Working Paper Series
Department of Economics
University of Verona

Measuring the interaction dimension of segregation: the Gini-Exposure index

Francesco Andreoli, Claudio Zoli

WP Number: 30 December 2015

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

Measuring the interaction dimension of segregation: the Gini-Exposure index*

Francesco Andreoli,
Claudio Zoli†

Luxembourg Institute of Socio-Economic Research and University of Verona

December 2015

*This paper benefitted from the comments and suggestions by Eugenio Peluso, Alain Trannoy, Dirk Van de gaer and participants to the IT10 Winter School (Canazei 2015) and PET Conference (Luxembourg 2015). The financial support from the Fonds National de la Recherche Luxembourg (AFR grant 5932132) is kindly acknowledged by Francesco Andreoli.

†**Contact details**: F. Andreoli is at Luxembourg Institute of Socio-Economic Research (LISER), 11 Porte des Sciences, L4366 Esch-sur-Alzette, Luxembourg. E-mail: `francesco.andreoli@liser.lu`. C. Zoli is at University of Verona, Department of Economics, Via dell'Artigliere 19, 37129 Verona, Italy. E-mail: `claudio.zoli@univr.it`.

Abstract

We study the heterogeneity of social interaction profiles among individuals and define the extent of the interaction dimension of segregation. An interaction profile quantifies the probabilities that one individual has to interact with different social groups. It can be inferred, for instance, from observation of social ties through networks data. Heterogeneity is minimal if everybody exhibit the same profile, and is maximal if everybody interacts with only one group. All the in-between configurations can be ordered on the bases of an intuitive principle based on operation that generate mixtures of interaction profiles. We proposes a characterization of the Gini-exposure index to assess heterogeneity in interaction patters in a society. One key advantage of this index is that overall heterogeneity can be decomposed into the segregation experienced by every individual with respect to other people in his own group (isolation) or in other groups (exposure). An preliminary empirical investigation of interaction patterns of natives and immigrants across Italian municipalities reveals connections and differences with other exposure measures.

Keywords: Interaction, segregation, dissimilarity, Gini index.

JEL Codes: J71, D31, D63, C16

1 Introduction

In their seminal analysis of segregation measures, Massey and Denton (1988) define five dimensions of analysis for residential segregation: evenness, exposure, clustering, centralization and concentration. The measurement of these phenomena requires to partition the population into groups and to know the distribution of these groups across *organizational units*, such as neighborhoods (Reardon and O'Sullivan 2004, Cutler and Glaeser 1997), school assignment (Frankel and Volij 2011, Echenique, Fryer and Kaufman 2006) or job types (Flückiger and Silber 1999, Hutchens 1991, Hutchens 2004). We focus on multi-group measures of segregation in the exposure dimension that can be used to assess *segregation in networks*.

This paper is interested in the distributional information that can be retained from a network, and not on the network's structure itself. We consider *interaction profiles*, that correspond to vectors of probabilities that every unit in a network has to interact with each of the groups that compose the society. We contribute to the literature on segregation measurement by proposing a multi-group index of segregation, the *Gini Exposure index*, which measures segregation in a network as a form of inequality in the distribution of interaction profiles.

Following Massey and Denton (1988), exposure measures should capture the differences across groups in the likelihood that any randomly selected individual from one of these groups interacts with a person/unit from his own group or from another group. Segregation is zero when the chances that any two randomly selected individuals interact are made independent on their respective groups of origin. On the contrary, segregation is maximized whenever every individual interactions are limited to the members of the same group.

Segregation measurement (Massey and Denton 1988, Reardon and Firebaugh 2002, Reardon and O'Sullivan 2004, Frankel and Volij 2011) has mainly focused on the rankings produced by segregation indices for populations partitioned into two or many groups. None of these indices, however, has been designed to deal with problems of segregation that use individual level data and the axiomatization of these indices, where it exists (see for instance Hutchens 1991, Flückiger and Silber 1999, Reardon and O'Sullivan 2004, Frankel and Volij 2011), cannot be meaningfully adapted to capture segregation patterns across individuals interaction profiles.

We fill this gap by proposing a framework to study segregation at individual level, conceptualized as a form of inequality in interaction profiles. We provide an axiomatic characterization of the Gini Exposure index that generalizes to the multi-group case the traditional Gini index of segregation. The index can be interpreted as the Gini volume index discussed in the multivariate inequality measurement literature (Koshevoy and Mosler 1996, Koshevoy and Mosler 1997, Arnold 2005). We also provide a decomposition result illustrating how the Gini Exposure index can be used to keep track of changes in group or individual specific patterns of segregation within the network.

The axiomatic characterization of the Gini Exposure index is mainly based on operations defined on interaction profiles that preserve or decrease segregation. Our analysis is grounded on a simple principle: when the number of units equals the number of the groups, if a portion of a unit interaction profile is merged with another unit, this mixture operation should not increase the segregation, and indeed should reduce it in proportion of the quota of the initial unit that is merged.

The following example clarifies this point. Consider, for instance, a large population that can be partitioned into two group of equal size, the "Reds" and "Greens", and interaction profiles can be inferred from network data. If every individual interacts with half of the remaining individuals, the degree of segregation depends exclusively on how different types of individuals interact among them. Two possible configurations are of particular interest. In the first configuration, each individual interacts with half of the Reds and half of the Greens. In this case the population is made of all homogeneous units that exhibit the same interaction profile, so there is no segregation. In a second configuration, every individual of the Reds interacts with all the Reds and exclusively with them, and analogously for the Greens. In this case we can consider the population as composed by two units that collect all the individuals that interact with a specific groups. Admittedly this is an highly segregated distribution.

If a proportion $1 - \alpha$ of the unit of Reds, is joining the unit of Greens and shares proportionally its interaction links then segregation is reduced in the proportion $1 - \alpha$. In fact as α tends to 1 the overall segregation should be eliminated because all the individuals will share the same average interaction profile. Our main result will show that this property will play a crucial role for the characterization of the Gini Exposure index for a large class of distribution matrices representing interactions profiles.

Alternative indicators have been proposed and adopted in the literature to measure the exposure dimension of segregation (see Hutchens 1991, Silber

1989, Flückiger and Silber 1999, Reardon and Firebaugh 2002). An extensive qualitative comparisons of these indicators with the Gini Exposure index on the base of the properties that they satisfy is not possible. We construct a quantitative analysis to recover empirical correlations in the rankings produce by different indices of exposure proposed in the literature and the Gini Exposure index. The closer is this correlation to zero, the more likely it is that the two indices capture very different segregation patterns underlying the data.

We make use of Italian data by ISTAT to study the degree of spatial segregation of immigrant groups across municipalities in Italy. We use a spatial model to identify interaction probabilities across Italian municipalities (nearly 8400), for each of the Italian provinces separately (101 provinces are considered in this study) in an interval of eight years (from 2003 to 2010). Our main assumption is that the chances for two individuals to interact decrease with the spatial distance between the area where the two individuals reside. We consider segregation among three groups: the groups of immigrants coming from low HDI and high HDI countries and the natives group.[1]

The empirical analysis reveals two broad categories of indicators: the indicators measuring the overall dissimilarity in interaction profiles and the entropy indicators, measuring how far profiles are from their average. The Gini Exposure index is mostly rank correlated with the dissimilarity-type indicators, and this correlation is fairly robust to the demographic variability of the data.

2 Notation

In this paper, we consider the problem of ranking *configurations* $A, B \in \mathcal{C}(G)$ according to the level of segregation in the exposure dimensions that they exhibit.

Definition 1 (Configuration) *A configuration $A \in \mathcal{C}(G)$ is a triplet*

$$\Big[\mathcal{N}(A),\ \mathcal{G},\ ((\pi_{gi}(A))_{g\in\mathcal{G}}, \xi_i(A))_{i\in\mathcal{N}(A)}\Big]$$

where $\mathcal{N}(A)$ is a finite, nonempty set of units of cardinality $N(A)$, $\mathcal{G}$ is a finite, nonempty set of G population groups, with variable demographic size

[1] In this setting, we treat municipalities as the basic units of our analysis.

denoted by $N_g(A)$. For each unit $i \in \mathcal{N}(A)$ and group $g \in \mathcal{G}$, the variable $\pi_{gi} \in [0,1]$ denotes the probability that i interacts with a randomly selected individual from group g. Unit i's demographic weight is denoted by $\xi_i(A)$, with $\sum_{i \in \mathcal{N}(A)} \xi_i(A) = 1$.[2]

To avoid cumbersome notation, references to the configuration A are dropped in what follows, unless disambiguation is needed. Thus, we denoted π_{ig}, N, N_g, $\mathbf{N}$ and ξ_i for configuration $A \in \mathcal{C}(G)$.

A configuration can be constructed, for instance, from empirical observation of the social connections between individuals, or from aggregate statistics of expected interaction patters. For a configuration $A \in \mathcal{C}(G)$, the interaction profile of $i \in \mathcal{N}$ is a column vector:

$$\boldsymbol{\pi}_{.i} := (\pi_{1i}, \ldots, \pi_{Gi})^t \in [0,1]^G,$$

such that $\sum_{g \in \mathcal{G}} \pi_{gi} = 1$ for any $i \in \mathcal{N}$. Hence, $\boldsymbol{\pi}_{.i}$ represents the social ties of unit i in terms of the probabilities that the individuals associated with this unit have to interact with members of each of the groups in $\mathcal{G}$. The $G \times N$ *interaction matrix* $\boldsymbol{\pi}$ represents a collection of the N interaction profiles (by column). The rows of the interaction matrix are denoted *group profiles* and are indicated with row vectors $\boldsymbol{\pi}_{g.} := (\pi_{g1}, \ldots, \pi_{gN}) \in [0,1]^N$.

The *expected interaction profile* associated with group g is the expected probability that a randomly drawn individual interacts with group g:

$$\pi_g^e(A) = \sum_{i \in \mathcal{N}(A)} \xi_i(A)\, \pi_{gi}(A).$$

Again, we write π_g^e in shorthand notation. For configuration A, we make use of expected interaction profiles to normalize the entries of the interaction matrix $\boldsymbol{\pi}$. This leads to define a $G \times N$ interaction matrix $\mathbf{A}$ (always denoted with boldface letters) such that $\mathbf{A} := (\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots \mathbf{a}_N)$ where $a_{gi} := \frac{\pi_{gi}(A)}{\pi_g^e(A)}$.

3 The Gini Exposure index

3.1 The index

The *Gini inequality* index of a univariate income distribution, represented by the N-dimensional vector $\mathbf{x}$, is defined as the average income gap between

[2] One particular case is the uniform weighting scheme, where $\xi_i(A) = 1/N(A)$ for all $i \in \mathcal{N}(A)$.

any pair of realizations in the income distribution $\mathbf{x}$, scaled by the overall average income:

$$G(\mathbf{x}) := \frac{1}{2N^2\left(\sum_i x_i/N\right)} \sum_{i=1}^{N} \sum_{j=1}^{N} |x_i - x_j|.$$

Alternatively, the Gini index can be related to the Lorenz curve: it is equal to twice the area between the Lorenz curve and the diagonal, representing the equal distribution.[3] As illustrated by Shephard (1974), the overall area delimited by the Lorenz curve can be represented as the sum of the areas spanned by every pair of vectors $(x_i, 1)$ and $(x_j, 1)$, corresponding to the determinant of a 2×2 matrix formed by these vectors. The gap $|x_i - x_j|$ corresponds, in fact, to the determinant of these matrices. It follows that the Gini inequality index rewrites:[4]

$$G(\mathbf{x}) := \frac{1}{2} \sum_{\forall \{i,j\} \subseteq \{1,\ldots,N\}} \frac{1}{N}\frac{1}{N} \left| \det \begin{pmatrix} x_i/\left(\sum_i x_i/N\right) & x_j/\left(\sum_i x_i/N\right) \\ 1 & 1 \end{pmatrix} \right|$$

In practice, the Gini inequality index can be conceptualized as a weighted average of the dissimilarity between the incomes of pairs of units and the two units' weights. The function measuring the intensity of this dissimilarity is the determinant, while the weights corresponds to the probability of drawing the pair of units i and j from the sample. Since every pair of incomes can be compared twice, the index must be standardized by two, so that its maximum is equal to one.

A similar logic can be adapted to the measurement of the degree of dissimilarity in interaction profiles, where income realizations have to be replaced by probabilities of interaction. Segregation assessments boil down to check

[3] The Lorenz Zonotope of distribution is defined as the area between the Lorenz curve and its *dual*. It can be written as a Minkowski sum of line segments, hence its area equals the sum of the areas spanned by each pair of bi-dimensional vectors $\left(\frac{x_i}{\sum x_i}, \frac{1}{N}\right)$ and $\left(\frac{x_j}{\sum x_i}, \frac{1}{N}\right)$, for all i, j. This area coincides with a parallelogram and it corresponds to a measure of inequality between incomes shares received by two individuals i, j equally weighted $\frac{1}{N}$ in the population.

[4] The terms $\frac{1}{N^2 \sum_i x_i/N}$ disappears at it is incorporated in the determinant calculation. The comparison is now expressed in relative, rather than absolute, incomes. Moreover, the determinant is a measure of linear dependence, and therefore similarity, between oriented vectors.

how much dissimilar is each group's interaction probability deviation from the mean from the value 1. A configuration exhibiting no segregation corresponds to the case in which i's interaction profile with group g is such that $\pi_{gi} = \pi_g^e$, for every $g \in \mathcal{G}$.

An obvious extension of the Gini inequality index is the *expected Gini* (EG) segregation index analyzed in Flückiger and Silber (1999) and Alonso-Villar and del Rio (2010). The EG index is an average of local Gini indices G_g, weighted by groups size:

$$EG(A) \ := \ \sum_{g \in G} s_g \, G_g(A),$$

where $s_g = \frac{N_g}{N}$ is the share of individuals in the network associated to group g. Each local Gini index is meant to capture the inequality in the distribution of interaction probability with a group, say g, across the population:

$$G_g(A) \ := \ \frac{1}{2} \sum_{\forall \{i_1, i_2\} \subseteq \mathcal{N}(A)} \xi_{i_1}(A) \xi_{i_2}(A) \left| \det \begin{pmatrix} \frac{\pi_{gi_1}(A)}{\pi_g^e(A)]} & \frac{\pi_{gi_2}(A)}{\pi_g^e(A)]} \\ 1 & 1 \end{pmatrix} \right|.$$

The expected Gini index assumes that evaluations of segregation can be separated across dimensions. This strong assumption leaves aside concerns about the composition of the interaction profiles. To overcome these limitations, we propose a multi-group extension of the local Gini index presented above, denoted the *Gini Exposure* index of segregation.

The Gini Exposure index of segregation, $G_E : \mathcal{C}(G) \to [0, 1]$ captures the dispersion in the normalized interaction profiles across units in the same configuration. The index is a weighted mean of a measure of dissimilarity between G-tuples of interaction profiles, as captured by the determinant of a square $G \times G$ matrix. The weight attached to each G-tuple corresponds to its probability of being observed. The index is standardized by $G!$, the overall number of possible G-tuples, so that the index maximum is equal to one.

