\}\} = \mathcal{V}_n$, one has that (9) is equivalent to $\forall \boldsymbol{\gamma} \in \mathcal{V}_n^{01} \exists \boldsymbol{\theta} \in \mathcal{V}_n: \mathbf{B}\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\theta}$. From Lemma 2, the latter condition implies i). □

Appendix A.2 Proof of Proposition 1

Proof i) \Rightarrow ii). If i) holds, then it follows from Donaldson and Weymark (1998) that there exist partial orders that rank \mathbf{B} at most as dissimilar as \mathbf{A} and that these orders are consistent with the operations underlying axioms MC, IPC, IEC and ISC. Matrix majorization \preceq^R is one of such partial orders. In fact, as highlighted in the preliminary results section, every operation underlying axioms MC, ISC, IEC and IPC can be represented by a row stochastic matrix transformation, respectively by $\mathbf{M} \in \mathcal{R}_{n_A}^{MC}$ for a merge of classes operation, by $\mathbf{S} \in \mathcal{R}_{n_A}^{ISC}$ for a split operation, by $\mathbf{E} \in \mathcal{R}_{n_A, n}^{IEC}$ for insertion/elimination of empty classes and by $\mathbf{\Pi} \in \mathcal{P}_{n_A}$ for permutations of classes. A sequence of these operations is for instance $\mathbf{X} = \mathbf{MSE\Pi} \in \mathcal{R}_{n_A, n}$ since the set of all row stochastic matrices \mathcal{R} is closed with respect to the product operation. Any \mathbf{X} obtained through these operations is row stochastic, thereby implying that ii) holds.

ii) \Rightarrow i). Assume that ii) holds, hence $\mathbf{B} = \mathbf{A}\boldsymbol{\Theta}$ with $\boldsymbol{\Theta} \in \mathcal{R}_{n_A, n_B}$. In shorthand notation

$$\mathbf{B} = \sum_{j=1}^{n_A} (\theta_j(1)\mathbf{a}_j, \dots, \theta_j(n_B)\mathbf{a}_j) \tag{10}$$

where $\theta_j(k)$ denotes the generic element θ_{jk} of $\boldsymbol{\Theta}$. Each addendum in (10) can be written as:

$$\left(\lambda_{j1}\mathbf{a}_j, \lambda_{j2}(1 - \lambda_{j1})\mathbf{a}_j, \dots, \lambda_{j(n_B-1)} \prod_{1 \leq k < n_B-1} (1 - \lambda_{jk})\mathbf{a}_j, \prod_{1 \leq k \leq n_B-1} (1 - \lambda_{jk})\mathbf{a}_j \right), \tag{11}$$

where each $\lambda \in [0, 1]$. 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:

$$\begin{aligned} & \theta(1) = \lambda_1 \in [0, 1] \\ & \theta(k) = \lambda_k \left(1 - \sum_{j=1}^{k-1} \theta(j) \right) \text{ with } \lambda_j \in [0, 1] \forall j = 2, \dots, n_B. \end{aligned} \tag{12}$$

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 (12) leads to (11) 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 $\mathbf{Z}_{[k]} \in \mathcal{R}_A^{ISC}$.¹⁹ 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 (11) can be equivalently written as:

$$\left(\lambda_{j1}\mathbf{a}_j, \dots, \prod_{1 \leq k < n_B-1} (1 - \lambda_{jk})\mathbf{a}_j, \mathbf{0}_d \right) \cdot \mathbf{Z}_{[n_B-1]} = (\mathbf{a}_j, \mathbf{0}_d, \dots, \mathbf{0}_d) \cdot \prod_{1 \leq k \leq n_B-1} \mathbf{Z}_{[k]}. \tag{13}$$

Extending the representation in (13) to all addends in (10) 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 classes to \mathbf{A} to perform the n splits. According to the summation operator in (10), the order of the classes of \mathbf{A} is irrelevant. Thus operations of *permutations of classes* are admitted.²⁰ 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:

$$\widehat{\mathbf{X}} := \Pi_{n_A} \cdot \mathbf{Y} \cdot \text{diag} \left(\prod_{k=1}^{n_B-1} \mathbf{Z}_{[k]}(1), \dots, \prod_{k=1}^{n_B-1} \mathbf{Z}_{[k]}(n_A) \right) \tag{14}$$

$$= \Pi_{n_A} \cdot \mathbf{Y} \cdot \prod_{j=1}^{n_A} \left(\prod_{k=1}^{n_B-1} \widetilde{\mathbf{Z}}_{[k]}(j) \right), \tag{15}$$

where $\mathbf{Z}_{[k]}(j)$ is indexed for j to highlight the relation with the class j in \mathbf{A} . Here $\widetilde{\mathbf{Z}}_{[k]}(j) := \text{diag}(\mathbf{I}, \mathbf{Z}_{[k]}(j), \mathbf{I}')$ and \mathbf{I} and \mathbf{I}' are two identity matrices of size $(j - 1)n_B$ and $(n_A - j)n_B$ respectively. Line (15) 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 $\Pi_{n_A n_B}$. 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 $(\lambda_{1j}\mathbf{a}_1, \dots, \lambda_{n_A j}\mathbf{a}_{n_A}, \dots)$.²¹ 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

¹⁹ See the preliminary results in Appendix A.1.

²⁰ The two operations of permutation and insertion of classes transform \mathbf{A} into

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

²¹ Formally: $\mathbf{A} \cdot \mathbf{X} \cdot \Pi_{n_A n_B} = (\lambda_{11}\mathbf{a}_1, \dots, \lambda_{n_A 1}\mathbf{a}_{n_A}, \dots, \lambda_{1n_B}\mathbf{a}_1, \dots, \lambda_{n_A n_B}\mathbf{a}_{n_A})$.

with class 3 and so on, up to the first n_A classes. The sequence of merge transformations can be modeled by matrices $\mathbf{M}_{[1]} \in \mathcal{R}_{n_A n_B}^{MC}$, $\mathbf{M}_{[2]} \in \mathcal{R}_{n_A n_B}^{MC}$ and so on, up to $\mathbf{M}_{[n_A-1]} \in \mathcal{R}_{n_A n_B}^{MC}$, 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:

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

All the matrices multiplying \mathbf{A} are row stochastic, and the resulting matrix is also row stochastic. Hence $\Theta \in \mathcal{R}_{n_A n_B}$ can be always decomposed in permutation transformations of the product of matrices originated exclusively by split, merge and insertion/deletion of empty classes. This is i), which concludes the proof. \square

Appendix A.3 Proof of Remark 2

Proof Assume $\mathbf{B}^j \preceq^R \mathbf{A}$, then (by Proposition 1) $\exists \mathbf{X}^j \in \mathcal{R}_n$ such that $\mathbf{B}^j = \mathbf{A}\mathbf{X}^j$, $\forall j$. Recall that axiom Strong-MixC requires that $\mathbf{B} = \sum_{j=1}^m w_j \mathbf{B}^j$, hence $\mathbf{B} = \sum_{j=1}^m w_j \mathbf{B}^j = \sum_{j=1}^m w_j \mathbf{A}\mathbf{X}^j = \mathbf{A} \sum_{j=1}^m w_j \mathbf{X}^j \preceq^R \mathbf{A}$ since $\sum_{j=1}^m w_j \mathbf{X}^j \in \mathcal{R}_n$ given that $\sum_{j=1}^m w_j = 1$. Thus, $\mathbf{B} \preceq^R \mathbf{A}$. \square

