Universidade Vigo

Departamento de Economía Aplicada
https://webx06.webs.uvigo.es

Working Paper: 26/02. June 2026

Building Multigroup Segregation Indices: Evenness and Representativeness

Olga Alonso-Villar

Coral del Río

Coordinator: Olga Alonso-Villar
ovillar@uvigo.es

Building Multigroup Segregation Indices: Evenness and Representativeness*

Olga Alonso-Villar[#] and Coral del Río

ECOBAS, Universidade de Vigo

Abstract

To quantify segregation in increasingly diverse societies, researchers need tools that enable them to measure the separation between more than two social groups. We address this issue by conceptualizing the measurement of overall segregation in a novel way. Our approach involves quantifying the degree to which each social group is unevenly distributed across organizational units using indicators that possess desirable properties and then aggregating those values in an appropriate manner. This ensures that the multigroup indices thus obtained satisfy basic properties well established in the literature. Using this procedure, we build a family of multigroup segregation indices Φ_α that captures the inequality notions of the generalized entropy family for values of the inequality aversion parameter $\alpha > 0$. This allows us to expand the range of available measures simply by changing the value of a parameter, making it easier to incorporate different value judgements into the measurement. This family generalizes the mutual information index and the unnormalized version of the square coefficient of variation, and also includes multigroup segregation indices that can be used when researchers are interested in inequality aversion values other than 1 and 2 (i.e., those associated with the aforementioned indices). In particular, this family comprises the $\Phi_{0.5}$ index, which shares the same inequality aversion as the square root index, but, unlike the latter, can be used in a multigroup context. These indices can be thought of in terms of both evenness and representativeness and are useful tools for empirical analysis thanks to their decomposability properties.

JEL Classification: D63

Keywords: Segregation; Multigroup Indices; Inequality; Evenness; Representativeness

---

* We want to thank participants at the XII EQUALITAS (Economics of Inequality and Poverty Analysis) Workshop at the *Universidad Complutense de Madrid* for helpful comments.

Correspondence address: Universidade de Vigo; Facultade de CC. Económicas; Departamento de Economía Aplicada; Campus Lagoas-Marcosende s/n; 36310 Vigo; Spain. Tel.: +34 986812507; fax: +34 986812401; e-mail: ovillar@uvigo.es.

1. Introduction

To address the measurement of segregation in increasingly diverse societies, we need tools that enable us to measure the separation among more than two population groups, as occurs in the case of segregation by race/ethnicity. The literature offers several segregation indices that can be used in multigroup contexts, primarily from an evenness perspective that focuses on the extent to which social groups are unevenly distributed across organizational units, such as schools, neighborhoods, or occupations (Boisso et al., 1994; Silber, 1992; Reardon and Firebaugh, 2002; Frankel and Volij, 2011).

Some of these multigroup segregation measures are based on inequality measures (Alonso-Villar and Del Río, 2025). Thus, the Gini coefficient is both behind the multigroup segregation index *G* developed by Reardon and Firebaugh (2002) and that proposed by Boisso et al. (1994). Likewise, two members of the generalized entropy family of inequality indices—which are obtained when the inequality aversion parameter α is equal to 1 and 2—are behind other multigroup segregation measures (Del Río and Alonso-Villar, 2022). So, the Theil 1 index underlies both the mutual information index *M* and the information theory index *H* proposed by Theil and Finizza (1971), while the Theil 2 index underlies the square coefficient of variation *C* developed by Reardon and Firebaugh (2002).[1]

The generalized entropy family has also been used to measure segregation in the case of two social groups. Thus, Hutchens (2001, 2004) proposed a family of binary segregation measures drawing on the generalized entropy family for inequality aversion values $\alpha \in (0,1)$, although only for $\alpha = 0.5$ the index treats the two groups symmetrically, a desirable property usually invoked to measure segregation. This index, called the square root index, has been widely used to measure school segregation between two groups.[2]

However, as far as we know, in a multigroup context, no indices have been put forward to quantity segregation based on the generalized entropy family with inequality aversion

[1] The *M* and *H* indices—the latter being a normalized version of the former—were proposed by Theil and Finizza following a different approach based on the concept of entropy. These authors did not call the former the mutual information index, although this name has become popular (Frankel and Volij, 2011).

[2] The binary indices that he develops for $\alpha = 1$ and 2 (Hutchens, 1991) also do not treat the two groups symmetrically and can only be used when the reference group against which the other group is compared does not have zero values in any unit, which may not be the case in some empirical analyses.

parameter values other than $\alpha = 1$ and 2. Filling this gap would enable researchers to have a broader set of value judgements in the measurement of segregation, as traditionally done in the measurement of economic inequality. Furthermore, beyond the theoretical interest this may hold, in empirical analysis this would allow us to assess the robustness of the results against different inequality views simply by changing the value of α.

This paper addresses this issue by conceptualizing the measurement of overall segregation in a novel way that departs from what the literature has done to derive multigroup indices. We propose measuring overall segregation as the weighted average of the unevenness of the mutually exclusive social groups into which society is partitioned. This involves quantifying each group's unevenness (i.e., the degree to which a group is unevenly distributed among the units) using indicators that satisfy desirable properties (Alonso-Villar and Del Río, 2010; Del Río and Alonso-Villar, 2022) and then aggregating those values in an appropriate manner. Our conceptualization goes a step beyond the methods proposed by Reardon and Firebaugh (2002) to systematize the derivation of multigroup segregation measures (most of which had been proposed in the literature from different traditions). Our approach, based on the appropriate measurement of the unevenness of social groups, ensures that the multigroup indices thus obtained satisfy basic properties well established in the literature for measuring segregation from an evenness perspective (James and Taeuber, 1985; Hutchens, 1991, 2001; Reardon and Firebaugh, 2002).

Following this procedure, we build a family of multigroup segregation indices Φ_α that captures the inequality notions of the generalized entropy family for values of the inequality aversion parameter $\alpha > 0$. In addition to generalizing the mutual information index M and the unnormalized version of C, this family includes, among other new indices, the index $\Phi_{0.5}$, which shares the same inequality aversion as the square root index, but can be used in a multigroup context. We show that the Φ_α indices can also be interpreted in terms of the unrepresentativeness of units (an approach posed by Frankel and Volij, 2011), which is very useful in empirical analysis.