Definition 2 (The Gini Exposure segregation index)

$$G_E(A) := \frac{1}{G!} \sum_{\forall \{i_1, \ldots, i_G\} \subseteq \mathcal{N}(A)} \xi_{i_1}(A) \cdot \ldots \cdot \xi_{i_G}(A) \left| \det \begin{pmatrix} \mathbf{a}_{i_1} & \ldots & \mathbf{a}_{i_G} \end{pmatrix} \right|.$$

3.2 A geometric illustration of the index

An equivalent way of assessing heterogeneity in interaction profiles consists in looking at the likelihood that any randomly chosen individual from group g interacts with individual i, given the original information about the distribution of interaction profiles across the population. For configuration A, define the *interaction likelihood* $\mathcal{L}^A_{gi} \in [0,1]$ as this probability. The sequence of probabilities $\mathcal{L}^A_{g1}, \ldots, \mathcal{L}^A_{gN}$ defines a distribution of interaction likelihoods of group g with all the units in the distribution, hence satisfying $\sum_{i\in\mathcal{N}} \mathcal{L}^A_{gi} = 1$ for every $g \in \mathcal{G}$. The interaction likelihood is tied to interaction profiles and individual weights through the Bayes' rule:

$$\mathcal{L}^A_{gi} \; := \; a_{gi}\xi_i \; = \; \frac{\pi_{gi}\xi_i}{\pi^e_g}.$$

Heterogeneity in interaction profiles always implies that a form of dissimilarity between interaction likelihoods prevails. When all interaction profiles coincide, then $a_{gi} = 1$ and $\mathcal{L}^A_{gi} = \xi_i$ for any i and g, meaning that the interaction likelihoods coincide across groups. This does not necessary imply, however, that the interaction likelihoods are constant across individuals. Conversely, when each individual interacts with exactly one group, say g, then the knowledge of the group allows to infer with certainty the individuals that will interact with it, because $\mathcal{L}^A_{gi} > \mathcal{L}^A_{g'i} = 0$ for all $g' \neq g$. All in-between situations display some form of dissimilarity between the rows of the interaction likelihood matrix $\mathcal{L}^A$ associated with configuration A and defined as:

$$\mathcal{L}^A := (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{N(A)}) \; = \; \begin{pmatrix} \mathcal{L}^A_{11} & \ldots & \mathcal{L}^A_{1N(A)} \\ \vdots & & \vdots \\ \mathcal{L}^A_{G1} & \ldots & \mathcal{L}^A_{GN(A)} \end{pmatrix},$$

where $\mathcal{L}^A$ is a *row stochastic matrix* (i.e. the entries add up to one by row, but not necessarily by column) of the type analyzed in Andreoli and Zoli (2014).

Andreoli and Zoli show that the dissimilarity between the rows of a $G \times N$ stochastic matrix (depicting sets of G discrete probabilities distributions defines over n classes of realizations) can be visually represented through the *Zonotope set* Z of the interaction likelyhood matrix $\mathcal{L}^A$, denoted $Z(\mathcal{L}^A)$. The Zonotope is a centrally symmetric polytope in the G-dimensional space representing the Minkowski sum of the matrix's columns (see Shephard 1974). More formally, it is defined as:

$$Z(\mathcal{L}^A) := \left\{ \mathbf{z} := (z_1, \ldots, z_G) : \ \mathbf{z} = \sum_{i=1}^{N(A)} \theta_i \cdot \boldsymbol{\ell}_i^A, \ \theta_i \in [0,1] \ \ \forall i \in \mathcal{N}(A) \right\},$$

where the units' weights $\xi_i \in [0,1]$ are such that $\sum_i \xi_i = 1$. A Zonotope can be seen as a multi-group extension of the *segregation curve*[5], where actual distributions of groups across organizational units are replaced by the interaction likelihoods these groups have with individuals. Since the Zonotope represents the extent od dissimilarity across interaction likelihoods, the order of distribution matrices produced by Zonotopes inclusion is always consistent with decreasing segregation.

Various works in linear algebra have studied the properties of the volume of the Zonotopes (McMullen 1971, Shephard 1974). It is shown, in particular, that the volume of any Zonotope of a $G \times N$ matrix can be written as the sum of the volumes of the Zonotope sets generated by every $G \times G$ (thus square) matrix obtained from the original one by considering distinct G-tuples of its columns. The volume of a square matrix is the absolute value of its determinant, as already noted in Koshevoy and Mosler (1997) and Arnold (2005). In our case, since the reference matrix is $\mathcal{L}^A$, the volume of $Z(\mathcal{L}^A)$ is the Gini-Exposure index of segregation. This immediately bears the following implication.

Remark 1 *For any $A, B \in \mathcal{C}(G)$, $Z(\mathcal{L}^B) \subseteq Z(\mathcal{L}^A) \Rightarrow G_E(B) \leq G_E(A)$.*

This remark is important for two reasons. First, because it shows that the analysis of the exposure dimension of segregation is associated with the analysis of dissimilarity between distributions, in this case consisting in interaction likelihoods of interactions. It follows that the exposure dimension of segregation can be studied by making use of methods developed in the context of dissimilarity analysis (Andreoli and Zoli 2014). The robust dissimilarity test based on Zonotopes inclusion, for instance, defines sufficient conditions for decreasing dissimilarity as picked up by the Gini-Exposure index.

Second, the remark provides the setting that can be used to study the decomposition properties of the Gini-Exposure index, discussed hereafter.

[5]Segregation curves were introduced by (Duncan and Duncan 1955) and further studied by (Hutchens 1991) and generalized by (Carrington and Troske 1997) and (Silber 1989).

3.3 Decomposition properties

In many instances, one would like to assess the degree of exposure as experienced only by some subgroups of the population. This can be done by computing the share of the overall segregation that can be attributed to each group $g \in \mathcal{G}$ through a suitable decomposition of the overall Gini Exposure index.

Similarly to the traditional Gini inequality index, the Gini Exposure index can be decomposed into a weighted average of the degree of segregation experienced by each subgroup, captured by a group specific Gini index, and an overlapping term. This linear decomposition allows to study separately the dynamics of segregation for the members of each group.

For a configuration A, consider a partition of the overall population in groups, where we use $\mathcal{N}_g(A) \subseteq \mathcal{N}(A)$ to identify the set of individuals belonging to a group $g \in \mathcal{G}$. The construction of the problem makes clear that within the same unit i, there are possibly individuals belonging to each of the groups in $\mathcal{G}$. To each of these individuals, we associate the same interaction profile irrespectively of their group of origin. For each group g, the Gini Exposure index $G_E(A|g)$ measures the overall degree of segregation as experienced exclusively by members of group g in the configuration. This index is defined as:

$$G_E(A|g) := \frac{1}{G!} \sum_{\forall \{i_1,\ldots,i_G\} \subseteq \mathcal{N}_g(A)} \xi_{i_1}(A|g) \cdots \xi_{i_G}(A|g) \left| \det \begin{pmatrix} \frac{\pi_{1i_1}(A)}{\pi_1^e(A|g)} & \cdots & \frac{\pi_{1i_G}(A)}{\pi_1^e(A|g)} \\ \vdots & & \vdots \\ \frac{\pi_{Gi_1}(A)}{\pi_G^e(A|g)} & \cdots & \frac{\pi_{Gi_G}(A)}{\pi_G^e(A|g)} \end{pmatrix} \right|.$$

where $\xi_i(A|g) = \frac{\xi_i(A)}{\sum_{i \in \mathcal{N}_g(A)} \xi_i(A)}$ is the relative weight of unit i as measured by all individual experiencing the interaction profile $\boldsymbol{\pi}_{.i}$ and belonging to group g, and $\pi_m^e(A|g) = \sum_{i \in \mathcal{N}_g(A)} \xi_i(A|g)\pi_{mi}(A)$ is the expected probability of interaction with group m for an individual in group g.

Note that the multi-group $G_E(A|g)$ index is logically different from the single group index $G_g(A)$ defined above. In fact, it performs multi-group segregation comparisons by assessing the inequalities in the interaction profiles involving only individuals in group g, rather than assessing how much the chances of interacting with g are unequally distributed in the population as a whole.

The overlapping set, denoted $\mathcal{O}$, gathers all the possible G-tuple of individuals, with at least two individuals coming from different groups. It is given by:

$$\mathcal{O} \ := \ \{\{i_1, \dots, i_G\} \subseteq \mathcal{N}(A) : \ \nexists \{i_1, \dots, i_G\} \subseteq \mathcal{N}_g(A) \text{ for any } \ g \in \mathcal{G}\}\,.$$

Based on this notation, we are now able to show an additive decomposition of the multi-group Gini Exposure index in a within groups and an overlapping component, to obtain a decomposition of the Gini-Exposure index that is analogous to that of Ebert (2010), developed in the context of income inequality analysis for univariate distributions.

Proposition 1 *The Exposure Gini index can be decomposed as follows:*

$$G_E(A) \ = \ \left(\sum_{g \in \mathcal{G}} \alpha_g\right) \sum_{g \in \mathcal{G}} \beta_g \ G_E(A|g) \ + \ G_E(A|\mathcal{O}),$$

where $\beta_g = \frac{\alpha_g}{\sum_{g \in \mathcal{G}} \alpha_g}$ *and* $\alpha_g = (\sum_{i \in \mathcal{N}_g(A)} \xi_i(A))^G \prod_{m \in \mathcal{G}} \frac{\pi^e_m(A|g)}{\pi^e_m(A)}$.

Proof. See Appendix B.1. ■

If there are no systematic differences between groups in the expected interaction profiles, even though there exist within group variability, then $\pi^e_m(A|g) = \pi^e_m(A)$ for all groups m is expected to hold also across all groups $g \in \mathcal{G}$. In this case the weighting scheme β_g would depend only on groups densities. Moreover, if one compares allocations with little or no variability in groups compositions, the unique sources of variation for the Gini index are given either by the variations in the conditional segregation captured by $G_E(A|g)$ or by changes in the degree of overlapping. Otherwise, differences in the structural composition of the groups populations may play a relevant role in determining the overall degree of segregation.

3.4 An illustrative example

Consider an allocation A with a population of 20 individuals, partitioned in three non-overlapping groups $\mathcal{G} = \{g_1, g_2, g_3\}$. Out of the 20 individuals, 10 belong to group g_1, 5 belong to group g_2 and the remaining 5 are of group g_3, such that $\mathcal{N}_{g_1}(A) = \{1, \dots, 10\}$, $\mathcal{N}_{g_2}(A) = \{11, \dots, 15\}$ and $\mathcal{N}_{g_3}(A) = \{16, \dots, 20\}$.[6]

[6] The elements of the three sets represent individuals of the population.

We consider two different frameworks. In the first, we analyze a bivariate model of segregation, based on the interaction profiles with groups g_1 and g_2, as experienced by the whole population. In Figure 1 we draw the related Segregation Zonotope, we identify the Gini Exposure index and we depict a graphical representation of the index decomposition. The second example extends the analysis to the multi-groups (three groups) case.

Consider a simplified setting where there are only three possible interaction profiles with groups g_1 and g_2, namely $\boldsymbol{\pi}(A)$, $\boldsymbol{\pi}'(A)$ and $\boldsymbol{\pi}''(A)$ and defined as follows:

$$\boldsymbol{\pi}(A) = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}; \quad \boldsymbol{\pi}'(A) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \quad \boldsymbol{\pi}''(A) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The third interaction profile, $\boldsymbol{\pi}''(A)$, is such that all the individuals allocated with that profile do not have any chance to interact with group g_2. The other profiles can be interpreted in a similar way.

In the following table we summarize the distribution of individuals across groups *and* interaction profiles. For instance, individual $i = 3$ in group g_1 is allocated with interaction profile $\boldsymbol{\pi}(A)$, while individual $i = 15$ is the unique individual in group g_2 allocated with profile $\boldsymbol{\pi}''(A)$.

	$\boldsymbol{\pi}(A)$	$\boldsymbol{\pi}'(A)$	$\boldsymbol{\pi}''(A)$
g_1	$\{1,2,3\}$	$\{4,5,6\}$	$\{7,8,9,10\}$
g_2	$\{11,12\}$	$\{13,14\}$	$\{15\}$
g_3	$\{16,17,18\}$	$\{19,20\}$	$\emptyset$

Assuming a uniform weighting scheme ($\xi_i(A) = \frac{1}{20}$), one can construct the expected interaction profiles with the two groups: $\pi^e_{g_1}(A) = \frac{8}{20}0.5 + \frac{7}{20}0 + \frac{5}{20}1 = 0.45$ and $\pi^e_{g_2}(A) = 0.55$. The distribution matrix obtained from these data defines the underlying information that is necessary to construct the Segregation Zonotope in Figure 1(a) (the Segregation Zonotope is defined by the solid contour of the figure). The Gini Exposure index corresponds to its area, and it amounts to $G_E(A) = 0.8383$.

Within this example, it is possible to provide a graphical representation of the decomposition of the Gini Exposure index into a weighted sum of Gini Exposure indices, each measuring the overall inequality in interaction profiles between individuals of the same group. In Figure 1(a) we represent three distinct areas (denoted by different scales of grey), that correspond to the contributions of each group to the overall exposure inequality. Each area is decomposed according to the distribution of individuals across the three groups. The values of the group-specific Gini Exposure indices are:

$$G_E(A|g_1) = 0.909, \quad G_E(A|g_2) = 0.8333, \quad G_E(A|g_3) = 0.5714, \quad G_E(A|\mathcal{O}) = 0.53,$$

and the associated weighting scheme is:

$$\alpha_{g_1} = 0.25, \qquad \alpha_{g_2} = 0.06, \qquad \alpha_{g_3} = 0.053.$$

Using a similar analysis, one can evaluate the multi-group exposure patterns for cases where the interaction profiles are defined on more than two dimensions. We consider the case where interaction takes place with respect to the three groups considered, thus redefining the new interaction patterns $\boldsymbol{\pi}(A)$, $\boldsymbol{\pi}'(A)$ and $\boldsymbol{\pi}''(A)$ as follows:

$$\boldsymbol{\pi}(A) = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.5 \end{pmatrix}; \quad \boldsymbol{\pi}'(A) = \begin{pmatrix} 0 \\ 0.7 \\ 0.3 \end{pmatrix}; \quad \boldsymbol{\pi}''(A) = \begin{pmatrix} 0.2 \\ 0 \\ 0.8 \end{pmatrix}.$$

The distribution of these interaction profiles across the population is set as before. In Figure 1(b) it is reported the Segregation Zonotope associated to this distribution of interaction profiles (shaded in grey). The overall Gini Exposure index coincides with the zonotope *volume*, and it is equal to $G_E(A) = 0.1138$. It is possible to replicate the previous exercise to obtain the decomposition of the Gini Exposure index and identify (also graphically) the segregation patterns of the three groups under analysis.

4 A characterization result for measuring the exposure dimension of segregation

In this section, we study the minimal transformations of the data that allow to *characterize the Gini Exposure index* as a measure of segregation for a set of configurations of interactions. We use the term *segregation* to indicate any departure from the situation where interaction profiles are equalized within the network.

For expositional purposes we consider likelihood matrices of dimension $G \times N$ where $N \geq G \geq 2$. Each matrix is by construction row-stochastic, and represent configurations where in each column there exists at least a positive element. The set of such matrices in denoted $\mathcal{M}_{GN}$, where $\mathcal{M}_{GG}$ is the subset of $\mathcal{M}_{GN}$ containing only square row-stochastic matrices of dimension G that do not include empty units (i.e. columns with all elements equal to 0). For technical purposes we will consider also an extended set $\mathcal{M}^0_{GG}$ that will include matrices belonging to $\mathcal{M}_{GG}$ and those where at most $G-1$ rows could contain all 0's while the other rows are stochastic and no column has all elements equal to 0.

Let $\mathcal{L}^A \in \mathcal{M}_{GN}$, denote a *likelihood matrix* obtained from configuration A, its generic element $\ell_{gj} \geq 0$ represents the probability that an individual in group g interacts with individuals associated with unit j.