Appendix A.4 Proof of Remark 3

Proof A direct verification of the remark can be obtained by considering matrices $\mathbf{A}, \hat{\mathbf{A}} \in \mathcal{M}_d$ such that $\hat{\mathbf{A}}$ is obtained by permuting columns j and $j + 1$ of \mathbf{A} . Thus, by IPC we have $\mathbf{A} \sim \hat{\mathbf{A}}$. Setting $m = 2$, with $\mathbf{B}^1 := \mathbf{A}$ and $\mathbf{B}^2 := \hat{\mathbf{A}}$ in the definition of Strong-MixC and letting $w_1 = w_2 = 1/2$, we obtain $\mathbf{B}^0 = 1/2 \cdot \mathbf{B}^1 + 1/2 \cdot \mathbf{B}^2 = 1/2 \cdot (\mathbf{A} + \hat{\mathbf{A}})$. That is, \mathbf{B}^0 coincides with \mathbf{A} except for columns j and $j + 1$ whose vectors are identical and coincide with $1/2 \cdot \mathbf{a}_j + 1/2 \cdot \mathbf{a}_{j+1}$. By Strong-MixC we have that $\mathbf{B}^0 \preceq \mathbf{A}$. Consider matrix $\mathbf{B} \in \mathcal{M}_d$ that is identical to \mathbf{B}^0 except for columns j and $j + 1$ where $\mathbf{b}_j = \mathbf{0}_d$ and $\mathbf{b}_{j+1} = \mathbf{a}_j + \mathbf{a}_{j+1}$. If we drop the empty column j from \mathbf{B} and split the class/column $j + 1$ with weights 1/2 and 1/2 into two adjacent classes we obtain matrix \mathbf{B}^0 . Applying IEC and ISC we have that $\mathbf{B} \sim \mathbf{B}^0$. By transitivity of the dissimilarity relation $\mathbf{B} \sim \mathbf{B}^0 \preceq \mathbf{A}$ implies that $\mathbf{B} \preceq \mathbf{A}$. Note that by construction matrices \mathbf{A} and \mathbf{B} are those in the definition of axiom MC that turns out to be satisfied. \square

Appendix A.5 Proof of Proposition 2

Proof i) \Rightarrow ii). The proof consists in showing that the set of all matrices \mathbf{B} such that $\mathbf{B} \preceq^R \mathbf{A}$ can be characterized using exclusively operations underlying axioms Strong-MixC, IEC, ISC and IPC. The result follows from the fact that if i) holds, then it follows from Donaldson and Weymark (1998) that there exist partial orders that rank \mathbf{B} at most as dissimilar as \mathbf{A} and that these orders are consistent with the operations underlying axioms ISC, IEC, IPC and Strong-MixC, and matrix majorization \preceq^R is one of such partial orders (by Proposition 1 and Remark 2).

The result holds for any matrix in \mathcal{M}_d . Let $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ with n_A not necessarily equal to n_B . First, consider matrices $\tilde{\mathbf{A}}, \tilde{\mathbf{B}} \in \mathcal{M}_d$ of size $d \times n$ obtained from \mathbf{A}, \mathbf{B} through split and permutation of classes and deletion of empty classes such that $\mathbf{1}'_d \tilde{\mathbf{A}} = \mathbf{1}'_d \tilde{\mathbf{B}} = \frac{d}{n} \mathbf{1}'_n$. Every ordering \preceq satisfying axioms IEC, ISC and IPC ranks $\tilde{\mathbf{A}} \sim \mathbf{A}$ and $\tilde{\mathbf{B}} \sim \mathbf{B}$.

We now investigate the implications for the transitive closure of the orderings \preceq (satisfying axioms IPC, IEC, ESC) deriving from the fact that it satisfies Strong-MixC. We do so by looking at all permutations of a matrix $\tilde{\mathbf{A}}$, that turn out to be indifferent to it in terms of dissimilarity for all orderings satisfying IPC²², and applying to them the mixing operations considered in Strong-MixC, thereby obtaining the set of matrices that are matrix majorized by $\tilde{\mathbf{A}}$. The result holds for any initial matrix $\mathbf{A} \in \mathcal{M}_d$. Denote by $J_k, k = 2, \dots, n$ a partition of the classes $\{1, \dots, n\}$ with cardinality $|J_k| = k$. For instance, $J_2 = \{1, 4\}$ when $n \geq 4$. There are $m_k = \binom{n}{k}$ partitions for any k , with $m = \sum_k m_k$, that are collected in the set \mathcal{J}_k , so that $J_k \in \mathcal{J}_k$ and $J_k \subseteq \{1, \dots, n\}$. For any such partition J_k , define $\mathcal{P}_n^{J_k}$ the set of $n \times n$ permutation matrices corresponding to all permutations of indices in J_k (there are $k!$ of them). The matrix $\tilde{\mathbf{A}}\Pi$, for $\Pi \in \mathcal{P}_n^{J_k} \forall k$ is obtained by permuting k columns indexed as in J_k . Given that \preceq satisfies IPC, then $\tilde{\mathbf{A}}\Pi \sim \tilde{\mathbf{A}}$. Consider a mix operation as defined in Strong-MixC axiom, that gives equal weight $w_h = \frac{1}{k!}$ to each permutation $h = 1, \dots, k!$ of indices in J_k , yielding $\sum_{h=1}^{k!} w_h \tilde{\mathbf{A}}\Pi = \sum_{\Pi \in \mathcal{P}_n^{J_k}} \frac{1}{k!} \tilde{\mathbf{A}}\Pi = \tilde{\mathbf{A}} \cdot \sum_{\Pi \in \mathcal{P}_n^{J_k}} \frac{1}{k!} \Pi$. There are m of such matrices $\sum_{\Pi \in \mathcal{P}_n^{J_k}} \frac{1}{k!} \Pi \in \mathcal{R}_n$, that can be denoted by \mathbf{X}^j , for $j = 1, \dots, m$. Note that matrices \mathbf{X}^j form the basis of the set \mathcal{R}_n (see Proposition 3.1 in Dahl 1999).²³ We hence have that $\tilde{\mathbf{A}}\mathbf{X}^j \preceq \tilde{\mathbf{A}} \forall j$ for all orderings \preceq satisfying Strong-MixC and IPC, moreover we have that $\tilde{\mathbf{A}}\mathbf{X}^j \preceq^R \tilde{\mathbf{A}}$ given that all \mathbf{X}^j are in \mathcal{R}_n . Consider now obtaining $\tilde{\mathbf{B}}$ by mixing (using the weights in axiom Strong-MixC) matrices $\tilde{\mathbf{A}}\mathbf{X}^j \forall j$, such as: $\tilde{\mathbf{B}} = \sum_{j=1}^m w_j \tilde{\mathbf{A}}\mathbf{X}^j = \tilde{\mathbf{A}} \sum_{j=1}^m w_j \mathbf{X}^j$. Since $\sum_j w_j = 1$, the summation term identifies the convex hull of all row-stochastic matrices that form the basis for the set \mathcal{R}_n . From Corollary 3.2 in Dahl (1999), the set of matrices $\tilde{\mathbf{B}} = \tilde{\mathbf{A}} \sum_{j=1}^m w_j \mathbf{X}^j$ obtained for all $w_j \in [0, 1]$ and $\sum_j w_j = 1$ identifies the whole set of matrices that are matrix majorized by $\tilde{\mathbf{A}}$ (which is called a Markotope), thus

²² Note that all orderings \preceq satisfying IEC, ISC and IPC rank at unanimity $\tilde{\mathbf{B}} \sim \tilde{\mathbf{A}}$ if and only if $\tilde{\mathbf{A}} = \tilde{\mathbf{B}}\mathbf{Y}$ and $\mathbf{Y} \in \mathcal{P}_n$. In fact, any other admissible matrix $\mathbf{Y} \in \mathcal{R}_n^{IEC} \cup \mathcal{R}_n^{ISC}$ would not guarantee comparability of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}\mathbf{Y}$, that is $\mathbf{1}'_d \tilde{\mathbf{A}} = \frac{d}{n} \mathbf{1}'_n$ whereas $\mathbf{1}'_d \tilde{\mathbf{B}}\mathbf{Y} \neq \frac{d}{n} \mathbf{1}'_n$.