This paper is structured as follows. In Section 2, we propose a new method for deriving multigroup overall segregation measures. In Section 3, we put forward and evaluate a family of indices parameterized with an inequality aversion parameter $\alpha > 0$. In Section 4, we show that the members of this family can also be written in terms of the unrepresentativeness of the units using indicators of unrepresentativeness that satisfy several properties that we develop here. In addition, we show the usefulness of this view of segregation when units are grouped into clusters. Finally, Section 5 offers the main conclusions.

2. A New Procedure for Building Multigroup Overall Segregation Measures

In this section, we provide a procedure for building multigroup segregation measures that differs from previous approaches in the literature (Boisso et al., 1994; Silber, 1992; Reardon and Firebaugh, 2002). As already mentioned, we propose to measure overall segregation as the weighted average of the unevenness of each social group into which the total population is partitioned (with weights equal to their population shares) using indicators of unevenness that satisfy basic properties.

In what follows, after introducing the notation, we first present the properties established in the literature for measuring the unevenness of a group in a multigroup context (Alonso-Villar and Del Río, 2010; Del Río and Alonso-Villar, 2022). Next, we show that the overall segregation measures we build using such unevenness indicators meet the basic properties established in the literature (Reardon and Firebaugh, 2002).

2.1 Notation

Suppose we have a society with $G \geq 2$ mutually exclusive social groups and $J > 1$ organizational units (schools, neighborhoods, occupations, etc.) among which each group g $(g = 1, \ldots, G)$ is distributed. We denote by vector $\boldsymbol{t} \equiv (t_1, t_2, \ldots, t_J)$ the distribution of the total population, T, across these units, where $t_j > 0$ represents the number of individuals in unit j $(j = 1, \ldots, J)$ and $T = \sum_j t_j$. Vector $\boldsymbol{c^g} \equiv (c_1^g, c_2^g, \ldots, c_J^g)$ represents the distribution of group g across these units, where $c_j^g \geq 0$ is the number of group g's individuals in unit j and $c_j^g \leq t_j$. The total number of individuals in unit j can be expressed as $t_j = \sum_g c_j^g$ whereas the total number of individuals of group g is $C^g = \sum_j c_j^g$.

Matrix E below represents the distributions of the groups across units. If we add up the elements in each row, we get the size of each group. If, instead, we add up the elements in each column, we get the size of each unit.

$$\begin{array}{c} G \text{ groups} \times J \text{ units} \\ E = \begin{bmatrix} c_1^1 & \cdots & c_J^1 \\ \vdots & \vdots & \\ c_1^G & \cdots & c_J^G \end{bmatrix} \end{array} \rightarrow \begin{bmatrix} \sum_j c_j^1 = C^1 \\ \vdots \\ \sum_j c_j^G = C^G \end{bmatrix}$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$\begin{bmatrix} \sum_g c_1^g = t_1 & \cdots & \sum_g c_J^g = t_J \end{bmatrix} \rightarrow T = \sum_g C^g = \sum_j t_j$$

2.2 Measuring the Unevenness of a Group in a Multigroup Context: What Do We Know?

Drawing on the literature on inequality, Alonso-Villar and Del Río (2010) propose measuring the unevenness of a group g by comparing its distribution across units, i.e., $\left(\frac{c_1^g}{C^g}, \ldots, \frac{c_J^g}{C^g}\right)$, with the distribution of the total population across the same units, i.e., $\left(\frac{t_1}{T}, \ldots, \frac{t_J}{T}\right)$, and put forward several basic requirements for any indicator $\phi: D \rightarrow \mathbb{R}$ that captures this unevenness, where $D = \bigcup_{J>1}\left\{(\boldsymbol{c^g}; \boldsymbol{t}) \in \mathbb{R}^J{}_+ \times \mathbb{R}^J{}_{++}: c_j^g \leq t_j \forall j\right\}$ (see also Del Río and Alonso-Villar, 2022). They call ϕ a local segregation index (from an evenness perspective) to distinguish it from overall segregation indices. We assume that ϕ is a continuous function that is equal to zero when and only when the group is evenly distributed among the units (i.e., $\frac{c_j^g}{C^g} = \frac{t_j}{T}\ \forall j$) and is positive otherwise. Below, we present this approach, upon which we subsequently construct our proposal.

To understand the connection between unevenness and inequality, it is helpful to keep in mind that measuring the unevenness of group g is like measuring the inequality of an "income" distribution $(\underbrace{\frac{c_1^g}{t_1},\dots,\frac{c_1^g}{t_1}}_{t_1 \text{ individuals}},\dots,\underbrace{\frac{c_J^g}{t_J},\dots,\frac{c_J^g}{t_J}}_{t_J \text{ individuals}})$ in a world of T individuals and a total "income" of C^g. Note that this income distribution is well-defined regardless of whether the group is present in a unit or not, since the reference we compare vector $\boldsymbol{c^g}$ to is vector $\boldsymbol{t}$, whose components are strictly positive.[3]

Basic Properties for Measuring the Unevenness of a Group

Next, we state the four properties required of any local segregation index ϕ (Alonso-Villar and Del Río, 2010) and provide examples to help with their interpretation using the following matrix E:

$$\begin{array}{cc}
 & \begin{array}{cccc}\text{Unit 1} & \text{Unit 2} & \text{Unit 3} & \text{Unit 4}\end{array} \\
\begin{array}{c}\text{Group 1}\\ \text{Group 2}\\ \text{Group 3}\end{array} & \begin{bmatrix} 2 & 3 & 5 & 10 \\ 3 & 7 & 10 & 20 \\ 10 & 10 & 5 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} C^1 = 20 \\ C^2 = 40 \\ C^3 = 30 \end{bmatrix} \\
 & \downarrow \qquad\qquad\qquad\qquad \downarrow \\
 & [t_1 = 15 \quad t_2 = 20 \quad t_3 = 20 \quad t_4 = 35] \rightarrow T = 90
\end{array}$$

- *Symmetry in units*. The units play a symmetrical role in the expression of the index. In other words, if $\big(\Pi(1),\dots,\Pi(J)\big)$ is a column vector representing a permutation of the units, then $\phi(\boldsymbol{c^g\Pi};\boldsymbol{t\Pi}) = \phi(\boldsymbol{c^g};\boldsymbol{t})$. This implies that the level of unevenness of group 1 in the scenario $(\boldsymbol{c^1};\boldsymbol{t}) \equiv (2,3,5,10;15,20,20,35)$ is the same as in the scenario $(\boldsymbol{c^1\Pi};\boldsymbol{t\Pi}) \equiv (3,2,5,10;20,15,20,35)$.