Given matrix $\mathcal{L}\in\mathcal{M}_{GN}$, we measure its exposure dimension segregation through the index $E^N(\mathcal{L})$, where $E^N : \mathcal{M}_{GN} \to [0,1]$ denotes a sequence of *continuous functions* from the set $\mathcal{M}_{GN}$ to the interval $[0,1]$. The index is increasing in the degree of segregation exhibited by a likelihood matrix and reaches its maximum value at 1.

We illustrate here some properties that should be satisfied by the E^N index.

Let Π_N denote a $N \times N$ permutation matrix. The set of all these matrices is $\mathcal{P}_N$. The property of *Units Anonymity* requires that the index is invariant with respect to permutations of the units (columns) of matrix $\mathcal{L}$.

Axiom 1 (UA: Units Anonymity) $E^N(\mathcal{L}) = E^N(\mathcal{L}\Pi_N)$ *for all* $\mathcal{L}\in\mathcal{M}_{GN}$, *all* $\Pi_N\in\mathcal{P}_N$.

The UA axiom can be interpreted equivalently also in terms of interaction matrices $\boldsymbol{\pi}$. It requires that segregation in exposure is not affected by columns permutations of $\boldsymbol{\pi}$. In this case also the weights of the units ξ should be permuted accordingly.

Next, the *Normalization* axiom identifies the reference case of maximal segregation. It is specified only for matrices in $\mathcal{M}_{GG}$. Let I_G denote the identity matrix of dimension G. When each unit is associated only to a group then the segregation is maximal and the index reaches the value of 1.

Axiom 2 (N: Normalization) $E^G(I_G) = 1$.

In terms of interaction matrices the maximal segregation is also associated with the case where the matrix $\boldsymbol{\pi}$ is an identity matrix.

In order to make segregation comparisons for matrices where $N > G$ we adopt a decomposition property. This property assumes that overall segregation evaluations can be based on a weighted combination of evaluations applied to square likelihood matrices in $\mathcal{M}_{GG}$ that are obtained by focussing only on G units ordered as in $\mathcal{L}$. According to this view the set $\mathcal{M}_{GG}$ is the minimal set of matrices that allow to "fully" express segregation evaluations. This is the smaller set of matrices where each unit could interact with only one group and all groups interact with at least one unit. This view is consistent with the N axiom that specify the reference case for maximal segregation in terms of matrices in $\mathcal{M}_{GG}$.

Let $\mathcal{N}_O(\mathcal{L})$ denote the set of *units ordered* according to the ranking in $\mathcal{L}$. Any G dimensional subset of ordered units is denoted by $\{i_1, i_2, ..., i_G\} \subseteq \mathcal{N}_O(\mathcal{L})$ where the index i_k denotes a unit in position $k \leq G$ in the units order obtained by eliminating $N - G$ units from the initial ordered set of units $\{1, 2, ..., N\}$ in $\mathcal{L}$. The obtained sub-matrix derived from $\mathcal{L}$ by keeping the units $\{i_1, i_2, ..., i_G\}$ is denoted $(\boldsymbol{\ell}_{i_1}, \boldsymbol{\ell}_{i_2}, ..., \boldsymbol{\ell}_{i_G})$. In general this square matrix of dimension G is not in $\mathcal{M}_{GG}$. This could be the case because all elements in a row are 0's, or more generally because for some/all rows the elements do not sum to 1. In order to accommodate the first case we consider matrices in $\mathcal{M}^0_{GG}$. In the second case, the matrix could however be made row-stochastic by dividing each row (except those made of all 0's) by the corresponding element of the vector $\boldsymbol{\lambda}^{\{i_1,i_2,...,i_G\}}$ obtained by calculating the product

$$\boldsymbol{\lambda}^{\{i_1,i_2,...,i_G\}} := (\boldsymbol{\ell}_{i_1}, \boldsymbol{\ell}_{i_2}, ..., \boldsymbol{\ell}_{i_G}) \cdot \mathbf{1}_G$$

where $\mathbf{1}_G$ denotes the G dimensional column vector of 1's. For simplicity of exposition we denote such row stochastic matrix as $(\tilde{\boldsymbol{\ell}}_{i_1}, \tilde{\boldsymbol{\ell}}_{i_2}, ..., \tilde{\boldsymbol{\ell}}_{i_G})$. Note that the generic element $\lambda_g^{\{i_1,i_2,...,i_G\}}$ of the vector $\boldsymbol{\lambda}^{\{i_1,i_2,...,i_G\}}$ corresponding to group g denotes the probability that an individual from group g interacts with one of the units in the set $\{i_1, i_2, ..., i_G\}$. The joint probability of interaction obtained taking into account all G groups is given by the product of all elements of $\boldsymbol{\lambda}^{\{i_1,i_2,...,i_G\}}$.

We are now in the position to formalize the *Decomposition* property that requires that the aggregate segregation evaluation could be decomposed in the weighted sum of all evaluations made over all ordered matrices in $\mathcal{M}^0_{GG}$ weighted according to the joint probability of interaction.

Axiom 3 (D: Decomposition)

$$E^N(\mathcal{L}) := \sum_{\{i_1,i_2,...,i_G\}\subseteq\mathcal{N}_O(\mathcal{L})} \left(\prod_{g=1}^{G} \lambda_g^{\{i_1,i_2,...,i_G\}} \right) \cdot E^G(\tilde{\boldsymbol{\ell}}_{i_1}, \tilde{\boldsymbol{\ell}}_{i_2}, ..., \tilde{\boldsymbol{\ell}}_{i_G})$$

for all $\mathcal{L}\in\mathcal{M}_{GN}$, *where* E^G *is defined over* $\mathcal{M}^0_{GG}$.

The decomposition property is consistent with the logic adopted by the Gini index to measure dispersions in income distributions, for the index in fact the aggregate dispersion is a weighted average of pairwise individuals comparisons of their distances.

Next axiom will allow to quantify changes in segregation moving from one matrix to another. To introduce it we consider a combination of two operations that should reduce the segregation for matrices in $\mathcal{M}_{GG}$ and we also quantify the cardinal level of this reduction. We first provide an intuition of these operations for interaction matrices $\boldsymbol{\pi}$ and then we express them in terms of transformations applied to likelihood matrices in $\mathcal{M}_{GG}$.

Consider a generic interaction matrix $\boldsymbol{\pi}$ and two units characterized by the column vectors of interactions $\boldsymbol{\pi}_{.i}$ and $\boldsymbol{\pi}_{.j}$, with demographic weights ξ_i and ξ_j respectively. Assume now that a proportion $0 \leq (1-\alpha) < 1$ of the individuals in unit i is joining unit j and shares its interaction probabilities. The new demographic weights then become respectively $\alpha\xi_i$ and $\xi_j+(1-\alpha)\xi_i$, the column vectors of interactions $\boldsymbol{\pi}_{.i}$ is unaffected but the one of group j is modified and is given by the weighted "mixture" of interaction probabilities of the merged proportions of units, it becomes

$$\boldsymbol{\pi}'_{.j} = (\xi_j \boldsymbol{\pi}_{.j} + (1-\alpha)\xi_i \boldsymbol{\pi}_{.i}) \left(\frac{1}{\xi_j + (1-\alpha)\xi_i} \right).$$

The axiom of "Exposure segregation reduction through Mixtures of units" postulates that these operations should not increase segregation, and more precisely also assumes that segregation should be reduced proportionally according to the coefficient $(1-\alpha)$. Thus, if the original configuration is associated to a positive level of segregation these operations should strictly reduce it proportion $(1-\alpha)$.

We formalize now the axiom in terms of matrices in $\mathcal{M}_{GG}$. This notation will allow us also to highlight more directly the connections of the property and the Gini measures of inequality.

Consider a likelihood matrix $\mathcal{L} = (\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \boldsymbol{\ell}_G) \in \mathcal{M}_{GG}$ represented by making explicit the G column vectors $\boldsymbol{\ell}_i$. This matrix can be transformed by taking a portion $(1-\alpha)$ where $0 < \alpha \leq 1$ of unit i and transferring it so to merge it with unit j, such that the resulting matrix $\mathcal{L}(\alpha, i, j)$ can be written as

$$\mathcal{L}(\alpha, i, j) = (\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \alpha\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_j + (1-\alpha)\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_G).$$

We call such transformations *elementary "mixture of units"*. Note that the matrix $\mathcal{L}(\alpha, i, j)$ is still in $\mathcal{M}_{GG}$ by construction. If $\boldsymbol{\ell}_j$ and $\boldsymbol{\ell}_i$ are not linearly dependent this operation should not increase the degree of segregation. The fact that $E^G(\mathcal{L}(\alpha, i, j)) \leq E^G(\mathcal{L})$ can be explained by combining two operations [split and merge of columns] discussed in Andreoli and Zoli (2014) and in the literature on dissimilarity measurement with permutable columns. On one hand the split of a unit into two units is supposed to keep dissimilarity/segregation unchanged, on the other hand the merge of two units that are linearly independent is supposed not to increase the dissimilarity/segregation, by leveling the disparities between these two units. Next axiom is also quantifying the reduction in segregation.

Axiom 4 (EM: Exposure segregation reduction through Mixtures of units) $E^G(\mathcal{L}(\alpha, i, j)) = \alpha E^G(\mathcal{L})$ *for all* $\mathcal{L} \in \mathcal{M}_{GG}$, $i, j \in \mathcal{N}$, $0 < \alpha \leq 1$.

Axiom EM is a generalization of a property satisfied by the Gini index when $G = 2$.

Consider for instance the Gini index derived from the matrix $L = \begin{bmatrix} p & 1-p \\ x & 1-x \end{bmatrix}$ where for expositional purpose we assume that p denotes the proportion of poor individuals whose income is a share $x < p$ of the total income, and $1-p$ is the proportion of rich individuals that own a share $1-x$ of the society income.

The Gini index for this society is $p-x$ as could be calculated by computing the Lorenz curve that in this case is piecewise linear with coordinates $(0,0)$, (p,x), $(1,1)$. Suppose that we apply an elementary mixture transformation to the data by post multiplying matrix L by $\begin{bmatrix} \alpha & 1-\alpha \\ 0 & 1 \end{bmatrix}$ the obtained new matrix $L(\alpha, 1, 2)$ will be

$$L(\alpha, 1, 2) = \begin{bmatrix} \alpha p & (1-\alpha)\,p + 1 - p \\ \alpha x & (1-\alpha)\,x + 1 - x \end{bmatrix} = \begin{bmatrix} \alpha p & 1 - \alpha p \\ \alpha x & 1 - \alpha x \end{bmatrix}.$$

The Gini index of matrix $L(\alpha, 1, 2)$ will be $\alpha p - \alpha x = \alpha \cdot (p - x)$, precisely as postulated by axiom EM.

An analogous result could be obtained by post multiplying L by $\begin{bmatrix} 1 & 0 \\ 1-\alpha & \alpha \end{bmatrix}$ which is the matrix related to the other possible elementary mixture operation created by splitting column 2 and merging it with column 1. In this case

$$L(\alpha, 2, 1) = \begin{bmatrix} 1 - \alpha\,(1-p) & \alpha\,(1-p) \\ 1 - \alpha\,(1-x) & \alpha\,(1-x) \end{bmatrix}$$

whose Gini index is $1-\alpha(1-p)-1+\alpha(1-x)=\alpha\cdot(p-x)$.

We are now ready to prove the main characterization result for the Gini Exposure index.

We apply our result to a set $\hat{\mathcal{M}}_{GG}\subseteq\mathcal{M}_{GG}$ of square matrices of dimension G that could be obtained as a combination of elementary mixture of units/columns operations and permutations of units/columns. As we will show when $G=2$ the two sets coincide, that is $\hat{\mathcal{M}}_{22}=\mathcal{M}_{22}$. However, the fact that $\hat{\mathcal{M}}_{GG}$ could be strictly included in $\mathcal{M}_{GG}$ for $G\geq 3$, is an open question. Making use of an equivalence result in Theorem 1 in Andreoli and Zoli (2014) it can be shown that any matrix in $\mathcal{M}_{GG}$ can be obtained from I_G through a finite sequence of splitting of units, merge of units and permutation of units.[7] Even though split and merge operations are the basis for the elementary mixture of units, there is no guarantee that any appropriate sequence could be decomposed combining all split and merge operations into elementary mixture operations so that the starting and arriving matrices are all square matrices of dimension G.

Proposition 2 *Let $\mathcal{L}\in\hat{\mathcal{M}}_{GG}$, the exposure segregation index $E^G:\mathcal{M}_{GG}\to[0,1]$ satisfies axioms UA, N and EM if and only if it is the absolute value of the determinant of $\mathcal{L}$, that is*

$$E^G(\mathcal{L}):=|\det\mathcal{L}|\,.$$

Proof. See Appendix B.2. ■

Before moving to the extension of the result to matrices where $N>G$ it is important to highlight the relevance of the restrictions applied in the result in Proposition 2, that holds for matrices in $\hat{\mathcal{M}}_{GG}$. As already stated, the fact that $\hat{\mathcal{M}}_{GG}$ and $\mathcal{M}_{GG}$ could not coincide when $G\geq 3$, is an open question. At this stage we are not in the position either to prove this fact or to disprove it. The degree of flexibility in expanding the set $\hat{\mathcal{M}}_{GG}$ so that it could coincide with $\mathcal{M}_{GG}$, is "large". In fact the construction of matrices

[7] This statement is obtained by considering the equivalence between statements (iv) and (i) in Theorem 1 in Andreoli and Zoli (2014) and adapting it to the notation adopted here. The result is more general as the one restated here, in fact it holds for matrices B and A where $B=AX$ for a generic row stochastic matrix X such that A and B could exhibit a different number of columns. It involves the possibility of adding or eliminating empty classes, the use of operations of splitting of columns and merging of columns and columns permutations. In the current setting it suffices to consider $A=I_G$.

in $\mathcal{M}_{GG}$ requires to identify $G \cdot (G-1)$ values, because matrices are row-stochastic. Moreover, the inequalities restrictions among all the values of the matrix are at most $\frac{1}{2}\left[(G^2-1) \cdot G^2\right]$. On the other hand, in order to construct $\hat{\mathcal{M}}_{GG}$ it is possible to use $G \cdot (G-1)$ transformation matrices as $T(\alpha, i, j)$ in (1) each one with a possibly different parameter α. Moreover, the product of these matrices could be permuted in $[G \cdot (G-1)]!$ configurations, and could be possibly integrated in the sequence by insertion of permutation matrices. This large flexibility in the number of parameters and operations behind the construction of $\hat{\mathcal{M}}_{GG}$ has however not yet allowed us to obtain a conclusive answer on whether $\hat{\mathcal{M}}_{GG} = \mathcal{M}_{GG}$ or $\hat{\mathcal{M}}_{GG} \subset \mathcal{M}_{GG}$ even for the $G = 3$ case.

The result in Proposition 2 goes beyond the simple proof of the sufficiency part, it shows also the necessity condition for the characterization that holds for a potentially large set of admissible matrices of interest.

Note that $E^G(\mathcal{L}) = |\det \mathcal{L}|$ is the *volume of the zonotope* associated with matrix $\mathcal{L} \in \hat{\mathcal{M}}_{GG}$. By construction the index is such that if two columns are linearly dependent then the index takes the value of 0. This is certainly a limitation for the use of this measure if one restricts attention to problems with G groups and G units.[8] However, as we are going to show by making use of axiom D, when $N > G$ the overall segregation measure boils down to 0 only if there are not G linearly independent column vectors among the N vectors associated to the units in $\mathcal{L}$. A sufficient case for this extreme result is obtained when two rows of $\mathcal{L}$ are identical.