²³ Note further that matrix $\tilde{\mathbf{A}}$ has uniform margins, implying that \mathbf{X}^j must be row and column stochastic.

guaranteeing that $\tilde{\mathbf{B}} \preceq^R \tilde{\mathbf{A}}$. Recall now (see the second set of preliminary results) that because of the transformations of split and permutation of classes and deletion of empty classes that have generated matrices $\tilde{\mathbf{A}}, \tilde{\mathbf{B}} \in \mathcal{M}_d$ such that $\tilde{\mathbf{A}} \sim \mathbf{A}$ and $\tilde{\mathbf{B}} \sim \mathbf{B}$ we have that it is possible to obtain $\tilde{\mathbf{A}}$ from \mathbf{A} through a row stochastic transformation and also the other way round making use of another row stochastic transformation, and similarly for $\tilde{\mathbf{B}}$ and \mathbf{B} , thus $\mathbf{B} \preceq^R \tilde{\mathbf{B}}$ and $\tilde{\mathbf{A}} \preceq^R \mathbf{A}$ that combined with $\tilde{\mathbf{B}} \preceq^R \tilde{\mathbf{A}}$ by transitivity gives condition ii).

ii) \Rightarrow i). Assume ii), which is equivalent to $\mathbf{B} \preceq \mathbf{A}$ by all dissimilarity orders satisfying MC, IPC, IEC, ISC from Proposition 1 and implies i) by Remark 3. \square

Appendix A.6 Proof of Remark 4

Proof The proof relies heavily on Remark 5 and Remark 7, which are demonstrated later. From Remark 7, $\mathbf{B} \preceq^R \mathbf{A}$ is equivalently represented by $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ when $d = 2$. By Remark 5, zonotope inclusion is consistent with the mixing operations underlying axiom MixC. \square

Appendix A.7 Proof of Remark 5

Proof Consider a $d \times n$ matrix $\mathbf{A} \in \mathcal{M}_d$ and a sequence of $d \times n$ matrices $\mathbf{B}^j \in \mathcal{M}_d$ for $j = 1, \dots, m$ such that $\mathbf{1}'_d \mathbf{A} = \mathbf{1}'_d \mathbf{B}^j = \frac{d}{n} \mathbf{1}'_n \forall j$. Note that for any sequence $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}^j \in \mathcal{M}_d$ one can find such matrices \mathbf{A}, \mathbf{B}^j defined as above such that $Z(\mathbf{A}) = Z(\tilde{\mathbf{A}})$ and $Z(\mathbf{B}^j) = Z(\tilde{\mathbf{B}}^j), \forall j$. Assume that $Z(\mathbf{B}^j) \subseteq Z(\mathbf{A}), \forall j$. By Lemma 2, $\forall \boldsymbol{\gamma} \in \mathcal{V}_n^{01} \exists \boldsymbol{\theta}^j \in \mathcal{V}_n$ such that $\mathbf{B}^j \boldsymbol{\gamma}^j = \mathbf{A} \boldsymbol{\theta}^j, \forall j$. Obtain $\mathbf{B} \in \mathcal{M}_d$ as per (4), then $\forall \boldsymbol{\gamma} \in \mathcal{V}_n^{01} \exists \boldsymbol{\gamma}^j \in \mathcal{V}_n^{01}$ such that $\mathbf{B} \boldsymbol{\gamma} \in \text{conv}\{\mathbf{B}^1 \boldsymbol{\gamma}^1, \dots, \mathbf{B}^m \boldsymbol{\gamma}^m\}$. There hence exist weights denoted $\omega_j^{\boldsymbol{\gamma}}$ that are specific to the choice of the mixing $\boldsymbol{\gamma}$, satisfying $\omega_j^{\boldsymbol{\gamma}} \in [0, 1]$ and $\sum_{j=1}^m \omega_j^{\boldsymbol{\gamma}} = 1$ such that $\mathbf{B} \boldsymbol{\gamma} = \sum_j \omega_j^{\boldsymbol{\gamma}} \mathbf{B}^j \boldsymbol{\gamma}^j = \sum_j \omega_j^{\boldsymbol{\gamma}} \mathbf{A} \boldsymbol{\theta}^j = \mathbf{A} \sum_j \omega_j^{\boldsymbol{\gamma}} \boldsymbol{\theta}^j = \mathbf{A} \boldsymbol{\theta}$ with $\boldsymbol{\theta}^j \in \mathcal{V}_n \forall j$. Since \mathcal{V}_n is closed to convex combinations, then $\boldsymbol{\theta} \in \mathcal{V}_n$, which by Lemma 2 is a sufficient condition for $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$. \square

Appendix A.8 Proof of Theorem 1

Proof i) \Rightarrow ii). The proof consists in showing that the set of all matrices \mathbf{B} such that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ can be characterized using exclusively operations underlying axioms MixC, IEC, ISC and IPC. The result follows from the fact that if i) holds then it follows from Donaldson and Weymark (1998) that there exist partial orders that rank \mathbf{B} at most as dissimilar as \mathbf{A} and that these orders are consistent with the operations underlying axioms ISC, IEC, IPC and MixC. The zonotope inclusion criterion entails one of such partial orders (by Remark 5). The result must hold for any matrix \mathbf{A} . Using the same procedure and notation as in the part i) \Rightarrow ii) of the proof of Proposition 2, we can construct matrices \mathbf{X}^j , for $j = 1, \dots, m$ such that $\tilde{\mathbf{A}} \mathbf{X}^j \preceq \tilde{\mathbf{A}} \forall j$ for all orderings \preceq consistent with axioms MixC, IPC, IEC, ISC (note that the set of orderings consistent with axioms MixC is a subset of those consistent with