- *Size invariance*. If we replicate both vector $\boldsymbol{c^g}$ and vector $\boldsymbol{t}$ λ times, the unevenness of the group does not change. In other words, $\phi(\lambda\boldsymbol{c^g};\lambda\boldsymbol{t}) = \phi(\boldsymbol{c^g};\boldsymbol{t})$, where $\lambda > 0$. This property means that the level of unevenness of group 1 in the above example, where $(\boldsymbol{c^1};\boldsymbol{t}) \equiv (2,3,5,10;15,20,20,35)$, is the same as the level the group would

[3] This differs from other approaches followed in the literature to derive overall segregation indices in the case of two groups using group 2 as a reference against which group 1 is compared (Hutchens, 1991, 2001, 2004).

have in a world where the size of each unit and the size of the group in each unit doubled, i.e., in the scenario $(2\boldsymbol{c}^{\mathbf{1}}; 2\boldsymbol{t}) \equiv (4, 6, 10, 20; 30, 40, 40, 70)$.

- *Insensitivity to proportional subdivisions of units* (a property that does not exist in the measurement of inequality). The unevenness of group *g* should not change if a unit is split into several units of equal size with identical number of individuals of the group. In other words, if for simplicity we split the last unit, *J*, in $\mathrm{K} > 0$ units and in each of them we have $\frac{c_J^g}{K}$ individuals of the group and a total of $\frac{t_J}{K}$ people (while keeping the other units unchanged), then $\phi(\boldsymbol{c}^{\boldsymbol{g}\prime}, \boldsymbol{t}^{\prime}) = \phi(\boldsymbol{c}^{\boldsymbol{g}}, \boldsymbol{t})$, where $\boldsymbol{c}^{\boldsymbol{g}\prime}$ and $\boldsymbol{t}^{\prime}$ represent the resulting distributions after the subdivision. This implies that the unevenness of group 2 in the example, where $(\boldsymbol{c}^{\mathbf{2}}; \boldsymbol{t}) \equiv (3, 7, 10, 20; 15, 20, 20, 35)$, is the same as the unevenness it would have if $(\boldsymbol{c}^{\mathbf{2}\prime}; \boldsymbol{t}^{\prime}) \equiv (3, 7, 5, 5, 20; 15, 20, 10, 10, 35)$ because this change does not bring any heterogeneity to the group.
- *Sensitivity towards disequalizing movements between units (type I)* (which is a kind of Pigou-Dalton transfers principle). If some members of group *g* move from one unit to another unit of the same size in which the group has a higher presence, and this movement is accompanied by an opposing movement from other groups that keeps the size of the units unchanged, the unevenness of group *g* increases. In other words, if we denote by $\boldsymbol{c}^{\boldsymbol{g}\prime}$ the distribution of group *g* obtained when δ individuals of group *g* move from *i* to *h*, where $c_i^g < c_h^g$ and $t_i = t_h$, then $\phi(\boldsymbol{c}^{\boldsymbol{g}\prime}; \boldsymbol{t}) > \phi(\boldsymbol{c}^{\boldsymbol{g}}; \boldsymbol{t})$. This implies, for example, that the level of unevenness when $(\boldsymbol{c}^{\mathbf{1}}; \boldsymbol{t}) \equiv (2, 3, 5, 10; 15, 20, 20, 35)$ is lower than the level when $(\boldsymbol{c}^{\mathbf{1}\prime}; \boldsymbol{t}) \equiv (2, 2, 6, 10; 15, 20, 20, 35)$. [In this case, an individual of group 1 moved from unit 2 to unit 3 and an individual from either group 2 or 3 moved in the opposite direction, so that the size of these units does not change.]

Note that the translation of the Pigou-Dalton transfers principle to a segregation context may consider other formulations. Del Río and Alonso-Villar (2022) define disequalizing movements of *type II* and *type III*, which are more demanding than those of *type I* but allow studying how the index change in more realistic scenarios:

- *Sensitivity towards disequalizing movements between units (type II)* requires that the index increase when some individuals in group *g* move from *i* to *h*, where $\frac{c_i^g}{t_i} < \frac{c_h^g}{t_h}$ (regardless of whether units *i* and *h* have the same size), replacing that movement with a movement in the opposite direction by individuals from other groups. Namely, if we denote by $\boldsymbol{c^g}'$ the distribution of group *g* after that change, then $\phi(\boldsymbol{c^g}', \boldsymbol{t}) > \phi(\boldsymbol{c^g}, \boldsymbol{t})$.
- *Sensitivity towards disequalizing movements between units (type III)* requires that the index increase when some individuals in group *g* move from *i* to *h*, where $\frac{c_i^g}{t_i} < \frac{c_h^g}{t_h}$ (regardless of whether units *i* and *h* have the same size), without replacing that movement with movements from other groups. Namely, if we denote by $\boldsymbol{c^g}'$ and $\boldsymbol{t}'$, respectively, the distribution of group *g* and that of total population after that change, then $\phi(\boldsymbol{c^g}', \boldsymbol{t}') > \phi(\boldsymbol{c^g}, \boldsymbol{t})$.

Although these two properties entail seeing how the ϕ index behaves in scenarios that are more realistic than the previous one, it is not necessary to require compliance with them, since a combination of several properties mentioned earlier ensures that the index meets *type II* and *III* properties, as we will explain later.

Below, we present two useful properties—*scale invariance* and *aggregation*—which help characterize a family of indicators, as shown later:

- *Scale invariance*. It is like size invariance but without requiring that vectors $\boldsymbol{c^g}$ and $\boldsymbol{t}$ be multiplied by the same number. Thus, if we replicate vector $\boldsymbol{c^g}$ λ times and vector $\boldsymbol{t}$ β times, the unevenness of the group does not change. In other words, $\phi(\lambda \boldsymbol{c^g}; \beta \boldsymbol{t}) = \phi(\boldsymbol{c^g}; \boldsymbol{t})$, where $\lambda, \beta > 0$ and $\lambda c_j^g \leq \beta t_j \; \forall j$. This property implies that the unevenness of a group should only depend on how the proportion of the group in each unit ($\frac{c_j^g}{C^g}$) compares with the population share of that unit ($\frac{t_j}{T}$), this comparison expressed as a ratio, and nothing else.
- *Aggregation*. Let us cluster units in two mutually exclusive categories such that the number of individuals in category 1 and 2 is denoted by T^1 and T^2, respectively, and the number of group *g*'s individuals in these categories are C^{g1} and C^{g2}. Let us

assume for convenience that the first units are in category 1 and the next units are in category 2 so that $(\boldsymbol{c^g}; \boldsymbol{t}) = (\boldsymbol{c^{g1}}, \boldsymbol{c^{g2}}; \boldsymbol{t^1}, \boldsymbol{t^2})$. Index ϕ is aggregative if there exists a continuous aggregator function A such that $\phi(\boldsymbol{c^g}, \boldsymbol{t}) = A\left(\phi(\boldsymbol{c^{g1}}; \boldsymbol{t^1}), \frac{C^{g1}}{T^1}, T^1, \phi(\boldsymbol{c^{g2}}; \boldsymbol{t^2}), \frac{C^{g2}}{T^2}, T^2\right)$, where A is strictly increasing in the first and fourth argument. In other words, the unevenness of the group depends on its unevenness within each category, the population size of each category, and the proportion of the group within each category.