We derive now the general formula for the exposure segregation index for matrices where $N > G$. Given the construction of Proposition 1, we consider the set of matrices $\hat{\mathcal{M}}_{GN} \subseteq \mathcal{M}_{GN}$, such for any matrix in $\hat{\mathcal{M}}_{GN}$ any of its square submatrices, where all rows have at least one positive element, that are obtained after eliminating $N - G$ columns is in $\hat{\mathcal{M}}_{GG}$. By direct application of axiom D in conjunction with the result of Proposition 2 it follows that:

Corollary 1 *Let $\mathcal{L} \in \hat{\mathcal{M}}_{GN}$, the exposure segregation index $E^N : \mathcal{M}_{GN} \rightarrow [0,1]$ satisfies axioms UA, N, EM and D if and only if it is the Gini Exposure index, that is*

$$E^N(\mathcal{L}) := \frac{1}{G!} \sum_{\{i_1, i_2, \ldots, i_G\} \subseteq \mathcal{N}(\mathcal{L})} \left| \det(\boldsymbol{\ell}_{i_1}, \boldsymbol{\ell}_{i_2}, \ldots, \boldsymbol{\ell}_{i_G}) \right|.$$

[8] These concerns are in line with those expressed in Shorrocks (1978) for the measures of intergenerational mobility that take into account the determinant of the square transition mobility matrix.

Proof. See Appendix B.3. ■

Next remark formalizes the fact that when two groups are considered the result applies to all matrices in $\mathcal{M}_{2N}$. In this case Proposition 2 and the related corollary provide a full characterization of the Gini segregation index in the two groups case.

Remark 2 $\hat{\mathcal{M}}_{22} = \mathcal{M}_{22}$.

In order to prove the remark, consider matrix

$$L = \begin{bmatrix} p & 1-p \\ x & 1-x \end{bmatrix}$$

where $p > x$, we show that it can be obtained as the product of matrices

$$T(\alpha_1, 1, 2) = \begin{bmatrix} \alpha_1 & 1-\alpha_1 \\ 0 & 1 \end{bmatrix}, \text{ and } T(\alpha_2, 2, 1) = \begin{bmatrix} 1 & 0 \\ 1-\alpha_2 & \alpha_2 \end{bmatrix}$$

and eventually the permutation matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Consider the product

$$T(\alpha_1, 1, 2) \cdot T(\alpha_2, 2, 1) = \begin{bmatrix} \alpha_1 + (\alpha_1 - 1)(\alpha_2 - 1) & \alpha_2 (1-\alpha_1) \\ 1-\alpha_2 & \alpha_2 \end{bmatrix}$$

it follows that $\alpha_2 = 1 - x$ and $\alpha_2 (1 - \alpha_1) = 1 - p$ that is $\alpha_1 = \frac{p-x}{1-x}$. This latter term is consistent with the definition of $\alpha_1 > 0$. If this was not the case then one has to permute the columns of the matrix to get the result.

Recall that the value of the index could be obtained by multiplying $\alpha_1 \cdot \alpha_2$ it follows that $E^2(L) = \frac{p-x}{1-x} \cdot (1-x) = p - x = |\det(L)|$ that also coincides with the Gini index of the distribution.

5 Comparison with other indices within the spatial interaction model

This section compares the Gini Exposure index with other measures of the exposure dimension of segregation. To do so, we use a qualitative analysis of the properties characterizing these different indices, as well as a quantitative analysis based on empirical correlations. The focus is on interaction at the urban level.

In many applications on urban networks, data are only available for (i) the demographic size of the groups, (ii) the distribution of groups across a well defined partition of the space into organizational units and, possibly, (iii) a measure of the proximity between the units, decreasing with their distance or diversity. Within this framework, interaction profiles can be inferred making use of a spatial model for interaction, approximating the actual unobservable network.

Spatial data have been often used in the sociological literature to assess the spatial dimensions of social interaction at the urban level. Under the postulate that social interactions frequency decreases with spatial distance, sociologists have combined demographic data with spatial information to design the probabilities that one individual has to interact with other groups, as a function of her location in the space where interactions take place. Here, we adopt a similar strategy to asses the social segregation patterns across Italian provinces. Each province gathers together many municipalities. We assume that the Italian municipalities are representative agents whose weight depends on the municipality demographic size.

5.1 Additional notation for the spatial model

A configuration A defines the distribution of individuals within a province. Consider the case in which the interaction space is partitioned into N_A non-overlapping *organizational units* $i = 1, \ldots, N_A$ and use $\mathcal{N}(A)$ to denote this set. Let $n_{gi}(A)$ be the *observed* number of individuals living in the same organizational unit i who are of group g. Each organizational unit is assumed to have a demographic weight $\widehat{\xi}_i(A) = \frac{n_{gi}(A)}{\sum_{g\in\mathcal{G}}\sum_{i\in\mathcal{N}(A)} n_{gi}(A)}$ where the “hat” symbol is used to denote an empirical weighting scheme.

The second ingredient in our analysis is a measure of *distance*, d, between organizational units' centroids. Coherently with a long stream of works in sociology (see for instance White 1983, Reardon and O'Sullivan 2004, Echenique and Fryer 2007), we assume that the spatial distance accounts for social distance between individuals, so that the likelihood that two individuals interact is a decreasing function of their spatial distance.

Let $\delta_A(i, h; d)$ denotes a measure of *proximity* of two units $i, h \in \mathcal{N}_A$ which is inversely related to the distance function $d(i, h)$. We impose the proximity to be maximal and equal to 1 when i and h coincide according to the distance criterion d (so that $\delta_A(i,i;d) = 1$), while the measure decreases and approaches the value 0 the larger is the distance between the two units.

The proximity-weighted counting indicator $\widehat{n}_{gi}(A) = \sum_{i\in\mathcal{N}_A} n_{gi}(A)\delta_A(i,h;d)$ measures the number of individuals of groups g with whom an individual in unit i may interact with. The overall interaction potential can be measured by the total amount of individuals that can be associated to unit i, $\sum_{g\in\mathcal{G}} \widehat{n}_{gi}(A)$. Combining together these two indicators, one obtains an empirical measure of the probability to interact with group g conditional on the fact that interaction takes place in organizational unit i: $\widehat{\pi}_{gi}(A) = \frac{\widehat{n}_{gi}(A)}{\sum_{g\in\mathcal{G}} \widehat{n}_{gi}(A)}$. An empirical interaction profile is now denoted by the vector $\widehat{\boldsymbol{\pi}}_i = (\widehat{\pi}_{1i}, \ldots, \widehat{\pi}_{Gi})^t$. The expected interaction profile is denoted $\widehat{\pi}^e_g(A)$ and the interaction matrix is denoted $\widehat{\mathbf{A}}$, whose entry g, i is equal to $\widehat{a}_{gi} = \frac{\widehat{\pi}_{gi}(A)}{\widehat{\pi}^e_g(A)}$.

5.2 Qualitative comparison with other indices

Reardon and O'Sullivan (2004) systematically analyzed the exposure segregation measures treated in the literature, and proposed some meaningful properties that these indices should satisfy.

The first property, *scale interpretability* is satisfied by construction of the Gini Exposure index. We interpret $G_E(A) = 0$ as the case where interaction profiles coincide across groups and units, while $G_E(A) = 1$ as the opposite case of perfect segregation, occurring only in the case where the interaction profile allocated to each of the units is degenerate, that is it assigns a probability of interaction with group g equal to zero for the remaining.

The implementable model is not exempted from the MAUP problem,[9] and therefore the *arbitrary boundary independence* property is not satisfied. This is a drawback of our empirical analysis (based on a pre-determined partition of the space into organizational units) rather than an issue related to the index itself.

The implementable Gini Exposure index meets the requirements of *location equivalence*. In fact, if two organizational units are associated to the same interaction pattern, the operations of mixing the two together into a new unit preserves the segregation order characterized by the merge axiom,

[9]The Modifiable Areal Unit Problem occurs when the partition of the space into organizational units is exogenously fixed.

and hence the Gini Exposure index.[10]

The implementable Gini Exposure index also satisfies *Population density invariance*, however it does not respect the requirements of *composition invariance*. In fact, the Gini Exposure index captures a form of *relative* inequality in the distribution of interaction patterns across the population. Therefore, the index is independent from the overall expected interaction profiles. In particular, the index is not independent from the *size* of the groups. The convenience of satisfying composition invariance is, nevertheless, debatable (see Frankel and Volij 2011). We consider in our empirical comparison the *Atkinson* multi-group segregation index in Frankel and Volij (2011) that, differently from the other exposure indicators, is composition invariant.

Finally, it is impossible to establish if *transfer* and *exchange* principles are satisfied by the Gini Exposure index. In fact, the mixture of interaction profiles is not defined in the form of a movement of population masses across organizational units (transfer) or groups (exchange) but rather as a convex combination of interaction profiles. We compute empirical correlations between indices satisfying the transfer/exchange principle and the Gini Exposure index, to recover a relation between mixtures, transfers and exchanges.

We compare the multi-group Gini Exposure index with other multi-group measures proposed in the literature. In the class of indices that do not satisfy composition invariance, the first index that we consider is a *spatial* version of the the *Mutual Information* index $M(A)$ characterized (among others) by Frankel and Volij (2011).

The entropy of the discrete probability distribution $(p_1, \ldots, p_G)$ is defined by:

$$E(p_1, \ldots, p_G) \;=\; \sum_{g \in \mathcal{G}} p_g \log_2 \left(\frac{1}{p_g} \right).$$

The Mutual Information index equals the entropy of an allocation's groups distribution minus the average entropy of individual interaction profiles:

$$M(A) \;=\; E\left(\widehat{\pi}^e_1(A), \ldots, \widehat{\pi}^e_G(A)\right) \;-\; \sum_{i \in \mathcal{N}(A)} \widehat{\xi}_i(A)\, E\left(\widehat{\pi}_{1i}(A), \ldots, \widehat{\pi}_{Gi}(A)\right).$$

Alternatively, we also consider other spatial indices studied in Reardon

[10] This is so because all the agents living in the two regions are endowed with the same interaction profiles.

and O'Sullivan (2004). All these indices satisfy the transfer and exchange principles, provided that some symmetry requirements on the proximity measure are imposed. The first index is the *Spatial Relative Diversity* index $R(A)$. It consists in a ratio between the average degree of heterogeneity in the composition of probabilities of interaction of each unit's interaction profile and the degree of heterogeneity in the the composition of probabilities of interaction of the expected interaction profile. The heterogeneity in the the composition of probabilities of interaction observed in unit i is measured by the *interaction* coefficient $I_i(A) := \sum_{g\in\mathcal{G}(A)} \widehat{\pi}_{gi}(A)\,(1-\widehat{\pi}_{gi}(A))$ for each organizational unit i and for the population as a whole, denoted by the coefficient $I(A) := \sum_{g\in\mathcal{G}} \widehat{\pi}^e_g(A)\,(1-\widehat{\pi}^e_1(A))$. The relative diversity amounts to:

$$R(A) \;=\; 1 \;-\; \sum_{i\in\mathcal{N}(A)} \widehat{\xi}_i(A)\frac{I_i(A)}{I(A)}.$$

The *spatial dissimilarity* index $D(A)$ is a measure of how different the composition of individuals' organizational units environments are, on average, from the composition of the population as a whole. It is defined as follows:

$$D(A) \;=\; \frac{1}{2\,I(A)}\sum_{g\in\mathcal{G}}\sum_{i\in\mathcal{N}(A)} \widehat{\xi}_i(A)\left|\widehat{\pi}_{gi}(A)-\widehat{\pi}^e_g(A)\right|.$$

The last two indices that we consider are the empirical counterpart of the expected Gini index, denoted $EG(A)$, and the *normalized spatial exposure* index $NE(A)$, which is defined as:

$$NE(A) \;=\; \sum_{g\in\mathcal{G}}\sum_{i\in\mathcal{N}(A)} \widehat{\xi}_i(A)\,\frac{\left(\widehat{\pi}_{gi}(A)-\widehat{\pi}^e_g(A)\right)^2}{1-\widehat{\pi}^e_g(A)}.$$

This index belongs to the class of the variance indicators. In the Reardon and Firebaugh (2002) taxonomy, the two indices fall into the class of the indicators measuring segregation as a form of distributional inequality of the interaction profiles.

5.3 Quantitative comparisons: immigrants segregation across Italian municipalities

In this section we study the empirical performances of the spatial segregation indices discussed above. We exploit a panel of nearly 8400 Italian municipalities, observed in the period 2003 to 2010. The municipalities are clustered at province level (110 provinces in 2010, of which 101 remain fixed over time, each made by 74 municipalities on average), covering on average a population of 551,000 inhabitants. Each municipality has an average demographic size of 6,400 individuals, comparable to the dimension of the US MSAs districts. In the analysis, each municipality corresponds to an organizational unit, with $\mathcal{N}_{pt}$ the set of municipalities that belongs to a given province p in time t. We exploit the patterns of segregation of immigrants and natives (for a total of $G = 3$ groups) for each province p in each year t. This can be done by calculating a segregation index for each pair p, t. In this way, we have sufficient time and space variability to construct and analyze segregation patters in Italy, while keeping a sufficiently refined spatial scale.

We propose to study the degree of segregation of three mutually exclusive social groups: Italian natives, immigrants from countries with *high* HDI levels, and immigrants from countries with *low* HDI levels.[11] This multi-group separation (compared to the traditional bivariate analysis of immigrants versus natives) is of particular relevance in Italy, since immigration is a recent and growing phenomenon, and the *type* of the country of origin (as measured by the HDI) is a relevant factor to account for.

The distribution of the Italian provinces is represented in Figure 2. For each province we construct a spatial model to measure interaction profiles at the municipality level. Then, we compute the values of the segregation indices G_E, M, NE, R, D, EG and A for each of the provinces, using municipalities as organizational units. These indicators are meant to summarize the information about the distribution of interaction probabilities *within provinces*. We end up with 808 data point for each of the indicators, varying across the 101 provinces and the 8 years considered.

We study the empirical rank correlations of the indices, and we asses the differences in the type of segregation patterns that can be captured according to the indicator used. Then, we apply the decomposition of the Gini Exposure index to the data to study the contribution of each group to the overall exposure.

[11]The *Human Development Indicator* (HDI) proposed by the UNDP department is a synthetic indicator computed on a country and year bases for evaluating the multivariate distribution of health, resources and educational indicators across the population in that year and country. The UNDP also provides a classification of countries according the their HDI profile.

5.3.1 Data

We build the spatial analysis using ISTAT demographic data[12] at municipality level. We obtain data on the demographic size of the resident population, partitioned according to the nationality. Municipalities are grouped into provinces, according to the official repartition of the Italian territory. Table 1 in the appendix collects information on the variability of the number and demographic size of provinces and municipalities across the period considered.

Two immigration groups have been created according to the definition of low level and medium/high level HDI countries provided by the UNDP for 2011. In 2010, the share of immigrants from low HDI countries amounts to 6.7% of the total population by province, on average (Table 1), while it is particularly high in the north of Italy (Figure 2).