Strong-MixC by Remark 1). Moreover, note that for such matrices it is also verified that for any $\gamma \in \mathcal{V}_n$ there exists a $\theta \in \mathcal{V}_n$ such that $\tilde{\mathbf{A}}\mathbf{X}^j\gamma = \tilde{\mathbf{A}}\theta$ which yields $\tilde{\mathbf{A}}(\mathbf{X}^j\gamma - \theta) = \mathbf{0}_d \forall j$. Using the fact that $\mathbf{X}^j \in \mathcal{R}_n \forall j$ then $\mathbf{X}^j\gamma \in \mathcal{V}_n$, so that it is sufficient to set $\theta = \mathbf{X}^j\gamma$ for every $\gamma \in \mathcal{V}_n$ to guarantee that $\theta \in \mathcal{V}_n$. As argued in the proof of Lemma 2 the previous condition implies that $Z(\tilde{\mathbf{A}}\mathbf{X}^j) \subseteq Z(\tilde{\mathbf{A}}) \forall j$. Note that only some classes (i.e., columns) of $\tilde{\mathbf{A}}\mathbf{X}^j$ identify vertices of $Z(\tilde{\mathbf{A}})$. To see this, note that for any matrix $\mathbf{X}^j = \sum_{\mathbf{n} \in \mathcal{P}_n^j} \frac{1}{k!} \mathbf{\Pi}$ the class h of $\tilde{\mathbf{A}}\mathbf{X}^j$ has coordinates $\mathbf{v} := \sum_{\ell \in J_k} \frac{(k-1)!}{k!} \tilde{\mathbf{a}}_\ell$ for any $h \in J_k$, whereas class h' of $\tilde{\mathbf{A}}\mathbf{X}^j$ is equal to $\tilde{\mathbf{a}}_{h'}$ for all $h' \notin J_k$. This holds $\forall j, k$. Any vector \mathbf{v} is a vertex of the zonotope $Z(\tilde{\mathbf{A}})$, since $k\mathbf{v} = \tilde{\mathbf{A}}\gamma$ with $\gamma \in \mathcal{V}_n^{01}$ such that $\gamma_j = 1$ whenever $j \in J_k$ whereas $\gamma_j = 0$ for any $j \notin J_k$. From Proposition 3.1 in Dahl (1999), the zonotope set $Z(\tilde{\mathbf{A}})$ is the convex hull of these vertices. To complete the proof, consider now a matrix $\tilde{\mathbf{B}} \in \mathcal{M}_d$ satisfying condition (4), where $\mathbf{B}^j = \tilde{\mathbf{A}}\mathbf{X}^j, j = 1, \dots, m$. Such a matrix is ranked $\tilde{\mathbf{B}} \preceq \tilde{\mathbf{A}}$ by all dissimilarity ordering satisfying axioms MixC, IEC, IPC and ISC. For any such matrix, we have that $\forall \gamma \in \mathcal{V}_n^{01} \exists \theta^j \in \mathcal{V}_n^{01}, j = 1, \dots, m: \tilde{\mathbf{B}}\gamma \in \text{conv}\{\tilde{\mathbf{A}}\mathbf{X}^1\gamma^1, \dots, \tilde{\mathbf{A}}\mathbf{X}^m\gamma^m\}$. Any element of the *conv* set writes: $\sum_{j=1}^m \omega_j \tilde{\mathbf{A}}\mathbf{X}^j\gamma^j = \tilde{\mathbf{A}} \sum_{j=1}^m \omega_j \mathbf{X}^j\gamma^j = \tilde{\mathbf{A}} \sum_{j=1}^m \omega_j \theta^j = \tilde{\mathbf{A}}\theta$ with $\theta^j \in \mathcal{V}_n$ since $\mathbf{X}^j \in \mathcal{R}_n \forall j$ and also $\theta \in \mathcal{V}_n$ given that $\omega_j \in [0, 1], \sum_j \omega_j = 1$. It follows that $\forall \gamma \in \mathcal{V}_n^{01} \exists \theta \in \mathcal{V}_n: \tilde{\mathbf{B}}\gamma = \tilde{\mathbf{A}}\theta$, which is equivalent to $Z(\tilde{\mathbf{B}}) \subseteq Z(\tilde{\mathbf{A}})$ from Lemma 2. The fact that $Z(\tilde{\mathbf{B}}) = Z(\mathbf{B})$ and $Z(\tilde{\mathbf{A}}) = Z(\mathbf{A})$ gives ii).

In order to highlight the differences with the analogous steps in the proof of Proposition 2, note that, as shown there, each matrix \mathbf{X}^j is a basis for \mathcal{R}_n : every row-stochastic matrix is obtained as a convex combination (with weights as in axiom Strong-MixC) of matrices \mathbf{X}^j for $j = 1, \dots, m$. However, the weighting schemes underlying axiom Strong-MixC are only a subset of the admissible weights according to axioms MixC. In particular, these weights are not capable of generating the convex hull of all vertices of $Z(\tilde{\mathbf{A}})$ and, as a consequence, they are not sufficient to characterize the full set of matrices $\tilde{\mathbf{B}}$ such that $Z(\tilde{\mathbf{B}}) \subseteq Z(\tilde{\mathbf{A}})$, as opposed to the weights considered in MixC. Nonetheless, the weights underlying axiom Strong-MixC provide sufficient structure to characterize the set of matrices $\tilde{\mathbf{B}}$ such that $\tilde{\mathbf{B}} \preceq^R \tilde{\mathbf{A}}$, as illustrated in Proposition 2.

ii) \Rightarrow i). We show that the zonotope inclusion condition can always be rationalized by the existence of matrices that are ranked consistently by all dissimilarity orderings in i). Assume that ii) holds, and note that $Z(\mathbf{B}) \subseteq Z(\mathbf{A}) \Leftrightarrow Z(\tilde{\mathbf{B}}) \subseteq Z(\tilde{\mathbf{A}})$ where $\tilde{\mathbf{A}}, \tilde{\mathbf{B}} \in \mathcal{M}_d$ are $d \times n$ matrices obtained through split and permutation of classes and deletion of empty classes from \mathbf{A} and \mathbf{B} respectively, such that $\mathbf{1}'_d \tilde{\mathbf{A}} = \mathbf{1}'_d \tilde{\mathbf{B}} = \frac{d}{n} \mathbf{1}'_{n_2}$ as done in the first part of the proof of Proposition 2. Consider further matrices $\tilde{\mathbf{B}}^j \in \mathcal{M}_d$, for $j = 1, \dots, m$ with $m = n$, of size $d \times n$ defined as: $\tilde{\mathbf{B}}^j := (\tilde{\mathbf{b}}_j, \frac{1}{n-1} \mathbf{r}_j, \dots, \frac{1}{n-1} \mathbf{r}_j) \cdot \mathbf{\Pi}_{1,j}$ with $\mathbf{\Pi}_{1,j} \in \mathcal{P}_n$ being a permutation matrix permuting classes 1 and j . Denote vector $\mathbf{r}_j := \mathbf{1}_d - \tilde{\mathbf{b}}_j$ the “residual”, where we recall that $\tilde{\mathbf{b}}_j$ denotes column j of matrix $\tilde{\mathbf{B}}$. By construction, $Z(\tilde{\mathbf{B}}^j) \subseteq Z(\tilde{\mathbf{A}}) \forall j$. Considering that because of the definition of zonotope inclusion we have that $\tilde{\mathbf{b}}_j = \tilde{\mathbf{A}}\theta$, for $\theta \in \mathcal{V}_n$ then $\mathbf{r}_j = \tilde{\mathbf{A}}(\mathbf{1}_n - \theta)$. Notice that the vector θ is specific to a matrix $\tilde{\mathbf{B}}^j$. It follows that:

$$\begin{aligned}
 \tilde{\mathbf{B}}^j &:= (\tilde{\mathbf{A}}\theta, \frac{1}{n-1}\tilde{\mathbf{A}}(\mathbf{1}_n - \theta), \dots, \frac{1}{n-1}\tilde{\mathbf{A}}(\mathbf{1}_n - \theta))\mathbf{\Pi}_{1,j} \\
 &= \tilde{\mathbf{A}}(\theta, \frac{1}{n-1}(\mathbf{1}_n - \theta), \dots, \frac{1}{n-1}(\mathbf{1}_n - \theta))\mathbf{\Pi}_{1,j} \\
 &= \tilde{\mathbf{A}}\mathbf{X}^j
 \end{aligned}
 \tag{16}$$