Characterizing Local Segregation Indices Based on These Criteria

Once the main properties have been established, we now discuss what type of indices we obtain by combining them. *Symmetry in units*, *scale invariance*, *insensitivity to proportional divisions of units*, and *sensitivity towards disequalizing movements between units (type I)* render local segregation indicators consistent with a dominance criterion based on local segregation curves, analogous to the Lorenz criterion used in the inequality literature (Alonso-Villar and Del Río, 2010).[4] In other words, the local segregation curve of a group is not at any point below another and is above it at some point (i.e., it is closer to the 45º line) if and only if the former group is more evenly distributed than the latter according to any local segregation index satisfying these four properties. Like Lorenz curves, local segregation curves are a powerful tool given that they allow us to conclude whether the unevenness of a group is greater in a scenario than in another for different formulations of the index (provided the curves do not intersect).

It follows from the above that if we want to use indicators of unevenness that are consistent with the dominance criterion given by the local segregation curves, our indicators should satisfy, at least, the above four properties.

Additionally, any local segregation index consistent with this dominance criterion meets *sensitivity to disequalizing movements type III*, whereas *insensitivity to proportional subdivisions of units* and *sensitivity to disequalizing movements type I* implies *sensitivity to disequalizing movements type II* (Del Río and Alonso-Villar, 2022). In other words,

[4] For the definition of this curve, see Appendix.

symmetry in units, *scale invariance*, *insensitivity to proportional divisions of units,* and *sensitivity towards disequalizing movements between units type I* are enough to ensure *sensitivity to disequalizing movements type II* and *type III.*

When adding the *aggregation* property to the above four properties, a family of local segregation indices related to the generalized entropy family is fully characterized (Alonso-Villar and Del Río, 2010):

$$\phi_\alpha(\boldsymbol{c^g};\boldsymbol{t}) = \begin{cases} \frac{1}{\alpha(\alpha-1)}\sum_j \frac{t_j}{T}\left[\left(\frac{\frac{c_j^g}{C^g}}{\frac{t_j}{T}}\right)^\alpha - 1\right] & \text{if } \alpha \neq 1 \\ \sum_j \frac{c_j^g}{C^g} ln\left(\frac{\frac{c_j^g}{C^g}}{\frac{t_j}{T}}\right) & \text{if } \alpha = 1 \\ \sum_j \frac{t_j}{T} ln\left(\frac{\frac{t_j}{T}}{\frac{c_j^g}{C^g}}\right) & \text{if } \alpha = 0 \end{cases} . \quad (1)$$

This family of unevenness indicators depends on an inequality aversion parameter, α. The lower its value, the greater the sensitivity of the index to what happens in the units where the group has a low presence (i.e., in units with low $\frac{c_j^g}{t_j}$ ratios). If $\alpha \leq 0$, the index can only be used if the group has presence in all units (i.e., if $c_j^g > 0 \; \forall j$).

When we use these local segregation indicators, the unevenness of a group can be broken down into unevenness within and between clusters (Del Río and Alonso-Villar, 2010). Thus, if we group the units into *K* clusters and denote by C^{gk} the number of individuals of group *g* in cluster *k* (*k*=1,…,*K*), by $\boldsymbol{c^{gk}}$ the distribution of the group across units of that cluster, by T^k the population of *k* and by $\boldsymbol{t^k}$ the distribution of the population across units of *k*, we get:

$$\phi_\alpha(\boldsymbol{c^g};\boldsymbol{t}) = \underbrace{\sum_k \left(\frac{C^{gk}}{C^g}\right)^\alpha \left(\frac{T^k}{T}\right)^{1-\alpha} \phi_\alpha(\boldsymbol{c^{gk}};\boldsymbol{t^k})}_{within\ term} + \underbrace{\phi_\alpha(C^{g1},\dots,C^{gK};T^1,\dots,T^K)}_{between\ term}. \quad (2)$$

The first term embodies the within component ($\phi_\alpha(\boldsymbol{c^{gk}};\boldsymbol{t^k})$ being the unevenness of group *g* in cluster *k*) and the second term represents the between component (i.e., the unevenness that the group would have if, within each cluster, the group were evenly distributed). Later on, we will use this within-between decomposition of local segregation indexes to decompose the corresponding overall segregation indices.

From all of the above, it follows that any indicator measuring the unevenness of a group should satisfy, at least, *symmetry in units*, *size invariance*, *insensitivity to proportional subdivisions of units,* and *sensitivity towards disequalizing movements between units type I*. If, in addition, the index satisfies *scale invariance*, we know that it also meets *sensitivity towards disequalizing movements between units type II* and *III* and is consistent with the dominance criterion given by the local segregation curves. If we add *aggregation*, the index takes the form of ϕ_α (or a monotonically increasing transformation of it) and can be decomposed into within- and between-cluster components.

2.3 Building Overall Segregation by Aggregating the Unevenness of Groups: A Proposal

After discussing how the literature measures the unevenness of a group, we now put forward a method for constructing overall segregation measures. We propose measuring multigroup overall segregation, Φ, as the weighted average of the unevenness of each group (with weights equal to the groups' population shares):

$$\Phi(\mathrm{E}) = \sum_{g=1}^{G} \frac{C^g}{T} \phi(\boldsymbol{c}^{\boldsymbol{g}}; \boldsymbol{t}), \tag{3}$$

using (unstandardized) indicators $\phi: D \to \mathbb{R}$ that satisfy, at least, *symmetry in units*, *size invariance*, *insensitivity to proportional subdivisions of units*, and *sensitivity towards disequalizing movements between units (type I)*.[5]

This way of deriving overall segregation indices—by aggregating the unevenness of groups using indices that satisfy desirable properties—is a very useful procedure since it ensures that those indices meet several criteria established in the literature for measuring overall segregation, whether in a binary context (James and Taueber, 1985; Hutchens, 1991, 2001) or one involving multigroup groups (Reardon and Firebaugh, 2002). Thus, index Φ satisfies:[6]

[5] If we want to work with standardized measures of multigroup segregation, we will divide Φ by its maximum value.