We use a spatial proximity index to empirically identify the interaction probabilities $\widehat{\pi}_{gi}$. We proceed as follows. Each municipality has been geocoded, so that latitude and longitude are now available for each municipality's *centroid*. We assume that the interaction probability decays with spatial distance. We construct a set of interaction profiles associated to each municipality *within* the same province (but not across provinces). To do so, we compute $\widehat{n}_{gi}$, assuming $\delta(i, h; d)$ to be a *biweight kernel* estimator of proximity, and we take d as the spatial distance between municipalities, censored at a 20km threshold.[13] Interaction probabilities and *expected* (by province) probabilities are calculated according to their definitions. We use the relative demographic size of a municipality i in a given province to infer $\widehat{\xi}_i$.

Interaction probabilities with immigrants from low HDI countries grew substantially but uniformly over the 8 years span, although the probability

[12]The municipality level composition (by nationality) of the resident population in Italy from 2003 to 2010 can be freely downloaded from the official ISTAT (the Italian Statistical Institute) webpage at the following link: `http://demo.istat.it/`.

[13]We use the biweight kernel to assign the distance weighted number of individual of each group to municipality i. The kernel has a Gaussian-like shape, although it is bounded, so that all the municipalities outside a given radius of length $r = 20km$ are assumed to have no weight in determining interaction probabilities for the population living in i. The weight decreases according to the spatial distance. The proximity weighting function is: $\delta(i, h; d) := \mathbf{1}(d(i, h) < 20km) \cdot \left(1 - \left(\frac{d(i,h)}{20km}\right)^2\right)^2$, where $\mathbf{1}(.)$ is the indicator function and $d(i, h)$ is the spatial distance based on the *cosine method*, and calculated by using latitude and longitude information for municipalities centroids.

of interaction with this group remains high in the north of Italy, both at the level of provinces (Figure 4) and municipalities (Figure 5).

The distribution of interaction profiles across municipalities within the same province defines the object of our study. Additional measures of disproportionality at municipality level, as well as indices at province level, can now be calculated.

5.3.2 Segregation patterns across Italian provinces

The distribution matrix associated to a given province provides information about the disproportion in interaction probabilities at municipality level versus the expected probabilities at province level. Figure 6 reports the spatial distribution of this *disproportionality* coefficient for the group of immigrants from low HDI countries, defined as $a_i = \widehat{\pi}_{\text{low HDI},i}(A)/\widehat{\pi}^e_{\text{low HDI}}(A)$. If a_i is larger than one, then the probability of interacting with immigrants from low HDI countries in i is larger than what is expected at province level.

In north-east and central Italy it is observed the largest within province variability in interaction disproportionality, which implies higher variability across municipalities in the type of interaction profiles. These macro-regions (the distribution of disproportionality coefficients at municipality level in the north Italy region is reported in Figure 7) are also characterized by large variability in their ranking position throughout the period, while the expected interaction level remains stable (see Figure 8 and the figures in Table 1, reporting the percentage of municipalities where $a_i > 1$, stable in the 2003/2010 period).

This particular pattern of (exposure) segregation across municipalities is captured by both by the Gini Exposure index and the Mutual Information index by Frankel and Volij (2011), which we take as a reference for the class of multi-group exposure indices that do not satisfy composition invariance. The position of all the 101 Italian provinces (for which data are available) in the ranking produced by the Gini Exposure index are depicted in Figure 9 for the year 2003 and 2010. Provinces are ranked according to increasing segregation. The top 20 segregated provinces are concentrated in the center and the north-east regions of Italy. This outcome is coherent with the fact that the Gini Exposure index captures the within province variability in interaction profiles, which is consistent in the two macro-regions.

The Mutual Information index provides a closely related (although not coincident) picture regarding the distribution of the entropy associated to each interaction profile within provinces (Figure 10). The changes in the ranking of the provinces by 2003 to 2010 (right panel of Figures 9 and 10) generated by the two indices, however, do not coincide.

5.3.3 Comparison of segregation indices

The graphs in Figure 12 suggest two well defined patterns of segregation that distinguish the spatial indices under analysis. For each of the six indices considered (G_E, M, NE, R, EG, D) calculated by province and year, we report three curves, identifying the dynamics of segregation across years associated to the province scoring at the first, median and third quartile of the ranking of provinces defined, for each year, by one given index.

The Gini Exposure index identifies a slightly decreasing pattern of segregation across years for the three reference provinces taken into account. One can interpret the graph in the following way: the segregation pattern measured for the most segregated province among the least 25%, 50% and 75% segregated provinces in a given year is decreasing across the time interval considered. A similar pattern emerges if we consider the multi-group Dissimilarity index. Also for the Expected Gini index the patterns is almost stable across time.

On the other hand, the Mutual Information index defines a different pattern: segregation is slightly increasing in time for the moderately (25% and median) segregated provinces, while the growth in segregation of the most segregated provinces is even more evident. Similar patterns are also exhibited by the Normalized Exposure or the Relative Diversity indices. A possible explanation of this divergence in segregation patterns is that the two families of indices obeys to different aggregation principles.

We study more in depth the ordinal relation between the six indices, along with the composition invariant Atkinson index characterized by Frankel and Volij (2011), by computing the rank correlations between the indices, reported in Table 2. The correlations are all positive and significant. As anticipated above, the Gini Exposure index is significantly positively rank correlated with the Dissimilarity index ($\tau_b = 0.593$ and $\rho = 0.773$), although the link with the Expected Gini is less evident. On the other hand, the Mutual Information index, the Relative Diversity and the Normalized Exposure measures generate significantly similar ranking of segregated distributions.

Differently from the majority of the composition invariant measures, the

Gini Exposure and the Dissimilarity indices are also (weakly) correlated with the Atkinson index, thus remarking that the two indices, in part, are affected by the distributions and account for the changes in overall composition.[14].

In Figure 13(a) we decompose this correlation across years. We identify two distinct patterns of correlation between the Gini Exposure index and the remaining composition invariant indices. The rank correlation between the Gini Exposure index and the Dissimilarity index remains fairly stable across time and persistently high. This pattern is distinct from other patterns, obtained by plotting rank correlations between the Gini Exposure index and other indices such as M, NE and R. We also tried to perform the inverse analysis, that is comparing for each province the correlation in ranking between years. However, it is not possible to disentangle any clear pattern among observed correlations.

Finally, we try to decompose the sources of correlation across periods by making use of the main demographic variables that we have studied, such as the share of groups in a province, the interaction profile associated to a province, or the size of the population. The objective is to assess the impact of the variability in the data on the rank correlation between pairs of indices. To do so, we focus on the correlation of the Gini Exposure with D (first pattern) and with M (second pattern). We use regression models to explain the contribution of each province in determining the Kendall's τ_b correlation measure used to construct Figure 13. In fact, the Kendall's index of rank correlation is an average of the degree of measured *concordance* associated to each observation.[15] Hence, traditional OLS methods are suitable to assess the association of variability in concordance across provinces with the characteristics of each province.

We perform six regressions and we report the results in Table 3. Each regression gives a list of coefficients that identify the impact of marginal variations in the independent variable on the rank correlation between the Gini

[14] The reported correlation, for the appropriate indices, are comparable to the one computed in Frankel and Volij (2011)

[15] The Kendall's τ_b can be written as $\tau_b/4 = \#\text{concordant pairs}/n(n-1) - (1 - a/n(n-1))$. Let $\rho_i(I)$ be the rank of province i in a given year produced by the index I. We say that, within a given year, provinces i and j are concordant with respect to two indices I, I' if $(\rho_i(I) - \rho_j(I))\,(\rho_i(I') - \rho_j(I')) > 0$. The number of *concordant pairs* is the sum of concordances for each i. The index linearly depends only on the concordant cases, since $n(n-1)$ is the total number of pairwise comparisons and a the number of cases in which provinces i, j are ranked in the same position by both indicators.

Exposure and the Dissimilarity index (models (1) to (3)) or alternatively the Mutual Information index (models (4) to (6)). Models (3) and (6) control for time dummies (where 2003 is the reference value). As shown before, correlation patterns do not substantially differ across years. Moreover, the association between common variables entering in the segregation indices is very low. We conclude that the rank association between the Gini Exposure index and the Dissimilarity index does not rely on the variability of the data considered. Moreover, the two indices produce very consistent rankings, and these rankings are not influenced by the structure of the data.

We repeat the same analysis by regressing the contribution of each observation in determining the rank correlation between the Gini Exposure index and the Mutual Information index. Results for the complete specification are reported in Model (6). As in the previous case, variables measuring the population (total or group level) distribution across provinces in absolute or relative terms have no impact in explaining changes in correlation. However, time fixed effects are significantly different from zero. This result, along with the fact that variability in inequality within interaction profiles (captured by the odds of interacting with an immigrant) have a significant negative impact on correlation between G_E and M, which let us conclude that the association between G_E and M is in part due to the variability in the data, and decreases sensibly when the odds of interacting with one of the groups are low. Therefore, the two indices may capture different information when faced with substantial within interaction profiles heterogeneity. The Gini Exposure index is, however, robust with respect to these differences.

6 Concluding remarks

We have proposed and analysed a new measure of multi-group segregation in networks: the Gini Exposure index. The index is designed to evaluate across the individuals in a network the inequality in the distribution of their interaction profiles with social groups. It can be interpreted as the volume of the zonotope of the matrix of the likelihood probabilities of interaction with the social groups in analogy with the generalization of the inequality Gini index in the multidimensional setting provided by the volume Gini index. In order to highlight the properties of the Gini Exposure index we have presented an axiomatic characterization of the index that holds for a large set of interactions configurations.

Adopting a spatial model of interaction based on Italian data on the distribution of immigrants across municipalities we have analyzed the behavior of the Gini Exposure index compared to other segregation indices which are not defined from an individual level perspective.

References

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

Andreoli, F. and Zoli, C. (2014). Measuring dissimilarity, *Working Papers Series, Department of Economics, Univeristy of Verona, WP23* .

Arnold, B. C. (2005). Inequality measures for multivariate distributions, *Metron - International Journal of Statistics* **LXIII**(3): 317–327.

Carrington, W. J. and Troske, K. R. (1997). On measuring segregation in samples with small units, *Journal of Business & Economic Statistics* **15**(4): 402–409.
URL: http://www.jstor.org/stable/1392486

Cutler, D. and Glaeser, E. L. (1997). Are ghettos good or bad?, *Quarterly Journal of Economics* **112**: 827–872.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes, *American Sociological Review* **20**(2): 210–217.
URL: http://www.jstor.org/stable/2088328

Ebert, U. (2010). The decomposition of inequality reconsidered: Weakly decomposable measures, *Mathematical Social Sciences* **60**: 94–103.

Echenique, F. and Fryer, R. G. (2007). A measure of segregation based on social interactions, *Quarterly Journal of Economics* **122**(2): 441–485.

Echenique, F., Fryer, Roland G., J. and Kaufman, A. (2006). Is school segregation good or bad?, *American Economic Review* **96**(2): 265–269.
URL: http://www.aeaweb.org/articles.php?doi=10.1257/000282806777212198

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

Frankel, D. M. and Volij, O. (2011). Measuring school segregation, *Journal of Economic Theory* **146**(1): 1 – 38.
URL: http://www.sciencedirect.com/science/article/B6WJ3-5161PK7-7/2/f3c23c21aeeb1e2ba8377436e3fa937b

Hutchens, R. M. (1991). Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations, *Mathematical Social Sciences* **21**(1): 31 – 51.
URL: http://www.sciencedirect.com/science/article/B6V88-4582FM1-1D/2/b1891c8954a6d40963da229a60acbffa

Hutchens, R. M. (2004). One measure of segregation, *International Economic Review* **45**(2): 555–578.
URL: http://www.jstor.org/stable/3663531

Koshevoy, G. A. and Mosler, K. (1997). Multivariate Gini indices, *Journal of Multivariate Analysis* **60**(2): 252–276.
URL: http://ideas.repec.org/a/eee/jmvana/v60y1997i2p252-276.html

Koshevoy, G. and Mosler, K. (1996). The Lorenz zonoid of a multivariate distribution, *Journal of the American Statistical Association* **91**(434): 873–882.
URL: http://www.jstor.org/stable/2291682

Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation, *Social Forces* **67**(2): 281–315.
URL: http://www.jstor.org/stable/2579183

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

Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation, *Sociological Methodology* **32**: 33–67.
URL: http://www.jstor.org/stable/3186153

Reardon, S. F. and O'Sullivan, D. (2004). Measures of spatial segregation, *Sociological Methodology* **34**(1): 121–162.
URL: `http://dx.doi.org/10.1111/j.0081-1750.2004.00150.x`

Shephard, G. C. (1974). Combinatorial properties of associated zonotopes, *Canadian Journal of Mathematics* **26**: 302–321.
URL: `http://cms.math.ca/cjm/v26/cjm1974v26.0302-0321.pdf`

Shorrocks, A. F. (1978). The measurement of mobility, *Econometrica* **46**(5): 1013–1024.
URL: `http://www.jstor.org/stable/1911433`

Silber, J. G. (1989). On the measurement of employment segregation, *Economics Letters* **30**(3): 237 – 243.
URL: `http://www.sciencedirect.com/science/article/B6V84-45DMS77-7B/2/b62c3ef322470f5af58ce78f5d1df60c`

White, M. J. (1983). The measurement of spatial segregation, *The American Journal of Sociology* **88**(5): pp. 1008–1018.
URL: `http://www.jstor.org/stable/2779449`

A Illustration

We introduce, with a simple example, the representation of the data that we use and the type of transformations involved in our analysis.

An interaction profile defines the conditional probability that a given population unit[16], denoted by i, interacts with each of the social groups, denoted by g, in which the population is partitioned. This probability is denoted by π_{gi}. Each unit is associated with its own interaction profile, that may depend on her network, location, or demographic attributes. In the example, we consider four individuals l_1, l_2, j and k. Two of them belong to group g_1 and the remaining to group g_2. We assume that profiles of interactions are estimated on the individual bases, so that the units of analysis coincide with the individuals. The interaction profile of individual i specifies the probability he or she has to interact with groups g_1 and g_2. Let assume for simplicity that individuals l_1 and l_2 share the same interaction profile, which is marked with an l. We can reduce the analysis to three profiles. We use the following data to fix ideas:

$$\begin{pmatrix} \pi_{g_1 l} \\ \pi_{g_2 l} \end{pmatrix} = \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}, \quad \begin{pmatrix} \pi_{g_1 j} \\ \pi_{g_2 j} \end{pmatrix} = \begin{pmatrix} 1/8 \\ 7/8 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \pi_{g_1 k} \\ \pi_{g_2 k} \end{pmatrix} = \begin{pmatrix} 3/8 \\ 5/8 \end{pmatrix}.$$

According to the first profile, the chances that l_1 or l_2 interact with a person of group g_1 is 25%, while 75% of the times they interact with members of group g_2.

To normalize the data and eliminate any form of heterogeneity *within* interaction profiles we use the vector of expected interaction probabilities π_g as the endogenously determined reference interaction profile.

It turns out that segregation can be measured as a form of *dissimilarity* (see Andreoli and Zoli 2014) between the likelihood that any randomly drawn individual of group g interacts with the demographic unit i, for any group g and any unit i. This likelihood, denoted $\mathcal{L}(i|g)$, should ideally equate the probability of interacting with unit i, namely $\Pr[i]$ if interaction profiles are equally distributed in the population. That is, $\mathcal{L}(i|g) = \mathcal{L}(i|g')$ for all is and

[16]A unit can be an individual, in which case it receives a weight equal to the inverse of the overall population size. It can also represent the minimum statistical unit used to empirically construct interaction profiles, for instance a class of student in a school, a neighborhood or a family unit. In this case, the weight of the unit is proportional to the group of individuals attached to that unit, and thus experiencing the same interaction profile

all groups $g \neq g'$. Any departure from this configuration leads to a form of segregation in the exposure dimensions.