with $\mathbf{X}^j \in \mathcal{R}_n$ for every $j = 1, \dots, m$. We conclude that condition ii) always implies that $\tilde{\mathbf{B}}^j \preceq^R \tilde{\mathbf{A}} \forall j$ for the n matrices $\tilde{\mathbf{B}}^j$ obtained as above. Now consider the set of weights w_j^k with the following features: $w_j^j = 1$ and $w_j^k = 0$ for any $j = 1, \dots, n$ and $k = 1, \dots, n$ such that $k \geq j$. By construction: $\tilde{\mathbf{b}}_k = \sum_j w_j^k \tilde{\mathbf{b}}_j^k$ for all $k = 1, \dots, n$. From Proposition 1, $\tilde{\mathbf{B}}^j \preceq \tilde{\mathbf{A}} \forall j$ for all orderings satisfying MC, ISC, IEC, IPC. These are also the orderings satisfying ISC, IEC, IPC and Strong-MixC from Proposition 2. By Remark 1, this set of orderings includes those that are in i) because those considered in the set satisfy the stronger version of the MixC axiom. To conclude, we should verify that matrix $\tilde{\mathbf{B}}$ is obtained through a weighting scheme that is consistent with condition (4). This is immediate, since for $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)' \in \mathcal{V}_n^{01}$ we have that $\tilde{\mathbf{B}}\boldsymbol{\gamma} = (\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_n)\boldsymbol{\gamma} = (\sum_j w_j^1 \tilde{\mathbf{b}}_j^1, \dots, \sum_j w_j^n \tilde{\mathbf{b}}_j^n)\boldsymbol{\gamma} = (\tilde{\mathbf{b}}_1^1, \dots, \tilde{\mathbf{b}}_n^n)\boldsymbol{\gamma} = (\tilde{\mathbf{B}}^1 \mathbf{i}_1, \dots, \tilde{\mathbf{B}}^n \mathbf{i}_n)\boldsymbol{\gamma} = (\tilde{\mathbf{B}}^1 \mathbf{i}_1 \gamma_1, \dots, \tilde{\mathbf{B}}^n \mathbf{i}_n \gamma_n)$ where we recall that \mathbf{i}_k is the k -th column of a $n \times n$ identity matrix and that the second equality follows from the choice of the weighting scheme. It is sufficient to set $\boldsymbol{\gamma}^j := \mathbf{i}_j \gamma_j \in \mathcal{V}_n^{01}$ to conclude that the condition $\forall \boldsymbol{\gamma} \in \mathcal{V}_n^{01} \exists \boldsymbol{\gamma}^j \in \mathcal{V}_n^{01}$ such that $\tilde{\mathbf{B}}\boldsymbol{\gamma} = (\tilde{\mathbf{B}}^1 \boldsymbol{\gamma}^1, \dots, \tilde{\mathbf{B}}^n \mathbf{i}_n \boldsymbol{\gamma}^n)$ satisfies condition (4).

As a result for the obtained matrix $\tilde{\mathbf{B}}$ we have that $\tilde{\mathbf{B}} \preceq \tilde{\mathbf{A}}$ for all orderings satisfying ISC, IEC, IPC and MixC. The same orderings rank $\tilde{\mathbf{A}} \sim \mathbf{A}$ and $\tilde{\mathbf{B}} \sim \mathbf{B}$, which implies i). □

Appendix A.9 Proof of Remark 6

Proof Recall the definition of the zonotope set inclusion, for any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, that is $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ if and only if $\forall \boldsymbol{\gamma} \in \mathcal{V}_{n_B} \exists \boldsymbol{\theta} \in \mathcal{V}_{n_A}$ such that $\mathbf{B}\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\theta}$.

Let $\mathbf{B} \preceq^R \mathbf{A}$. By Proposition 1, $\mathbf{B} = \mathbf{A}\mathbf{X}$ with $\mathbf{X} \in \mathcal{R}_{n_B, n_A}$, which after substituting in $\mathbf{B}\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\theta}$ yields $\mathbf{A}(\mathbf{X}\boldsymbol{\gamma} - \boldsymbol{\theta}) = \mathbf{0}_d$. Given that $\mathbf{X} \in \mathcal{R}_{n_B, n_A}$, this implies that $\mathbf{X}\boldsymbol{\gamma} \in \mathcal{V}_{n_A}$ for any $\boldsymbol{\gamma} \in \mathcal{V}_{n_B}$, it is therefore sufficient to set $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\gamma}$ for $\boldsymbol{\gamma} \in \mathcal{V}_{n_B}$ to guarantee that $\boldsymbol{\theta} \in \mathcal{V}_{n_A}$, which thus implies $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$. □

Appendix A.10 Proof of Remark 7

Proof Let $\tilde{\mathbf{A}}, \tilde{\mathbf{B}} \in \mathcal{M}_2$ denote two distribution matrices obtained respectively from $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$ using split and permutation of classes and deletion of empty classes, such that $Z(\mathbf{A}) = Z(\tilde{\mathbf{A}})$, $Z(\mathbf{B}) = Z(\tilde{\mathbf{B}})$, with $n_{\tilde{\mathbf{A}}} = n_{\tilde{\mathbf{B}}} = \tilde{n}$, $\sum_j \mathbf{a}_j = \sum_j \tilde{\mathbf{a}}_j = \sum_j \tilde{\mathbf{b}}_j = \sum_j \mathbf{b}_j$ and $\sum_i \tilde{a}_{ij} = \sum_i \tilde{b}_{ij} = \frac{1}{\tilde{n}} \forall j = 1, \dots, \tilde{n}$. Assume that $Z(\tilde{\mathbf{B}}) \subseteq Z(\tilde{\mathbf{A}})$ which, from Lemma 2, implies that $\forall \boldsymbol{\gamma} \in \mathcal{V}_{\tilde{n}}^{01}, \exists \boldsymbol{\theta} \in \mathcal{V}_{\tilde{n}} : \tilde{\mathbf{B}}\boldsymbol{\gamma} = \tilde{\mathbf{A}}\boldsymbol{\theta}$. The vector $\boldsymbol{\gamma}$ selects column vectors

from $\tilde{\mathbf{B}}$ and aggregates these vectors with elementwise summations. For every $i = 1, 2$ we can equivalently write the condition as:

$$\sum_{k=1}^h \tilde{b}_{ik} = \sum_{j=1}^h \tilde{a}_{ij} \theta_{jk}, \quad \forall h = 1, \dots, \tilde{n}, \tag{17}$$