[6] Φ is a continuous function that is equal to zero if and only if the groups are evenly distributed among the units (otherwise it is positive), given that ϕ satisfies these properties.

- *Symmetry in units* (Hutchens, 1991): If we swap the units, the index remains the same. It is straightforward to prove that Φ fulfills this requirement given that the indicators we use to quantify the unevenness of each group satisfy that $\phi(\boldsymbol{c^g\Pi};\boldsymbol{t\Pi})=\phi(\boldsymbol{c^g};\boldsymbol{t})$ for any permutation of units given by Π.
- *Symmetry in social groups* (Hutchens, 2001): The index does not change if we permute the social groups. Trivially, Φ fulfills this criterion given that it is the sum of what happens to each social group.
- *Size invariance* (James and Taeuber, 1985; Reardon et al., 2002): If the number of individuals of each social group in each unit is multiplied by a constant, segregation remains the same. Φ also meets this property given that the weighting scheme does not change and $\phi(\lambda\boldsymbol{c^g};\lambda\boldsymbol{t})=\phi(\boldsymbol{c^g};\boldsymbol{t})$ for any group *g* and $\lambda>0$.
- *Organizational equivalence* (James and Taeuber, 1985; Reardon et al., 2002): When two units with the same size and group composition are combined into a single unit, overall segregation does not change. Φ trivially satisfies this property given that ϕ satisfies *insensitivity to proportional subdivisions of units* when applying it to each $(\boldsymbol{c^g};\boldsymbol{t})$.
- *Principle of exchanges* (Reardon et al., 2002): If an individual of group *g* in unit *i* is exchanged with another of group *g'* in unit *h*, where $\frac{c_i^g}{t_i}<\frac{c_i^{g'}}{t_i}$ and $\frac{c_h^g}{t_h}>\frac{c_h^{g'}}{t_h}$, multigroup segregation increases. It is easy to prove that Φ satisfies this property because this exchange involves two *disequalizing movements type II* for groups *g* and *g′*. Note that after that change, the unevenness of groups *g* and *g′* given by ϕ increases (as mentioned above, *sensitivity to disequalizing movements (type I)* and *insensitivity to proportional subdivisions of units* implies *sensitivity to disequalizing movements (type II)*), while there are no changes for the other groups.[7]

This way of aggregating the unevenness of the groups also implies giving the Φ indices another characteristic. Ceteris paribus, the greater the demographic weight of a group, the

[7] Note that, as we will show later, the fact that ϕ satisfies the property of *sensitivity towards disequalizing movements between units (type III)* does not guarantee that Φ fulfills the *principle of transfers* (according to which the index always decreases when an individual from a group moves to a unit where the group has a smaller representation).

greater the impact of its unevenness on the value of the index. In other words, the situation of large groups determines overall segregation to a greater extent, and consequently, the situation of very small groups is overshadowed. All of this seems reasonable if we want to quantify total segregation in a society, although it makes it difficult to understand the distinctive situation of the groups.[8] In any case, note that the decomposition shown in expression (3) allows us to separate the contribution of each group to the overall segregation into two components: the unevenness of each group and its population weight.

Our method bears similarities to the first of the methods proposed by Reardon and Firebaugh (2002) for deriving multigroup indices, since $\Phi(\mathrm{E})$ also aggregates the disproportionality in group proportions of the different social groups. However, our proposal allows us to go a step further since it takes as its starting point functions that guarantee an adequate measurement of the unevenness of the groups, $\boldsymbol{\phi(c^g;t)}$, giving that metric its own significance and placing it at the heart of the process. It is not simply a matter of aggregating the levels of disproportionality of the groups (associated with their distribution among the units), but rather of requiring that the functions we use for that purpose satisfy desirable properties that allow us to adequately quantify the unevenness of each group. As we have just seen, this ensures that the overall indices obtained are good measures of segregation.

3. Proposing a Family of Multigroup Segregation Indices $\Phi_{\boldsymbol{\alpha}}$

Using our procedure, we now derive a parameterized family of multigroup segregation indices, $\Phi_\alpha(\mathrm{E}) = \sum_{g=1}^{G} \frac{C^g}{T} \phi_\alpha(\boldsymbol{c^g}; \boldsymbol{t})$, based on the unevenness indicators ϕ_α defined earlier. Namely,

[8] To avoid this, we could also consider other ways of aggregating the groups' unevenness. For example, we could use the arithmetic mean, but this would be difficult to justify since it implicitly assumes that all groups have the same importance, regardless of their relative size. Furthermore, adding up unevenness without using demographic weights would prevent us from deriving some of the multigroup segregation indices already proposed in the literature.

$$\Phi_\alpha(E) = \begin{cases} \frac{1}{\alpha(\alpha-1)} \sum_g \frac{C^g}{T} \sum_j \frac{t_j}{T} \left[\left(\frac{\frac{c_j^g}{C^g}}{\frac{t_j}{T}} \right)^\alpha - 1 \right] & \text{if } \alpha \neq 0, 1 \\ \sum_g \frac{C^g}{T} \sum_j \frac{c_j^g}{C^g} ln \left(\frac{\frac{c_j^g}{C^g}}{\frac{t_j}{T}} \right) & \text{if } \alpha = 1 \\ \sum_g \frac{C^g}{T} \sum_j \frac{t_j}{T} ln \left(\frac{\frac{t_j}{T}}{\frac{c_j^g}{C^g}} \right) & \text{if } \alpha = 0 \end{cases} . \tag{4}$$

As already mentioned, indicators ϕ_α satisfy *symmetry in units*, *scale invariance*, *insensitivity to proportional subdivisions of units*, and *sensitivity towards disequalizing movements between units (type I).* Since *scale invariance* implies *size invariance*, ϕ_α meets the requirements that ensure that Φ_α satisfies *symmetry in units*, *symmetry in social groups*, *size invariance*, *organizational equivalence*, and the *principle of exchanges*. Note that for $\alpha \leq 0$, these indices can only be used if the social groups are present in all units (i.e., if $c_j^g > 0 \; \forall g, j$). For this reason, in what follows we will consider only the case where $\alpha > 0$.