The Bayes rule ties interaction probabilities to the likelihood of interaction in the following way:

$$\mathcal{L}(i|g) \;=\; \frac{\Pr[i] \cdot \pi_{gi}}{\pi_g}.$$

In our example, suppose that weights are defined as follows: $\Pr[l] = 2/4$, $\Pr[j] = 1/4$ and $\Pr[k] = 1/4$. The expected interaction profile can be computed as follows:

$$\begin{pmatrix} \pi_{g_1} \\ \pi_{g_2} \end{pmatrix} = \frac{2}{4}\begin{pmatrix} \pi_{g_1 l} \\ \pi_{g_2 l} \end{pmatrix} + \frac{1}{4}\begin{pmatrix} \pi_{g_1 j} \\ \pi_{g_2 j} \end{pmatrix} + \frac{1}{4}\begin{pmatrix} \pi_{g_1 k} \\ \pi_{g_2 k} \end{pmatrix} = \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}.$$

The interaction profiles are not equally distributed. In fact, one obtains:

$$\begin{pmatrix} \mathcal{L}_{lg_1} \\ \mathcal{L}_{lg_2} \end{pmatrix} = \begin{pmatrix} 2/4 \\ 2/4 \end{pmatrix}, \quad \begin{pmatrix} \mathcal{L}_{jg_1} \\ \mathcal{L}_{jg_2} \end{pmatrix} = \begin{pmatrix} 1/8 \\ 7/24 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mathcal{L}_{kg_1} \\ \mathcal{L}_{kg_2} \end{pmatrix} = \begin{pmatrix} 3/8 \\ 5/24 \end{pmatrix},$$

or, in matrix notation,

$$\mathcal{L} := \begin{pmatrix} 2/4 & 1/8 & 3/8 \\ 2/4 & 7/24 & 5/24 \end{pmatrix},$$

which shows that the sources of exposure are units j and k, given that $\mathcal{L}(l|g_1) = \mathcal{L}(l|g_2)$ holds. In fact, unit l_1 and l_2 interaction profiles coincide with the expected profile.

B Proofs

B.1 Proof of Proposition 1

Proof. The proof is made by construction. We first partition the set of all possible G-tuple $\{i_1, \ldots, i_G\}$ of individuals into two groups. There are some G-tuples gathering individuals that exclusively belong to subpopulation $\mathcal{N}_g(A)$, for each of the groups $g \in \mathcal{G}$. The remaining G-tuples belong, instead, to the overlapping set $\mathcal{O}$. This originate the first result: the Exposure Gini index is linearly separable into a within component plus the overlapping term. The latter is representable itself as a Gini index (because the whole population is taken into consideration in calculating it):

$$G_E(A) \;:=\; \text{Within term} \;+\; G_E(A|\mathcal{O}).$$

We now turn to the *within term*. Again, by linearity of the Gini index it is possible to separate the different observations by group, defined by $\mathcal{N}_g(A)$, such that for $i \in \mathcal{N}_g(A)$ it holds that $i \in \{i_1, \ldots, i_G\}$ only if $\{i_1, \ldots, i_G\} \in \mathcal{N}_g(A)$. An obvious requirement, always satisfied by definition of an allocation, is that $\mathcal{N}_g(A) \cap \mathcal{N}_m(A) = \emptyset$ for all groups $g \neq m$. As a result one obtains a comparison of G-tuples for all groups separately, for a total of G factors adding up to the within component.

Each of the G factors can be written as a sum of absolute values of determinants of a squared matrix of size G, which for simplicity is referred to by $\mathbf{D}$. For the G-tuple $\{i_1, \ldots, i_G\}$, this matrix is defined as $\mathbf{D} = (\boldsymbol{\ell}_{i_1}, \ldots, \boldsymbol{\ell}_{i_G})$. Note that within a chosen group g, $\mathbf{D}$ only depends upon the chosen G-tuple in $\mathcal{N}_g(A)$. The within term can be written as:

$$\text{Within term} \;=\; \sum_g \frac{1}{G!} \sum_{\{i_1,\ldots,i_m,\ldots,i_G\} \in \mathcal{N}_g(A)} |\det(\mathbf{D})|$$

For a chosen group (say g) and a given G-tuple (say the one including i_m), an element of the matrix $\mathbf{D}$ chosen in any position (say the one corresponding to row g and column i_m) is given by $\pi_{gi_m}(A)\mathbf{a}_{i_m}$.

Multiplication and division of the interaction probability vector by an appropriate conversion factor $\frac{\pi_g^e(A)}{(\sum_{i\in\mathcal{N}_g(A)} \xi_i(A))\pi_g^e(A|g)}$ does not produce any effect. The operation gives a new matrix, where a generic element in row g, column i_m is defined by

$$\frac{(\sum_{i\in\mathcal{N}_g(A)} \xi_i(A))\pi_g^e(A|g)}{\pi_g^e(A)} \frac{\xi_{i_m}(A|g)\pi_{gi_m}(A)}{\pi_g^e(A|g)}.$$

In compact form, one can substitute $\mathbf{D}$ with $\widehat{\mathbf{D}}$ in the calculation of the within term, to obtain:

$$\text{Within term} \;=\; \sum_g \frac{1}{G!} \sum_{\{i_1,\ldots,i_m,\ldots,i_G\} \in \mathcal{N}_g(A)} |\det(\boldsymbol{\alpha}_g \cdot \widehat{\mathbf{D}})|,$$

where

$$\boldsymbol{\alpha}_g := \operatorname{diag}\left(\frac{(\sum_{i\in\mathcal{N}_g(A)}\xi_i(A))\pi_1^e(A|g)}{\pi_1^e(A)},\dots,\frac{(\sum_{i\in\mathcal{N}_g(A)}\xi_i(A))\pi_G^e(A|g)}{\pi_G^e(A)}\right),$$

and $\widehat{\mathbf{D}} = \boldsymbol{\alpha}_g^{-1}\cdot\mathbf{D}$.

The determinant of the product of two matrices is the product of the respective determinants of the factors. Moreover, the determinant of a diagonal matrix is the product of elements on the diagonal. Few calculations show that $\det(\boldsymbol{\alpha}_g) = \alpha_g$, defined in the proposition. The value α_g only depends on the group index. Hence, the following result applies, which concludes the proof:

$$\text{Within term} \;=\; (\sum_{g\in\mathcal{G}}\alpha_g)\sum_{g\in\mathcal{G}}\frac{\alpha_g}{\sum_{g\in\mathcal{G}}\alpha_g}G_E(A|g).$$

■

B.2 Proof of Proposition 2

Proof. *Necessity part.* Consider matrix $\mathcal{L}\in\hat{\mathcal{M}}_{GG}$, by construction it could be obtained from I_G applying a finite sequence of elementary mixture of units/columns operations and permutation of units/columns. Note that $\mathcal{L} = I_G\mathcal{L}$, it follows that there is a finite sequence of elementary mixture of units/columns transformations and permutations of units/columns that allow to construct $\mathcal{L}$ starting from I_G. These transformations can be summarized in terms of matrices multiplications, by considering permutations matrices in $\mathcal{P}_G$ and matrices $T(\alpha,i,j)\in\mathcal{M}_{GG}$ such that

$$T(\alpha,i,j) = \begin{array}{c} \begin{array}{cccccc} 1 & \dots & i\dots & \dots & j & \dots \end{array} \\ \begin{bmatrix} 1 &&&&& \\ & 1 &&&& \\ && \alpha && 1-\alpha & \\ &&& 1 && \\ &&&& 1 & \\ &&&&& 1 \end{bmatrix} \end{array} \tag{1}$$

where all empty cells should be occupied by zeros. Let index by k the elements of the sequence of transformations $T(\alpha_k,i_k,j_k)$, where $\alpha_k\in(0,1]$ denotes the mixture coefficient and i_k,j_k denote the columns involved in the mixture at stage k. It follows that either

$$\mathcal{L} = I_G \Pi_G^1 \cdot \prod_k T(\alpha_k, i_k, j_k) \cdot \Pi_G^2 \tag{2}$$

or matrix $\mathcal{L}$ can be decomposed similarly by also inserting some permutations matrices among the elements of the sequence of transformations $T(\alpha_k, i_k, j_k)$.

We consider first the case in (2). The first permutation matrix Π_G^1 in the sequence can be considered as a column permutation of the matrix I_G. Note that Π_G^1 could also coincide with an identity matrix and therefore lead to no effect on the sequence of operations. Thus, $E^G(\mathcal{L}) = E^G\left(I_G \Pi_G^1 \cdot \prod_k T(\alpha_k, i_k, j_k) \cdot \Pi_G\right)$. Making use of axiom UA it follows that $E^G(\mathcal{L}) = E^G\left(I_G \Pi_G^1 \cdot \prod_k T(\alpha_k, i_k, j_k)\right)$. Note that if we apply the transformation associated with $T(\alpha_k, i_k, j_k)$ to matrix $I_G \Pi_G^1$ we obtain $I_G \Pi_G^1 \cdot T(\alpha_k, i_k, j_k)$. Recall that according to axiom UA in combination with axiom N, it follows that $E^G(I_G \Pi_G^1) = 1$. Then by applying axiom EM we obtain that $E^G[I_G \Pi_G^1 \cdot T(\alpha_k, i_k, j_k)] = \alpha_k \cdot E^G(I_G \Pi_G^1) = \alpha_k$. By repeated application of axiom EM one obtains that

$$E^G(\mathcal{L}) = E^G\left(\prod_k T(\alpha_k, i_k, j_k)\right) = \prod_k \alpha_k. \tag{3}$$

Note that by construction $\alpha_k = \det\left(T(\alpha_k, i_k, j_k)\right)$. Moreover, by making use of the property that the determinant of the product of square matrices is equal to the product of the determinant of the matrices, one obtains that

$$\prod_k \alpha_k = \det\left[\prod_k T(\alpha_k, i_k, j_k)\right]. \tag{4}$$

Note however that all elements α_k are positive, and thus the above relation holds also if we consider the absolute value of the determinant. This in general should be the case if we consider matrix $\mathcal{L}$ that could be obtained permuting either matrix I_G and/or matrix $\prod_k T(\alpha_k, i_k, j_k)$. These operation may invert the sign of the determinant and therefore we may have that combining (3) and (4) one obtains

$$|\det \mathcal{L}| = \det\left[\prod_k T(\alpha_k, i_k, j_k)\right] = \prod_k \alpha_k = E^G(\mathcal{L}).$$

The above consideration could be extended to the case where permutation matrices in $\mathcal{P}_G$ are inserted into the sequence of operations $T(\alpha_k, i_k, j_k)$. These insertions do not affect the final result, but allow to enrich the set of matrices that can be reached by the combinations of operations.

So far we have considered the case where $E^G(\mathcal{L}) > 0$, by continuity of E^G we can also approach the situations where $E^G(\mathcal{L}) = 0$, these can be obtained as limiting cases where some $\alpha_k \to 0$.

In order to complete the proof it is left to verify the uniqueness of the index and the sufficiency part.

Suppose that the sequence of matrices $T(\alpha_k, i_k, j_k)$ and of the permutation matrices Π_G is not unique. By construction then either it leads to the same value $|\det \mathcal{L}|$ or it not possible that the sequence leads to $\mathcal{L}$ because $|\det \mathcal{L}|$ is uniquely defined.

Sufficiency. The index $E^G(\mathcal{L}) = |\det \mathcal{L}|$ satisfies all the three axioms UA, N and EM. In fact the absolute value of the determinant is not affected by permutation of the columns of a matrix (axiom UA) and it equals 1 if $\mathcal{L} = I_G$ (axiom N). To prove that also axiom EM is satisfied, we need to combine some properties of the determinants.

Recall in the definition of axiom EM the notation for

$$\mathcal{L}(\alpha, i, j) = (\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \alpha\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_j + (1-\alpha)\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_G).$$

First note that if one column of a matrix is multiplied by α also its determinant is multiplied by α. It then follows that

$$\begin{aligned}\det \mathcal{L}(\alpha, i, j) &= \det(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \alpha\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_j + (1-\alpha)\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_G) \\ &= \alpha \det(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_j + (1-\alpha)\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_G).\end{aligned}$$

Then recall that the determinants are multilinear functionals and therefore

$$\begin{aligned}& \det(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_j + (1-\alpha)\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_G) \\ = \;& \det(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, , ..., \boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_j, ..., \boldsymbol{\ell}_G) + (1-\alpha)\det(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_G).\end{aligned}$$

Note that the last determinant equals 0 because two columns are identical to $\boldsymbol{\ell}_i$. It then follows that

$$\begin{aligned}& \det(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_j + (1-\alpha)\boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_G) \\ = \;& \det(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \boldsymbol{\ell}_i, ..., \boldsymbol{\ell}_j, ..., \boldsymbol{\ell}_G) = \det \mathcal{L}.\end{aligned}$$

To summarize, $\det \mathcal{L}(\alpha, i, j) = \alpha \cdot \det \mathcal{L}$. The above considerations are not affected if one considers the absolute value of the determinant, thereby leading to $E^G(\mathcal{L}(\alpha, i, j)) = |\det \mathcal{L}(\alpha, i, j)| = \alpha \cdot |\det \mathcal{L}| = \alpha E^G(\mathcal{L})$ as required by axiom EM. ■

B.3 Proof of Corollary 1

Proof. By direct application of the definition of axiom D to the result in Proposition 2, note first that is not necessary to consider matrices where at least one row is made of zeros, because in this case the associated $\left(\Pi_{g=1}^{G} \lambda_g^{\{i_1,i_2,\ldots,i_G\}}\right) = 0$. We should therefore only focus on matrices in $\hat{\mathcal{M}}_{GG}$. For these matrices by combining axiom D with the result in Proposition 2 we obtain

$$E^N(\mathcal{L}) = \sum_{\{i_1,i_2,\ldots,i_G\}\subseteq\mathcal{N}_O} \left(\Pi_{g=1}^{G} \lambda_g^{\{i_1,i_2,\ldots,i_G\}}\right) \cdot \left|\det(\tilde{\boldsymbol{\ell}}_{i_1}, \tilde{\boldsymbol{\ell}}_{i_2}, \ldots, \tilde{\boldsymbol{\ell}}_{i_G})\right|.$$

However, note that by construction

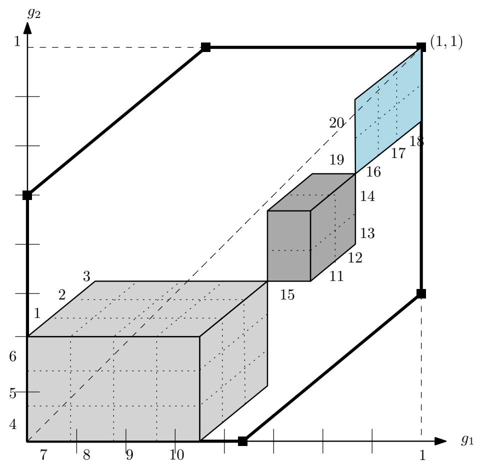
$$\left|\det(\tilde{\boldsymbol{\ell}}_{i_1}, \tilde{\boldsymbol{\ell}}_{i_2}, \ldots, \tilde{\boldsymbol{\ell}}_{i_G})\right| = |\det(\boldsymbol{\ell}_{i_1}, \boldsymbol{\ell}_{i_2}, \ldots, \boldsymbol{\ell}_{i_G})| \cdot \left(\Pi_{g=1}^{G} \lambda_g^{\{i_1,i_2,\ldots,i_G\}}\right)^{-1}.$$

After simplifying in the $E^N(\mathcal{L})$ formula we obtain

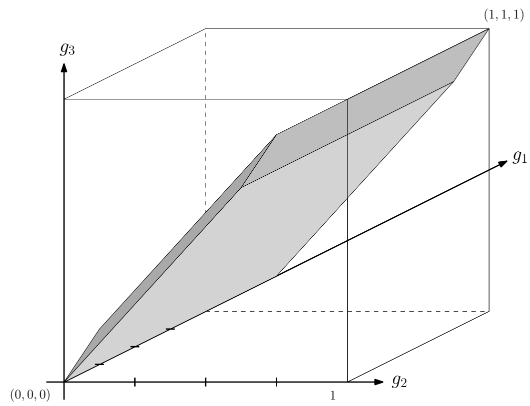
$$E^N(\mathcal{L}) = \sum_{\{i_1,i_2,\ldots,i_G\}\subseteq\mathcal{N}_O} |\det(\boldsymbol{\ell}_{i_1}, \boldsymbol{\ell}_{i_2}, \ldots, \boldsymbol{\ell}_{i_G})|.$$

Here the ordered distribution of the G units in $\mathcal{N}_O$ is taken into account. If we allow all possible permutations of these units as in $\mathcal{N}(\mathcal{L})$ then we obtain for each ordered set of units $G!$ times the same index $|\det(\boldsymbol{\ell}_{i_1}, \boldsymbol{\ell}_{i_2}, \ldots, \boldsymbol{\ell}_{i_G})|$ that, being expressed in absolute terms is not modified by permutation of the columns. These considerations lead to the final result. ■

C Figures and tables



(a) Case g_1 and g_2



(b) Case g_1, g_2 and g_3

Figure 1: Segregation Zonotope and the Gini Exposure index. In the first graph, the population of 20 individuals is partitioned according to the group of origin: $\mathcal{N}_{g_1}(A) = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $\mathcal{N}_{g_2}(A) = \{11, 12, 13, 14, 15\}$ and $\mathcal{N}_{g_3}(A) = \{16, 17, 18, 19, 20\}$. The share of overall segregation as experienced exclusively by the members of the three groups is given by the area of the three polytopes identified in panel (a).