with $\theta_{jk} \in [0, 1] \forall j, k$. It is necessary and sufficient that (17) holds for an arbitrary group $i = 1$ to guarantee that (17) also holds for $i = 2$, given that by construction $\tilde{a}_{2j} = \frac{2}{\tilde{n}} - \tilde{a}_{1j}$ and $\tilde{b}_{2j} = \frac{2}{\tilde{n}} - \tilde{b}_{1j} \forall j = 1, \dots, \tilde{n}$. After rearranging the terms in (17) in increasing order, so that $\tilde{b}_{1(k)} \leq \tilde{b}_{1(k+1)}$ and $\tilde{a}_{1(k)} \leq \tilde{a}_{1(k+1)}$, for $k = 1, \dots, \tilde{n} - 1$, and using the fact that $\theta_{jk} \in [0, 1]$ it follows that $\sum_{k=1}^h \tilde{b}_{1(k)} \geq \sum_{j=1}^h \tilde{a}_{1(j)}$, for any $h = 1, \dots, \tilde{n}$. From Marshall et al. (2011), if $d = 2$, this condition is equivalent to uniform majorization, i.e. $\exists \mathbf{X} \in \mathcal{D}_{\tilde{n}} : (\tilde{b}_{11}, \dots, \tilde{b}_{1\tilde{n}}) = (\tilde{a}_{11}, \dots, \tilde{a}_{1\tilde{n}})\mathbf{X}$. Hence $\tilde{\mathbf{B}} = \tilde{\mathbf{A}}\mathbf{X}$, since the same matrix \mathbf{X} guarantees that $(\tilde{b}_{21}, \dots, \tilde{b}_{2\tilde{n}}) = (\tilde{a}_{21}, \dots, \tilde{a}_{2\tilde{n}})\mathbf{X}$. Furthermore, notice that the indifference class of \leq^R is also characterized by the existence of row stochastic matrices: for any $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, $\mathbf{B} \sim \mathbf{A}$ for all orderings \leq satisfying axioms MC, ISC, IEC and IPC if and only if $\exists \mathbf{X} \in \mathcal{R}_{n_A, n_B}$ and $\exists \mathbf{X}' \in \mathcal{R}_{n_B, n_A}$ such that $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ and $\mathbf{A} = \mathbf{B} \cdot \mathbf{X}'$. We can hence write $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}\mathbf{Z}$ with $\mathbf{Z} \in \mathcal{R}_{n_A, \tilde{n}}$ and $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{Y}$ with $\mathbf{Y} \in \mathcal{R}_{\tilde{n}, n_B}$. Thus $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{Y} = \tilde{\mathbf{A}}\mathbf{X}\mathbf{Y} = \tilde{\mathbf{A}}\mathbf{Z}\mathbf{X}\mathbf{Y} = \tilde{\mathbf{A}}\Theta$ with $\Theta \in \mathcal{R}_{n_A, n_B}$ since $\mathcal{D}_{\tilde{n}} \subseteq \mathcal{R}_{\tilde{n}}$ and the set \mathcal{R}_n is closed with respect to matrix product. This concludes the proof for $d = 2$. When $d > 2$, there is no guarantee that the row stochastic matrix \mathbf{X} satisfying (17) for group i also verifies $\tilde{\mathbf{B}} = \tilde{\mathbf{A}}\mathbf{X}$. \square

Appendix A.11 Proof of Proposition 3

Proof From Proposition 1, i) is equivalent to $\mathbf{B} \leq^R \mathbf{A}$. From Lemma 1 in Appendix A.1, we have that $\mathbf{B} \leq^R \mathbf{A}$ if and only if

$$\sum_{j=1}^{n_B} g(b_{1j}, \dots, b_{dj}) \leq \sum_{j=1}^{n_A} g(a_{1j}, \dots, a_{dj}), \tag{18}$$

for $g : \mathbb{R}^d \rightarrow \mathbb{R}$ convex and homogeneous with $g(\mathbf{0}'_d) = 0$. Note that $\mathbf{B} \leq^R \mathbf{A}$ is equivalent to $(\mathbf{B}', \bar{\mathbf{b}})' \leq^R (\mathbf{A}', \bar{\mathbf{a}})'$, where $\bar{\mathbf{b}} = \mathbf{1}'_d \cdot \mathbf{B}$ and $\bar{\mathbf{a}} = \mathbf{1}'_d \cdot \mathbf{A}$, that is if both matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ are "expanded" by adding one more row whose elements are given by the sum of the elements of the associated column in the original matrix. Condition (18) hence rewrites $\sum_j g(\mathbf{b}'_j, \bar{b}_j) \leq \sum_j g(\mathbf{a}'_j, \bar{a}_j)$ with g defined on \mathbb{R}^{d+1} . Given that g is convex and homogeneous, then $g(\mathbf{a}'_j, \bar{a}_j) = \bar{a}_j g(\mathbf{a}'_j / \bar{a}_j, 1) = \bar{a}_j h(\mathbf{a}'_j / \bar{a}_j)$ where $h \in \mathcal{H}$, while for convenience empty classes receive weight $\bar{a} = 0$. Moreover, adding $|n_A - n_B|$ empty classes preserves the relation in (18). We have therefore obtained the index D_h . Thus, $D_h(\mathbf{B}) \leq D_h(\mathbf{A}) \forall h \in \mathcal{H}$ is equivalent to (18) and to condition ii). \square

Appendix A.12 Proof of Proposition 4

Proof Recall that from Theorem 1 for $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ we have that $\mathbf{B} \preceq \mathbf{A}$ for all orderings \preceq satisfying axioms MixC, ISC, IEC, IPC if and only if $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$.

Consider matrices $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}$ of dimension $d \times n$ obtained from \mathbf{A}, \mathbf{B} through split and permutation of classes and deletion of empty classes, so that $\mathbf{1}'_d \tilde{\mathbf{B}} = \mathbf{1}'_d \tilde{\mathbf{A}} = \frac{d}{n} \mathbf{1}'_n$. Then, by construction, $Z(\mathbf{B}) \subseteq Z(\mathbf{A}) \Leftrightarrow Z(\tilde{\mathbf{B}}) \subseteq Z(\tilde{\mathbf{A}})$. According to Lemma 3 the latter zonotope inclusion condition is equivalent to $n\mathbf{p}'\tilde{\mathbf{B}} = n\mathbf{p}'\tilde{\mathbf{A}}\Theta$, for $\Theta \in \mathcal{C}_n$ and for any $\mathbf{p} \in \mathbb{R}^d$. From Hardy et al. (1934), this condition is equivalent to $\sum_j \frac{1}{n} \phi(n\mathbf{p}\tilde{\mathbf{b}}_j) \leq \sum_j \frac{1}{n} \phi(n\mathbf{p}\tilde{\mathbf{a}}_j)$ for all convex functions $\phi : \mathbb{R} \rightarrow \mathbb{R}$. Consider a set of indices \mathcal{K}_j such that the classes of $\tilde{\mathbf{B}}$ are obtained from those of \mathbf{B} by split transformations. Thus for any class j in \mathbf{B} there is a non-empty set \mathcal{K}_j of associated classes in $\tilde{\mathbf{B}}$. Let $k \in \mathcal{K}_j$ denote a generic element of the set \mathcal{K}_j associated with the class k in \mathbf{B} and consider a related "splitting" weight α_k such that $\sum_{k \in \mathcal{K}_j} \alpha_k = 1$ for any $j = 1, \dots, n_B$, we can then write $\tilde{\mathbf{b}}_k = \alpha_k \mathbf{b}_j$, with by construction $n = \frac{1}{\bar{b}_k} = \frac{1}{\alpha_k \bar{b}_j}$ where $\bar{b}_j := \mathbf{1}'_d \mathbf{b}_j$ and similarly for \bar{b}_k . It follows that the argument of the evaluation function ϕ defined for matrix $\tilde{\mathbf{B}}$, becomes $\frac{1}{n} \phi(n\mathbf{p}\tilde{\mathbf{b}}_k) = \bar{b}_k \phi(\mathbf{p}\tilde{\mathbf{b}}_k / \bar{b}_k) = \alpha_k \bar{b}_j \phi(\alpha_k \mathbf{p}\mathbf{b}_j / (\alpha_k \bar{b}_j)) = \alpha_k \bar{b}_j \phi(\mathbf{p}\mathbf{b}_j / \bar{b}_j)$. It follows that $\sum_j \frac{1}{n} \phi(n\mathbf{p}\tilde{\mathbf{b}}_k) = \sum_j \sum_{k \in \mathcal{K}_j} \alpha_k \bar{b}_j \phi(\mathbf{p}\mathbf{b}_j / \bar{b}_j) = \sum_j \bar{b}_j \phi(\mathbf{p}\mathbf{b}_j / \bar{b}_j)$. An analogous sequence of transformations applies to $\sum_j \frac{1}{n} \phi(n\mathbf{p}\tilde{\mathbf{a}}_j)$ that leads to $\sum_j \bar{a}_j \phi(\mathbf{p}\mathbf{a}_j / \bar{a}_j)$, thereby yielding $D_{\mathbf{p}, \phi}(\mathbf{B}) \leq D_{\mathbf{p}, \phi}(\mathbf{A})$ for all $\mathbf{p} \in \mathbb{R}^d$ and for every ϕ convex that is equivalent to $Z(\tilde{\mathbf{B}}) \subseteq Z(\tilde{\mathbf{A}})$ that in turns is equivalent to $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ that is equivalent to unanimity for all orderings satisfying MixC, ISC, IEC, IPC. \square