This family of indicators generalizes the mutual information index M since $M = \Phi_1$. Therefore, its normalized version, the H index, is also related to this family.[9] In addition, the unnormalized version of the squared coefficient of variation, C_u, is part of this family since $C_u = 2\,\Phi_2$, and thus its normalized version, the C index proposed by Reardon and Firebaugh (2002), is also related to this family. To the best of our knowledge, the other members of this family are new multigroup segregation indices. In particular, Φ_α includes multigroup segregation indices with inequality aversion values lying between those of Theil 0 and Theil 1, values considered in inequality studies. Note that when $\alpha = 0.5$,

$$\Phi_{0.5} = 4 \sum_g \frac{C^g}{T} \left[1 - \sum_j \sqrt{\left(\frac{c_j^g}{C^g}\right)\left(\frac{t_j}{T}\right)} \right]. \tag{5}$$

[9] The mutual information index is the only member of this family that satisfies the *principle of transfers* mentioned above (see Appendix). Del Río and Alonso-Villar (2022) discuss why it may not be necessary to require this property of overall multigroup segregation indices if the corresponding local segregation indices satisfy *sensitivity towards disequalizing movements between units (type III)*.

Thus, $\Phi_{0.5}$ can be regarded as a multigroup segregation index that shares with the binary index proposed by Hutchens (2001, 2004), the square root index, an inequality aversion of 0.5.[10]

It follows from all of the above that Φ_α, with $\alpha > 0$, is a family of multigroup segregation indices that satisfy basic properties, remains valid even when some groups have zeros in some units, and is compatible with a wide range of views on inequality (encapsulated in α). Having a parameterized family of segregation indices is useful. On the one hand, it allows us to identify empirical situations in which the results achieved present a high degree of robustness with respect to the choice of index. On the other hand, if there are discrepancies depending on the index used, these could be interpreted based on the value judgments represented by the parameter associated with each of them.

Is There a Within-Between Decomposition for Multigroup Indices Φ_α?

Next, we show that, of our family of multigroup segregation indices, Φ_α, only the *M* index can be broken down into a between-cluster term and a within-cluster term, although for the remaining indices, a between component can also be disentangled from the unevenness that the groups experience within clusters.

To prove this, first note that, since the overall segregation Φ_α is expressed in terms of the unevenness of each group, and given that this unevenness can be decomposed into within- and between-cluster terms (see expression (2)), we can write:

$$\Phi_\alpha(E) = \sum_g \frac{C^g}{T}\left[\sum_k \left(\frac{C^{gk}}{C^g}\right)^{\alpha}\left(\frac{T^k}{T}\right)^{1-\alpha} \phi_\alpha\left(\boldsymbol{c^{gk}}; \boldsymbol{t^k}\right) + \phi_\alpha\left(C^{g1}, \dots, C^{gK}; T^1, \dots, T^K\right)\right], \qquad (6)$$

[10] In the case of two groups, index $\Phi_{0.5}$ differs from the square root index because the "income" distributions underlying these two segregation indices are not the same. $\Phi_{0.5}$ involves calculating the inequality of distribution $(\underbrace{\frac{c_1^g}{t_1}, \dots, \frac{c_1^g}{t_1}}_{t_1 \text{ individuals}}, \dots, \underbrace{\frac{c_J^g}{t_J}, \dots, \frac{c_J^g}{t_J}}_{t_J \text{ individuals}})$ for $g = 1$ and 2 (where distributions $\boldsymbol{c^1}$ and $\boldsymbol{c^2}$ are compared to a common distribution $\boldsymbol{t}$, which is strictly positive) and then averaging those two values. The square root, on the other hand, measures the inequality of distribution $(\underbrace{\frac{c_1^1}{c_1^2}, \dots, \frac{c_1^1}{c_1^2}}_{c_1^2 \text{ individuals}}, \dots, \underbrace{\frac{c_J^1}{c_J^2}, \dots, \frac{c_J^1}{c_J^2}}_{c_J^2 \text{ individuals}})$, if $c_j^2 > 0\ \forall j$, which results from comparing one group with the other. If instead $c_j^2 = 0$ in a unit, c_j^2 would be replaced by an infinitesimally small number and the square root index would be approximated by the inequality of the corresponding distribution.

where, as already mentioned, C^{gk} is the number of individuals of group g in cluster k and $\boldsymbol{c^{gk}}$ is the distribution of the group across units in that cluster. Next, after some simple calculations, we get:

$$\Phi_\alpha(E) = \sum_k \frac{T^k}{T}\left[\sum_g \left(\frac{\frac{C^{gk}}{T^k}}{\frac{C^g}{T}}\right)^{\alpha-1}\left(\frac{C^{gk}}{T^k}\right)\phi_\alpha(\boldsymbol{c^{gk}};\boldsymbol{t^k})\right] + \sum_g \frac{C^g}{T}\phi_\alpha(C^{g1},\dots,C^{gK};T^1,\dots,T^K). \qquad (7)$$

Therefore, overall segregation can be broken down into two components.[11] The first component comprises the unevenness of each group in each cluster, $\phi_\alpha(\mathbf{c^{gk}};\boldsymbol{t^k})$, weighted by the group's proportion in the cluster, $\frac{C^{gk}}{T^k}$, and by how this compares to the group's proportion in the total, $\frac{C^g}{T}$. The second component, which we call the between-cluster component, reflects the segregation society would have if, within each cluster, the groups were evenly distributed among its units. This decomposition allows us to disentangle the segregation that arises from differences in the situation of the groups among clusters from the unevenness experienced by the groups within those clusters.

Note that if $\frac{\frac{C^{gk}}{T^k}}{\frac{C^g}{T}}$ did not depend on g (i.e., if each group had the same weight in all clusters), this ratio would be equal to 1 and, therefore, the bracketed term in expression (7) would reflect the overall segregation in the cluster (i.e., the first term in expression (7) would properly be a within-cluster component for any Φ_α). However, this is a pattern that in general does not occur. In any case, for $\alpha = 1$, the above decomposition is properly a within-between decomposition, regardless of how the groups are distributed among the clusters, given that:

$$M = \Phi_1(E) = \sum_k \frac{T^k}{T}\Phi_1(E^k) + \Phi_1(E_B), \qquad (8)$$

where the first term is the weighted sum of the overall segregation within each cluster (whose corresponding matrix is denoted by E^k) and the second term is the overall segregation we would obtain if we treated each cluster as a unit (the corresponding matrix is denoted by E_B).