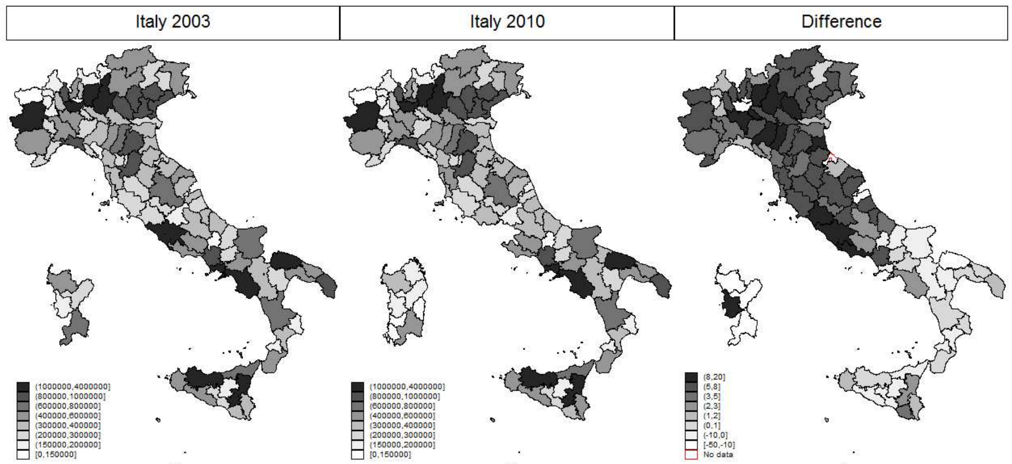


Figure 2: Population by province and year, and its growth rate (in %).

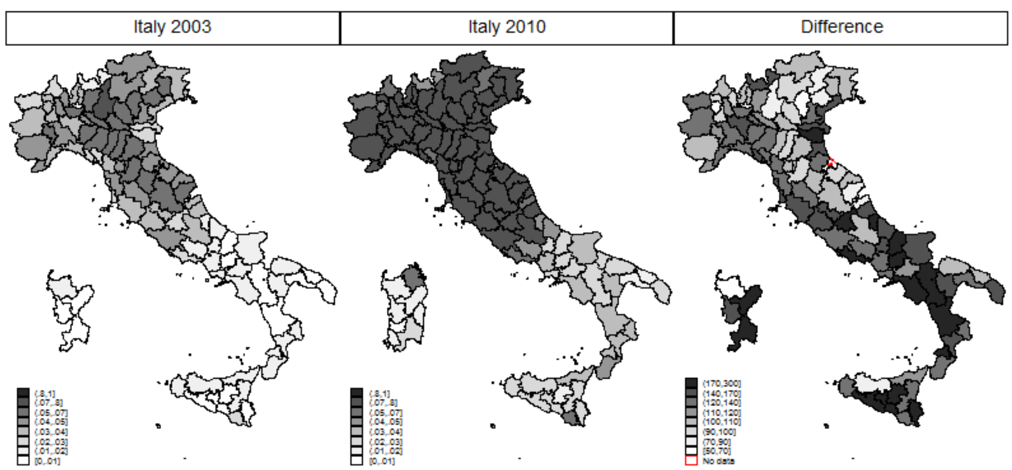


Figure 3: Share of immigrants by province and year, and its growth rate (in %).

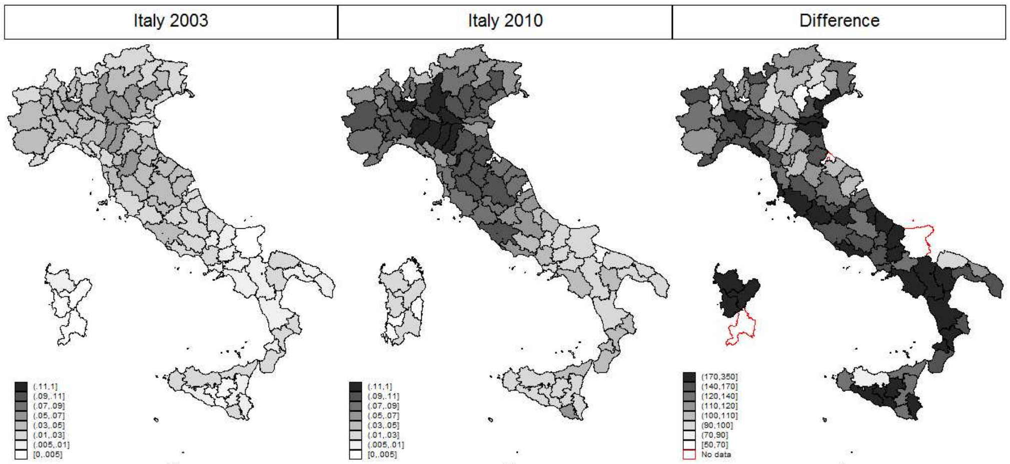


Figure 4: Expected interaction probability with immigrants from *low HDI* countries, by province and year, and its growth rate (in %).

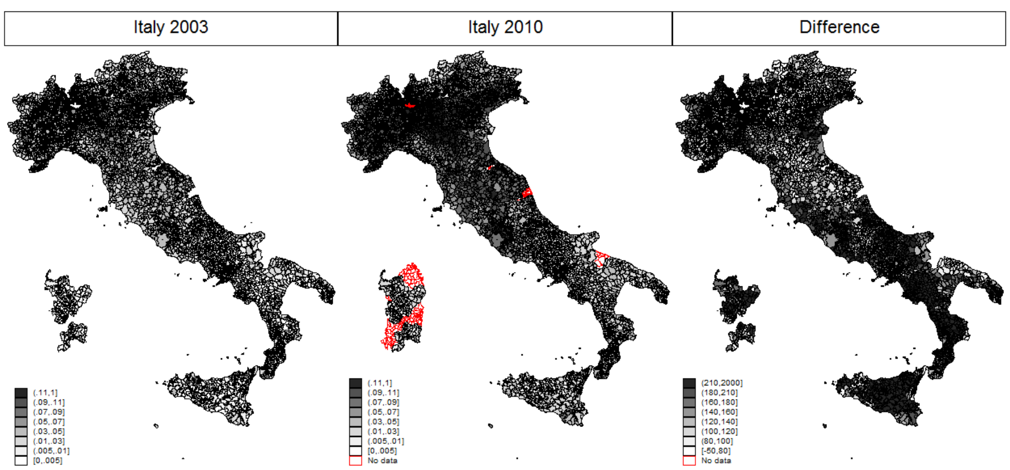


Figure 5: Interaction probability with immigrants from *low HDI* countries, by municipality and years, and its growth rate (in %).

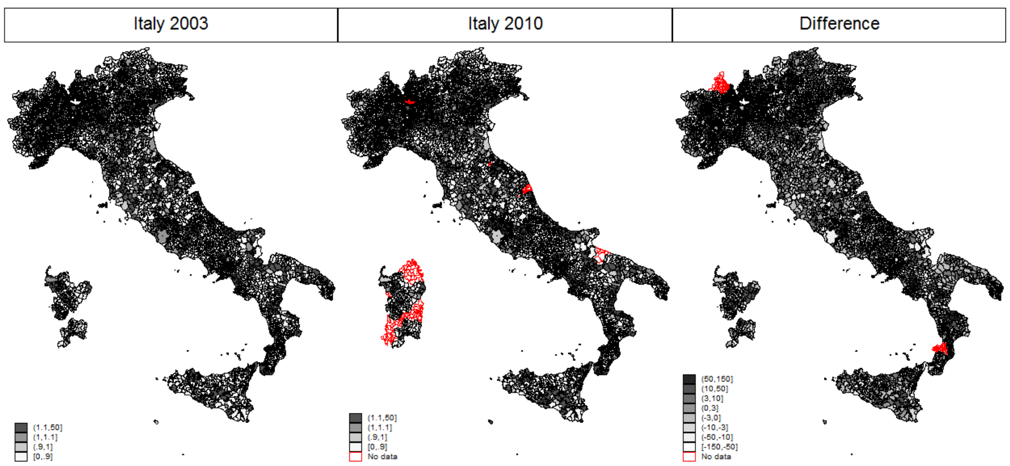


Figure 6: Disproportionality in interaction probability (with respect to the expected interaction) with immigrants from *low HDI* countries, by municipality and year, and the change in ranking (relative to the position of the municipalities in 2003 within the same province).

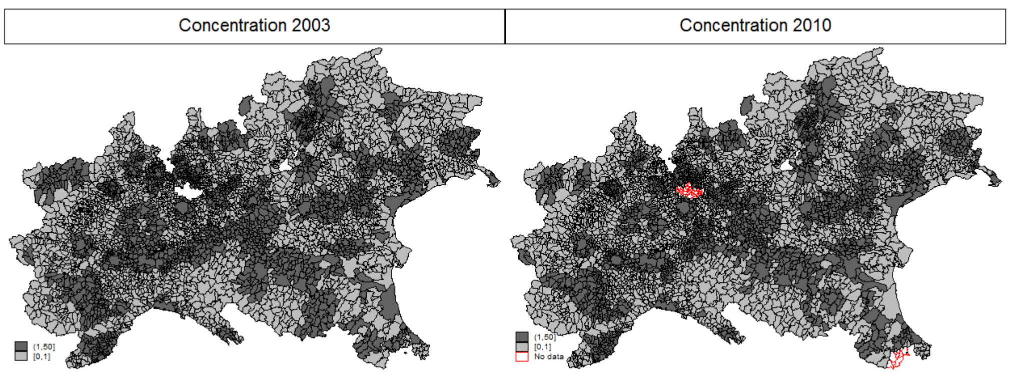


Figure 7: Disproportionality in interaction probability (with respect to the expected interaction) with immigrants from *low HDI* countries, by municipality in 2010 for the North Italy macro-region. Higher concentration ($a_k > 1$) is interpreted as the disproportion between the interaction with immigrants from low HDI countries with respect to the expected interaction. The expected interaction probability is calculated at province level.

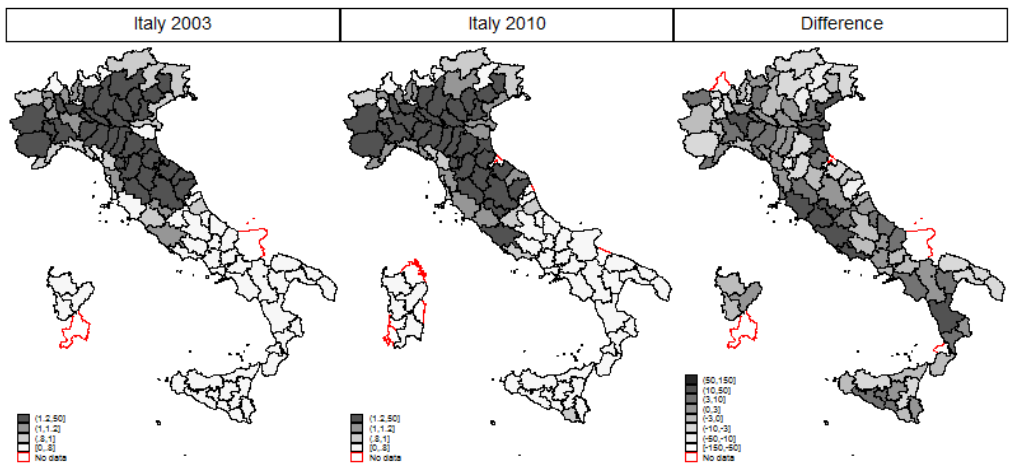


Figure 8: Disproportionality in *expected* interaction probability with immigrants from *low HDI* countries, by province and year, and it growth rate (in %).

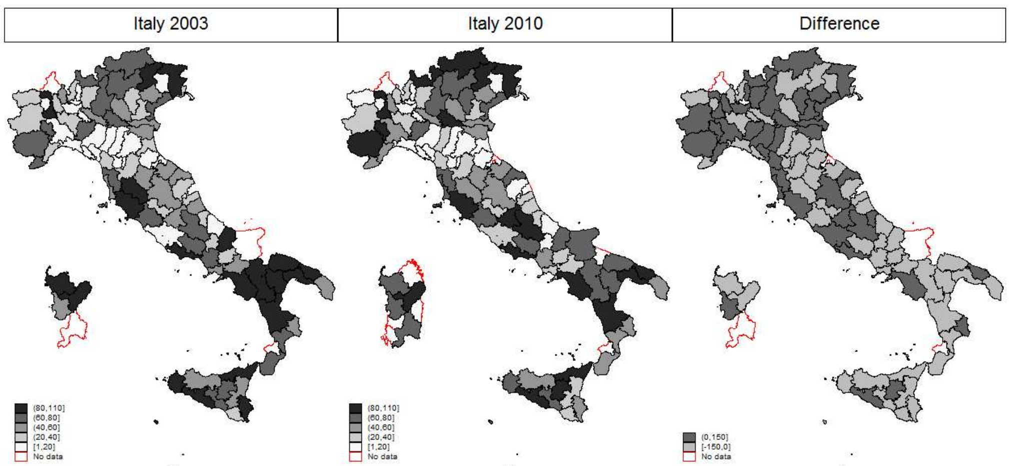


Figure 9: Ranking of Italian provinces according to the multi-group Gini Exposure index, by year. Differences are reported for provinces where segregation is increased (positive rank changes, in dark gray) and where segregation is decreased (negative rank changes, in pale gray).