Appendix A.13 Proofs of Corollary 2

Proof By Theorem 1, (6) is equivalent to $\tilde{\mathbf{B}} = \tilde{\mathbf{A}}\mathbf{X}$ for $\mathbf{X} \in \mathcal{R}_{n_A, n_B}$, which gives condition (i). Each entry in the first row of $\tilde{\mathbf{A}}$ is a constant equal to $1/n_A$, so it can be transformed by \mathbf{X} into the corresponding element in $\tilde{\mathbf{B}}$, equal to $1/n_B$, only by multiplying each single entry by n_A/n_B , thus (ii) should also hold. \square

Appendix A.14 Proofs of Corollary 3

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}' towards \mathbf{b}' involving a *T-transform* matrix $\mathbf{T}(\lambda, k, j)$, such that $\mathbf{b}' = \mathbf{a}' \cdot \mathbf{T}(\lambda, k, j)$, 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:

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

where $\lambda \in [0, 0.5]$ and $\mathbf{\Pi}_{n+1,k} \in \mathcal{P}_{n+1}$ is a $n + 1$ dimensional permutation matrix obtained from \mathbf{I}_{n+1} by 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 $\lambda' = \frac{1-2\lambda}{1-\lambda}$. □

Appendix A.15 Proof that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ from example (1)

We first show that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ for distribution matrices \mathbf{A}, \mathbf{B} in example (1). Note that after splitting equally column 1 of matrix \mathbf{B} into two columns, one obtains matrix $\tilde{\mathbf{B}} = (\frac{1}{2}\mathbf{b}_1, \frac{1}{2}\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ so that both matrices $\tilde{\mathbf{B}}, \mathbf{A}$ share the same distribution for one of the groups (the first row), which is uniform. The zonotopes of these matrices are three-dimensional objects, but the inclusion $Z(\mathbf{B}) = Z(\tilde{\mathbf{B}}) \subseteq Z(\mathbf{A})$ can be verified by fixing the dimension related to group 1 and then focusing on inclusion in terms of facets of the zonotope obtained by the intersection of each zonotope with the hyperplanes identified by levels of population of group 1 and parallel to the orthants of groups 2 and 3. Among these facets, only three are relevant: those corresponding to proportions of group 1 equal to $\frac{1}{4}, \frac{2}{4}$ and $\frac{3}{4}$. If inclusion of the facets of $Z(\mathbf{B})$ into those of $Z(\mathbf{A})$ is verified for each relevant share of group 1, then it is verified for any share of group 1.²⁴

The facets of $Z(\mathbf{B})$ (dark gray) and $Z(\mathbf{A})$ (light gray) at given proportions of group 1 are represented in Fig. 3. They are taken from figure 1(b), representing the three-dimensional zonotope of \mathbf{A} . The scaling of the axis eases the visual representation of the projections coordinates. Zonotope inclusion is granted for each projection (moving north-east implies larger shares of group 1 population), implying $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$.

Next, we show that every class of \mathbf{B} can be obtained through merge and split operations from the classes of \mathbf{A} and yet $\mathbf{B} \not\prec^R \mathbf{A}$. Consider the following row-stochastic matrices:

$$\mathbf{X}^1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix}; \quad \mathbf{X}^2 = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.5 & 0.25 \\ 0.5 & 0 & 0.5 \end{pmatrix}; \quad \mathbf{X}^3 = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.375 & 0.375 & 0.25 \\ 0.125 & 0.125 & 0.75 \\ 0.5 & 0.5 & 0 \end{pmatrix}.$$

Matrix \mathbf{X}^1 is such that column \mathbf{x}^1_1 displays a merge of classes 1 and 2 of matrix \mathbf{A} yielding \mathbf{b}_1 , whereas the remaining classes are merged and split uniformly into two classes. Every column \mathbf{x}^j_i is the result of split and merge operations that determine

²⁴ In fact, zonotopes are convex hulls of the underlying vertices, being defined by the sequential cumulation of the classes of the distribution matrices (see also Koshevoy 1995).

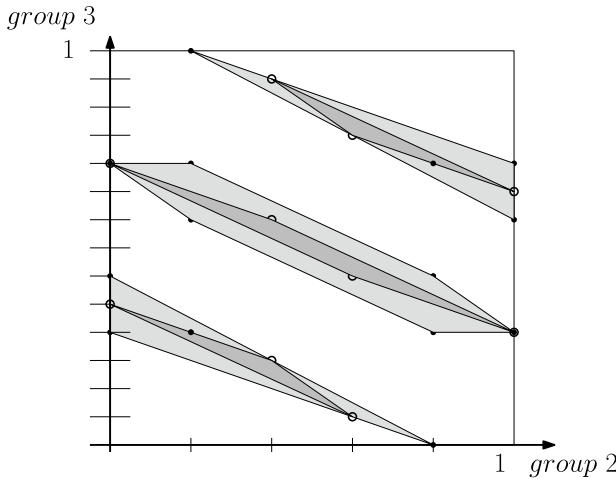


Fig. 3 Geometric proof that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$

class \mathbf{b}_j of the distribution matrix \mathbf{B} , whereas the remaining classes are merged and then split uniformly. The products $\mathbf{A}\mathbf{X}^j$ give the following matrices \mathbf{B}^j :

$$\mathbf{B}^1 := \mathbf{A}\mathbf{X}^1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{5}{7} & \frac{1}{7} & \frac{1}{7} \end{pmatrix}; \quad \mathbf{B}^2 := \mathbf{A}\mathbf{X}^2 = \begin{pmatrix} \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ 0.3 & 0.4 & 0.3 \\ \frac{11}{28} & \frac{3}{14} & \frac{11}{28} \end{pmatrix}; \quad \mathbf{B}^3 := \mathbf{A}\mathbf{X}^3 = \begin{pmatrix} \frac{3}{8} & \frac{3}{8} & \frac{1}{4} \\ 0.2 & 0.2 & 0.6 \\ \frac{13}{28} & \frac{13}{28} & \frac{1}{14} \end{pmatrix}.$$

It is clear that $\mathbf{B}^j \preceq^R \mathbf{A}$ which implies $Z(\mathbf{B}^j) \subseteq Z(\mathbf{A})$ (from Remark 6) for any $j = 1, 2, 3$. Now consider obtaining classes of matrix \mathbf{B} as a weighted average of classes of \mathbf{B}^j using the weights $w_j^k, j = 1, 2, 3$ and $k = 1, 2, 3$ such that $w_j^j = 1 \forall j$ and $w_j^k = 0$ when $k \neq j$. These weights are consistent with condition (4) (and hence with axiom MixC, see the second part of the proof of Theorem 1), but not with those implied by axiom Strong-MixC (they are not constant for given j). Matrix \mathbf{B} can be obtained from \mathbf{B}^j using $\mathbf{b}_k = \sum_j w_j^k \mathbf{b}_k^j = \mathbf{b}_k^k$ for any $k = 1, 2, 3 \forall k$. This implies $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ (from Lemma 5). Yet, using the weighting scheme we obtain $\mathbf{B} = \mathbf{A}\mathbf{X}$ with $\mathbf{x}_k = \sum_j w_j^k \mathbf{x}_k^j = \mathbf{x}_k^k \forall k$, but $\mathbf{X} \notin \mathcal{R}_{4,3}$ since there is a unique \mathbf{X} with non-negative entries that transforms \mathbf{A} into \mathbf{B} as in (3) that is not row-stochastic, thus $\mathbf{B} \not\preceq^R \mathbf{A}$. This counterexample shows that matrix majorization is not consistent with the axiom MixC.