[11] This decomposition holds not only for $\alpha > 0$, but also for $\alpha \leq 0$ (provided that $c_j^g > 0\ \forall j$ and $\forall g$).

This is a breakdown of the mutual information index already known in the literature (Frankel and Volij, 2011).[12]

4. From Evenness to Representativeness

In Section 3, following an unevenness view of segregation, we put forward a family of multigroup segregation indices, Φ_α, which generalizes the mutual information index and the unnormalized version of the squared coefficient of variation. In this section, we show that these multigroup indices can be also expressed as weighted averages of the unrepresentativeness of each unit.

First, we relate the measurement of unrepresentativeness to the measurement of inequality, and then we discuss what properties the indicators of unrepresentativeness should satisfy. Next, we express the overall segregation indices Φ_α in terms of unrepresentativeness indicators, which is useful for empirical analysis as it allows us to determine the contribution of each unit or cluster of units (identified according to criteria the researcher deems relevant) to the overall segregation.

4.1 Properties for Measuring the Unrepresentativeness of a Unit

Frankel and Volij (2011) define a unit as representative if the proportion of each group in that unit is equal to its proportion in the total (i.e., if the composition of that unit matches the composition of the total) and they consider society to be fully integrated if every unit is representative.

Therefore, to measure the unrepresentativeness of a unit *j,* we must measure the discrepancy between the composition of the unit and that of the total population, i.e., we must compare $\left(\frac{c_j^1}{t_j},\dots,\frac{c_j^G}{t_j}\right)$ with $\left(\frac{C^1}{T},\dots,\frac{C^G}{T}\right)$. Alonso-Villar and Del Río (2010) established the properties that any local segregation index measuring the unevenness of a group must satisfied and pointed out that these properties could be reinterpreted in terms of unrepresentativeness, given the duality between rows and columns in the E matrix, although they did not elaborate further on this issue.

[12] In the context of school segregation, these authors refer to this property as *strong school decomposability*.

In what follows, we state these properties. As in the measurement of unevenness, some of them result from adapting others invoked in the inequality literature, since the unrepresentativeness of a unit j can be measured as the inequality of "income" distribution $(\underbrace{\frac{c_j^1}{C^1}, \dots, \frac{c_j^1}{C^1}}_{C^1 \text{ individuals}}, \dots, \underbrace{\frac{c_j^G}{C^G}, \dots, \frac{c_j^G}{C^G}}_{C^G \text{ individuals}})$ in a world of T individuals and a total income of t_j.

Let $\psi: \widehat{D} \to \mathbb{R}$ be an indicator of the unrepresentativeness of a unit j, where the domain is given by $\widehat{D} = \bigcup_{G>1}\{(\boldsymbol{c_j}; \boldsymbol{C}) \in \mathbb{R}^G{}_+ \times \mathbb{R}^G{}_{++}: c_j^g \le C^g \ \forall g\}$, $\boldsymbol{c_j} \equiv (c_j^1, c_j^2, \dots, c_j^G)$ being the vector that represents the number of individuals of each group in unit j and $\boldsymbol{C} \equiv (C^1, C^2, \dots, C^G)$ being the vector that represents the size of each group in the total. We can think of ψ as a local segregation index from a representativeness perspective. We require that this index meet the following criteria:

- *Size invariance*. If we replicate both vector $\boldsymbol{c_j}$ and vector $\boldsymbol{C}$ λ times, the unrepresentativeness of unit j does not change. In other words, $\psi(\lambda \boldsymbol{c_j}; \lambda \boldsymbol{C}) = \psi(\boldsymbol{c_j}; \boldsymbol{C})$, where $\lambda > 0$.
- *Symmetry in groups*. Groups play a symmetrical role in the expression of the index. In other words, if $(\Pi(1), \dots, \Pi(G))$ is a column vector representing a permutation of the groups, then $\psi(\boldsymbol{c_j \Pi}; \boldsymbol{C \Pi}) = \psi(\boldsymbol{c_j}; \boldsymbol{C})$.
- *Sensitivity towards disequalizing movements between groups (type I)*. If g and g' are two groups of equal size (i.e., $C^g = C^{g'}$) and the latter has a higher presence in unit j than the former (i.e., $c_j^g < c_j^{g'}$), the unrepresentativeness of unit j should increase when δ individuals of group g are replaced by δ individuals of group g' (without changing the size of each group in the total population). In other words, $\psi(\boldsymbol{c_j}'; \boldsymbol{C}) > \psi(\boldsymbol{c_j}; \boldsymbol{C})$, where vector $\boldsymbol{c_j}'$ represents the composition of unit j after the above movement.
- *Insensitivity to proportional subdivisions of groups*. The unrepresentativeness of a unit should not change if a group is split into several subgroups of equal size, both in the total population and in the unit. Namely, if for simplicity we split the last group, G, in $K > 0$ subgroups of equal size $\frac{C^G}{K}$ and the presence of each of them in unit j is

given by $\frac{c_j^G}{K}$ (while keeping the other groups unchanged), then $\psi(\boldsymbol{c_j}'; \boldsymbol{C}') = \psi(\boldsymbol{c_j}; \boldsymbol{C})$, where $\boldsymbol{c_j}'$ and $\boldsymbol{C}'$ represent the resulting distributions after the subdivision.

We may also consider including an additional property, *scale invariance*:

- *Scale invariance*. The index should not change when we multiply vector $\boldsymbol{c_j}$ by $\lambda_1 > 0$ and vector $\boldsymbol{C}$ by $\lambda_2 > 0$. Namely, $\psi(\lambda_1 \boldsymbol{c_j}; \lambda_2 \boldsymbol{C}) = \psi(\boldsymbol{c_j}; \boldsymbol{C})$ where $\lambda_1, \lambda_2 > 0$ and $(\lambda_1 \boldsymbol{c_j}; \lambda_2 \boldsymbol{C}) \in \widehat{D}$. This implies that what should matter to the index is how each group is represented in the unit ($\frac{c_j^g}{t_j}$) in relation to its representation in the total ($\frac{C^g}{T}$), expressed as a ratio, and nothing else. Note that *scale invariance* implies *size invariance*.

As we state in the next theorem, *scale invariance* (together with the last three previous properties) provides consistency with a dominance criterion based on curves of unrepresentativeness, which makes it useful. [To make the text flow more smoothly, we define the unrepresentativeness curve, which follows a similar logic to the unevenness curve, in the Appendix.]