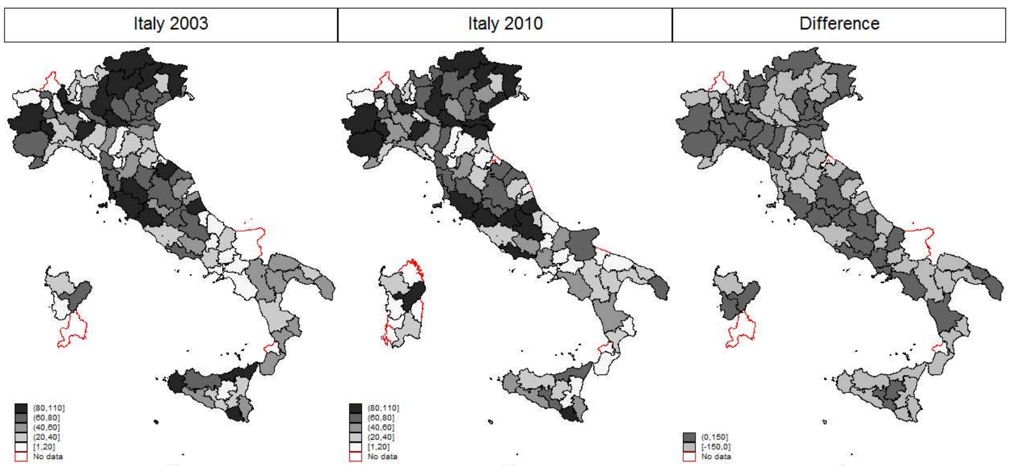


Figure 10: Ranking of Italian provinces according to the multi-group Mutual information index, by year. Differences are reported for provinces where segregation is increased (positive rank changes, in dark gray) and where segregation is decreased (negative rank changes, in pale gray).

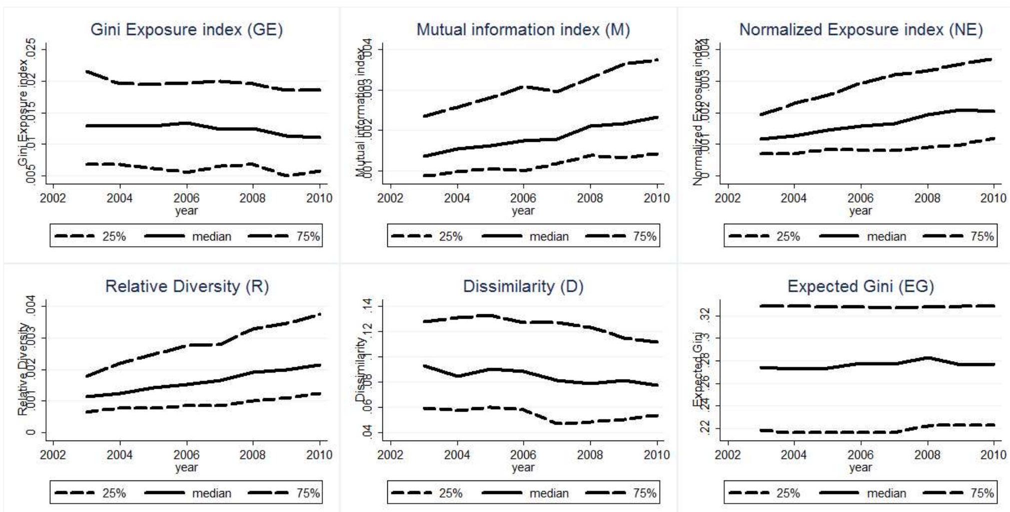


Figure 11: Dynamics of the six exposure segregation indices in the period 2003 to 2010, for the 101 Italian provinces. For each index and each year are reported the values of the index associated to the provinces in the first quartile (25%), *median* and third quartile (75%) of the ranking produced by the index in that year.

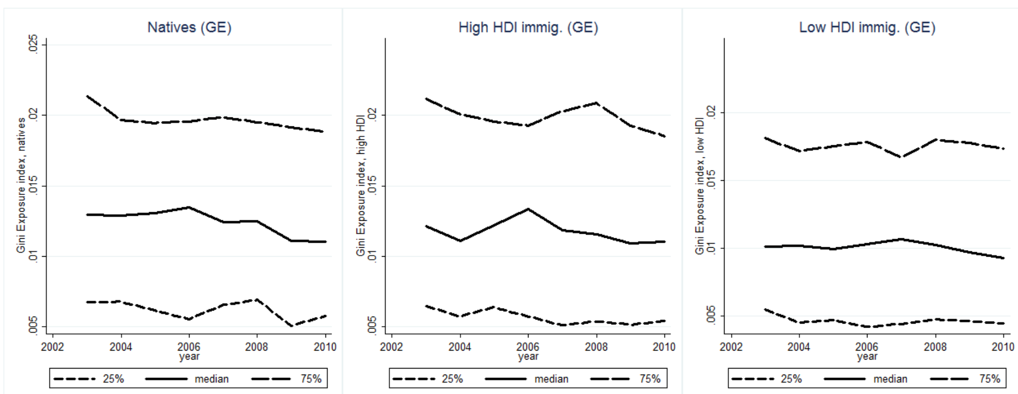
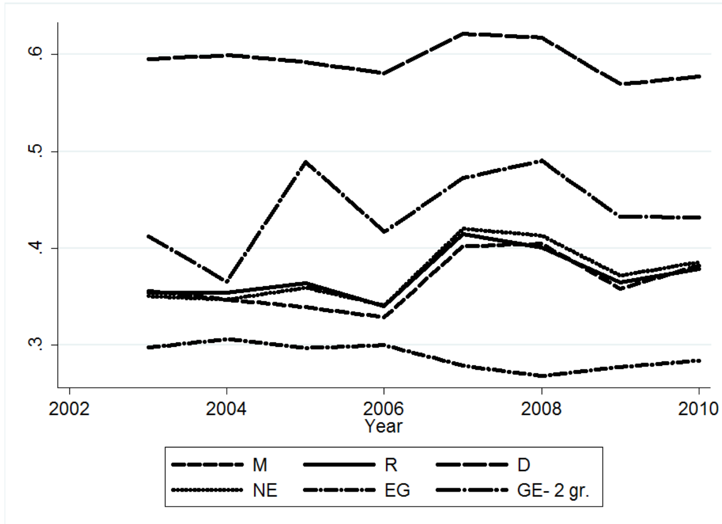
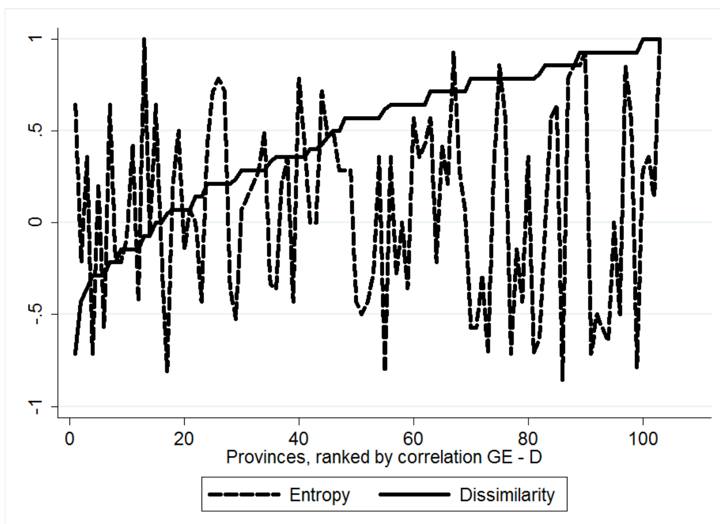


Figure 12: Dynamics of the decomposition of the G_E index in the period 2003 to 2010, for the 101 Italian provinces. For each subgroups and each year are reported the values of the index associated to the provinces in the first quartile (25%), *median* and third quartile (75%) of the ranking produced by the index in that year.



(a) Across periods



(b) Across provinces

Figure 13: Kendall's τ_b correlation coefficient of Mutual Information index (M), Relative interaction (R), Dissimilarity (D), Normalized Exposure (NE), Expected Gini (EG) and Gini Exposure for two groups (GE) indices with the multi-group Gini Exposure index. See the note of Table 2 for further details. Correlations in panel (a) are calculated for each year for the whole set of realizations of the indices across provinces (on average 102 observations per year), while correlations in panel (b) are calculated for each province using the data of the years 2003/2010 (eight years, negative correlations are statistically zero at 5% confidence level). Provinces are ordered by increasing magnitude of the correlation between G_E and D.

Table 1: Descriptive statistics

Year	2003	2004	2005	2006	2007	2008	2009	2010
Provinces (N)	103	103	103	107	107	107	107	110
Municipalities (N)	8100	8101	8101	8101	8101	8101	8100	8094
Municipalities by province	79	79	79	76	76	76	76	74
Population (total, mil)	57.9	58.5	58.8	59.1	59.6	60.0	60.3	60.6
Population by municipality	6267	6342	6375	6441	6500	6549	6586	6417
Population by province	562022	567596	570405	552629	557190	561169	563928	551149
Low HDI Immigrants								
Share	0.029	0.035	0.039	0.043	0.050	0.057	0.062	0.067
	0.017	*0.021*	*0.023*	*0.025*	*0.028*	*0.031*	*0.032*	*0.035*
Interaction probability	0.029	0.036	0.040	0.044	0.045	0.052	0.063	0.066
	0.018	*0.022*	*0.025*	*0.027*	*0.030*	*0.033*	*0.034*	*0.037*
High HDI immigrants								
Share	0.005	0.006	0.006	0.007	0.007	0.008	0.008	0.008
	0.003	*0.004*	*0.004*	*0.004*	*0.005*	*0.005*	*0.005*	*0.005*
Interaction probability	0.005	0.006	0.006	0.007	0.006	0.007	0.008	0.008
	0.004	*0.004*	*0.004*	*0.004*	*0.005*	*0.005*	*0.005*	*0.005*
Immigrants concentration as a proportion of cases where $a > 1$:								
By municipality	0.419	0.409	0.407	0.406	0.419	0.419	0.415	0.413
	0.493	*0.492*	*0.491*	*0.491*	*0.493*	*0.493*	*0.493*	*0.492*
By province	0.447	0.456	0.437	0.458	0.570	0.570	0.495	0.536
	0.500	*0.501*	*0.498*	*0.501*	*0.497*	*0.497*	*0.502*	*0.501*
Polarization ($.9 < a < 1.1$)	0.078	0.097	0.107	0.112	0.131	0.140	0.131	0.155
	0.269	*0.298*	*0.310*	*0.317*	*0.339*	*0.349*	*0.339*	*0.363*

Data: ISTAT, demographic statistics, years 2003/2010.

Notes: Interaction probabilities constructed with a spatial *biweighted quadratic kernel*, boundary distance is 20km. Standard deviations are reported in italics.

Table 2: Rank correlation between segregation indices

Index	G_E	M	R	D	NE	EG	G_E2	A
Gini Exposure (G_E)	1	0.484	0.501	0.773	0.506	0.410	0.614	0.407
Mutual Information (M)	0.347	1	0.901	0.498	0.859	0.251	0.335	0.073
Relative Diversity (R)	0.356	0.750	1	0.665	0.988	0.305	0.499	0.121
Dissimilarity (D)	0.593	0.356	0.486	1	0.686	0.311	0.767	0.379
Normalized Exposure (NE)	0.357	0.709	0.919	0.502	1	0.325	0.534	0.109
Expected Gini (EG)	0.285	0.168	0.207	0.222	0.221	1	0.272	0.140
Gini Exposure, pair (G_E2)	0.437	0.230	0.350	0.576	0.377	0.184	1	0.341
Atkinson (A)	0.278	0.050	0.078	0.259	0.070	0.091	0.233	1
Mean (diagonal excluded)	0.379	0.373	0.427	0.506	0.221	0.321	0.341	0.151

Data: by ISTAT, demographic statistics, years 2003/2010.

Notes: Kendall (τ_b, below the diagonal) and Spearman (ρ, above the diagonal) rank correlation coefficients of spatial segregation indices calculated at province level, years 2003 to 2010. The total number of observations is 808 (708 for the Atkinson index, year 2010 is chosen to set the index weighting scheme). All coefficients are significantly different from zero at 1% level. Universe is set according to the ISTAT statistical definition of Italian provinces (reduced to 101 here), and indices are computed with information at municipality level. Provinces created or destroyed after 2003 are excluded from the sample. Social groups are mutually exclusive: natives (Italian nationality), immigrants from low HDI countries and immigrants from high HDI countries.

Table 3: The impact of demographic factors on the rank correlation

Dependent var.:	Num. of concordances (G_E and D)			Num. of concordances (G_E and M)		
	(1)	(2)	(3)	(4)	(5)	(6)
Pop. total	-0.000	-0.000	-0.000+	0.000	0.000	0.000+
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pop. natives	0.000	0.000	0.000+	-0.000	-0.000	-0.000+
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pop immigrants	0.000	0.000	0.000*	-0.000	-0.000	-0.000+
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share imm.		-124.199	556.237		151.561	-638.273
		(86.62)	(436.70)		(123.50)	(569.17)
Proportion imm.		5.353	-596.959		-164.355	676.974
		(83.73)	(432.16)		(119.84)	(565.11)
Proportion ratio			16.256			-133.832**
			(44.00)			(65.59)
Rank (prop. ratio)			-14.540			125.770**
			(41.89)			(63.28)
Rank by G_E			0.081***			-0.046*
			(0.02)			(0.02)
year==2004		-0.626	-0.782		6.876***	6.691***
		(2.32)	(2.28)		(2.45)	(2.32)
year==2005		-0.394	-0.458		1.758	1.160
		(2.37)	(2.31)		(2.46)	(2.26)
year==2006		0.539	-0.796		9.410***	8.059***
		(2.38)	(2.36)		(2.63)	(2.41)
year==2007		-0.130	-0.921		8.784***	8.325***
		(2.45)	(2.46)		(2.60)	(2.44)
year==2008		-0.578	-1.812		7.226***	6.987***
		(2.51)	(2.50)		(2.59)	(2.47)
year==2009		4.802*	2.222		10.515***	9.260***
		(2.47)	(2.51)		(2.66)	(2.51)
year==2010		-0.027	-2.109		7.852***	6.243**
		(2.69)	(2.73)		(2.82)	(2.73)
Constant	60.236***	65.082***	55.332***	54.055***	47.755***	54.232***
	(0.82)	(2.15)	(3.69)	(0.95)	(2.34)	(4.07)
Provinces (8 years)	847	847	795	847	847	795
R^2	0.004	0.03	0.03	0.007	0.04	0.06
p-value model	0	0	0	0	0	0

$+\ p < 0.15$, $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Data: by ISTAT, demographic statistics, years 2003/2010.

Notation: Regression (OLS) of the number of *concordant* observations on predictors, controlling by year FE. An observation is a province in a given year. Let $\rho_i(I)$ be the rank of province i in a given year produced by the index I. We say that, within a given year, provinces i and j are concordant with respect to indices G_E and D if $(\rho_i(G_E) - \rho_j(G_E))\,(\rho_i(D) - \rho_j(D)) > 0$. The dependent variable is the number of cases of concordance of each province i in a given year, for both pairs (G_E, D) (models (1)-(3)) and (G_E, M) (models (4)-(6)). These values, normalized by the maximum number of comparisons, give the Kendall's τ_b coefficient.