Funding Open access funding provided by Università degli Studi di Verona within the CRUI-CARE Agreement. This paper forms part of the research projects *The Measurement of Ordinal and Multidimensional Inequalities* (Contract No. ANR-16-CE41-0005-01 - ORDINEQ) of the French National Agency for Research, the research project MOBILIFE (grant RBVR17KFHX) and PREOPP (grant RBVR19FSFA) of the University of Verona Basic Research scheme and the NORFACE/DIAL project *IMCHILD: The impact of childhood circumstances on individual outcomes over the life-course* (grant INTER/NORFACE/16/11333934/IMCHILD) of the Luxembourg National Research Fund (FNR), whose financial support is gratefully acknowledged.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alonso-Villar O, del Rio C (2010) Local versus overall segregation measures. *Math Soc Sci* 60:30–38
- Andreoli F, Zoli C (2020) From unidimensional to multidimensional inequality: a review. *METRON* 78:5–42
- Andreoli F, Havnes T, Lefranc A (2019) Robust inequality of opportunity comparisons: theory and application to early childhood policy evaluation. *Rev Econ Stat* 101(2):355–369. https://doi.org/10.1162/rest_a_00747
- Atkinson AB, Bourguignon F (1982) The comparison of multi-dimensioned distributions of economic status. *Rev Econ Stud* 49(2):183–201
- Blackwell D (1953) Equivalent comparisons of experiments. *Ann Math Stat* 24(2):265–272
- Blyth CR (1972) On Simpson's paradox and the sure-thing principle. *J Am Stat Assoc* 67(338):364–366
- Chakravarty SR, Silber J (2007) A generalized index of employment segregation. *Math Soc Sci* 53(2):185–195
- Dahl G (1999) Matrix majorization. *Linear Algebra Appl* 288:53–73
- Decancq K (2012) Elementary multivariate rearrangements and stochastic dominance on a Fréchet class. *J Econ Theory* 147(4):1450–1459
- Dekel E (1986) An axiomatic characterization of preferences under uncertainty: weakening the independence axiom. *J Econ Theory* 40(2):304–318
- Donaldson D, Weymark JA (1998) A quasi ordering is the intersection of orderings. *J Econ Theory* 78(2):382–387
- Duncan OD, Duncan B (1955) A methodological analysis of segregation indexes. *Am Soc Rev* 20(2):210–217
- Ebert U, Moyes P (2003) Equivalence scales reconsidered. *Econometrica* 71(1):319–343. <https://doi.org/10.1111/1468-0262.00397>
- Epstein LG, Tanny SM (1980) Increasing generalized correlation: a definition and some economic consequences. *Can J Econ* 13(1):16–34
- Flückinger Y, Silber J (1999) The measurement of segregation in the labour force. Physica-Verlag, Heidelberg
- Frankel DM, Volij O (2011) Measuring school segregation. *J Econ Theory* 146(1):1–38
- Gajdos T, Weymark JA (2005) Multidimensional generalized Gini indices. *Econ Theor* 26(3):471–496. <https://doi.org/10.1007/s00199-004-0529-x>
- Gini C (1914) Di una misura di dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche. *Atti del Regio Istituto Veneto di Scienze Lettere e Arti*, LXXIII
- Giovagnoli A, Marzioletti J, Wynn H (2009) Bivariate dependence orderings for unordered categorical variables. In: Pronzato L, Zhigljavsky A (eds) *Optimal design and related areas in optimization and statistics*, vol 28 of, optimization and its applications. Springer, New York
- Grant S, Kajii A, Polak B (1998) Intrinsic preference for information. *J Econ Theory* 83(2):233–259
- Hardy GH, Littlewood JE, Polya G (1934) *Inequalities*. Cambridge University Press, London
- Hasani A, Radjabalipour M (2007) On linear preservers of (right) matrix majorization. *Linear Algebra Appl* 423(2–3):255–261
- Hutchens RM (1991) Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations. *Math Soc Sci* 21(1):31–51
- Hutchens RM (2015) Symmetric measures of segregation, segregation curves and Blackwell's criterion. *Math Soc Sci* 73:63–68

- James DR, Taeuber KE (1985) Measures of segregation. *Sociol Methodol* 15:1–32
- Kolm S-C (1977) Multidimensional egalitarianisms. *Q J Econ* 91(1):1–13
- Koshevoy G (1995) Multivariate Lorenz majorization. *Soc Choice Welf* 12:93–102. <https://doi.org/10.1007/BF00182196>
- Koshevoy G, Mosler K (1996) The Lorenz zonoid of a multivariate distribution. *J Am Stat Assoc* 91(434):873–882
- Lasso de la Vega C, Volij O (2014) Segregation, informativeness and Lorenz dominance. *Soc Choice Welf* 43(3):547–564. <https://doi.org/10.1007/s00355-014-0801-3>
- Marshall AW, Olkin I, Arnold BC (2011) *Inequalities: theory of majorization and its applications*. Springer, Berlin
- Martínez Pería FD, Massey PG, Silvestre LE (2005) Weak matrix majorization. *Linear Algebra Appl* 403:343–368
- McMullen P (1971) On zonotopes. *Trans Am Math Soc* 159:91–109
- Mosler K (2012) *Multivariate dispersion, central regions, and depth: the lift zonoid approach*, vol 165. Springer Science and Business Media, Berlin
- Moulin H (2016) Entropy, desegregation, and proportional rationing. *J Econ Theory* 162:1–20
- Nehring K, Puppe C (2002) A theory of diversity. *Econometrica* 70(3):1155–1198
- Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoret Popul Biol* 21(1):24–43
- Reardon SF (2009) Measures of ordinal segregation. In: Yves JS, Sean FR (eds) *Occupational and residential segregation research on economic inequality*, vol 17. Emerald Group Publishing Limited, Bingley, pp 129–155
- Reardon SF, Firebaugh G (2002) Measures of multigroup segregation. *Sociol Methodol* 32:33–67
- Roemer JE, Trannoy A (2016) Equality of opportunity: theory and measurement. *J Econ Lit* 54(4):1288–1332
- Torgersen E (1992) Comparison of statistical experiments, vol 36 of *encyclopedia of mathematics and its applications*. Cambridge University Press, Cambridge
- Tsui K-Y (1995) Multidimensional generalizations of the relative and absolute inequality indices: the Atkinson-Kolm-Sen approach. *J Econ Theory* 67(1):251–265
- Weymark JA (2006) The normative approach to the measurement of multidimensional inequality. *Inequality and economic integration*. Routledge, London, p 26
- Ziegler G (1995) *Lectures on polytopes*, number 152 in *graduate texts in mathematics*. Springer-Verlag, New York

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.