Theorem. A curve of unrepresentativeness dominates another (i.e., the former is above the latter at some points and is not below it at any point) if and only if the level of unrepresentativeness is higher in the dominated distribution for any index $\psi: \widehat{D} \to \mathbb{R}$ satisfying *scale invariance*, *symmetry in groups*, *sensitivity towards disequalizing movements between groups (type I)*, and *insensitivity to proportional subdivisions of groups*.

Proof. This follows from the duality between the measurement of evenness and that of representativeness and from Theorem 1 in Alonso-Villar and Del Río (2010), where the relationship between the unevenness indicators that satisfy the corresponding properties and the unevenness curves is established (following steps analogous to those used in the inequality literature with Lorenz curves).

To quantify the unrepresentativeness of a unit, Alonso-Villar and Del Río (2010) proposed this family of indices:

$$\psi_\alpha(\boldsymbol{c_j};\boldsymbol{C}) = \begin{cases} \frac{1}{\alpha(\alpha-1)}\sum_g \frac{C^g}{T}\left[\left(\frac{\frac{c_j^g}{t_j}}{\frac{C^g}{T}}\right)^\alpha - 1\right] & \text{if } \alpha \neq 0,1 \\ \sum_g \frac{c_j^g}{t_j} ln\left(\frac{\frac{c_j^g}{t_j}}{\frac{C^g}{T}}\right) & \text{if } \alpha = 1 \\ \sum_g \frac{C^g}{T} ln\left(\frac{\frac{C^g}{T}}{\frac{c_j^g}{t_j}}\right) & \text{if } \alpha = 0 \end{cases}. \tag{9}$$

It is easy to see that ψ_α satisfies *size invariance*, *scale invariance*, *symmetry in groups*, and *insensitivity to proportional subdivisions of groups*. After simple calculations, we can also prove that it satisfies *sensitivity towards disequalizing movements between groups type I* (see Appendix).

4.2. Expressing Multigroup Indices Φ_α in Terms of Unrepresentativeness

As acknowledged in the literature (see, for example, Elbers, 2023), the mutual information index can be expressed as the weighted average of "local scores of units" given that

$$M = \sum_g \frac{C^g}{T}\sum_j \frac{c_j^g}{C^g} ln\left(\frac{\frac{c_j^g}{C^g}}{\frac{t_j}{T}}\right) = \sum_j \frac{t_j}{T}\sum_g \frac{c_j^g}{t_j} ln\left(\frac{\frac{c_j^g}{t_j}}{\frac{C^g}{T}}\right). \tag{10}$$

What we add to this literature is that *M* can be expressed both as the average unevenness of the groups and as the average unrepresentativeness of the units:

$$M = \Phi_1(E) = \sum_g \frac{C^g}{T}\phi_1(\boldsymbol{c^g};\boldsymbol{t}) = \sum_t \frac{t_j}{T}\psi_1(\boldsymbol{c_j};\boldsymbol{C}), \tag{11}$$

using the indicators $\phi_1(\boldsymbol{c^g};\boldsymbol{t})$ and $\psi_1(\boldsymbol{c_j};\boldsymbol{C})$, respectively, which fulfill desirable properties. Moreover, this occurs not only for Φ_1 but for the other members of the family Φ_α. For $\alpha \neq 0, 1$, the index can be expressed as:

$$\Phi_\alpha(E) = \sum_g \frac{C^g}{T} \underbrace{\frac{1}{\alpha(\alpha-1)} \sum_j \frac{t_j}{T} \left[\left(\frac{\frac{c_j^g}{C^g}}{\frac{t_j}{T}} \right)^\alpha - 1 \right]}_{\phi_\alpha(\boldsymbol{c^g};\boldsymbol{t})} = \sum_j \frac{t_j}{T} \underbrace{\frac{1}{\alpha(\alpha-1)} \sum_g \frac{C^g}{T} \left[\left(\frac{\frac{c_j^g}{t_j}}{\frac{C^g}{T}} \right)^\alpha - 1 \right]}_{\psi_\alpha(\boldsymbol{c_j};\boldsymbol{C})}, \qquad (12)$$

and for $\alpha = 0$ we can do the same. In other words, all members of the family of multigroup segregation indices Φ_α can be interpreted in terms of unrepresentativeness (in addition to unevenness), since they can be expressed as the weighted average of the unrepresentativeness of each unit using indicators that comply with basic criteria:

$$\Phi_\alpha(E) = \sum_g \frac{C^g}{T} \phi_\alpha(\boldsymbol{c^g}; \boldsymbol{t}) = \sum_j \frac{t_j}{T} \psi_\alpha(\boldsymbol{c_j};\ \boldsymbol{C}). \qquad (13)$$

Expressing the overall segregation in terms of the unrepresentativeness of each unit can be useful in empirical analysis. Let us consider, for example, that we are interested in school segregation in a district and that schools are clustered according to some criterion considered relevant. We can express the overall school segregation in that district as the weighted average of the segregation that exists within each cluster (denoted by *k*), taking the district as the distribution of reference. Namely,

$$\Phi_\alpha(E) = \sum_k \underbrace{\sum_{j\in k} \frac{t_j}{T} \psi_\alpha(\boldsymbol{c_j}; \boldsymbol{C})}_{\substack{Contribution \\ of\ cluster\ k}} = \sum_k \frac{T^k}{T} \sum_{j \in k} \frac{t_j}{T^k} \psi_\alpha(\boldsymbol{c_j}; \boldsymbol{C}), \qquad (14)$$

where $T^k = \sum_{j\in k} t_j$ and $T = \sum_k T^k$. In other words, if we cluster the units according to a criterion that we consider relevant to the analysis, we can determine the contribution of each cluster to overall segregation.

This decomposition allows researchers to move beyond the popular within-between decomposition. Using the within-between decomposition, we can determine, for example, whether grouping units by region is relevant to explain segregation in a country (i.e., whether there are significant differences in the distribution of the groups among regions, which is reflected on the between component) or, on the contrary, segregation in that country mainly arises from what happens within regions (within component). However, this decomposition does not allow us to determine how each region contributes to explain the country's segregation, a question that could be certainly answered using the decomposition in expression (14).[13] Furthermore, identifying which units exhibit the highest and lowest levels of unrepresentativeness allows us to explore the characteristics of those units, which may be useful for defining the clusters and, consequently, for explaining segregation.[14]

5. Conclusions

This paper develops a bottom-up method for deriving multigroup segregation measures from an evenness perspective. First, the unevenness of each group is measured using local segregation indices that satisfy basic properties, and then these values are aggregated to determine overall segregation. This procedure allows us to bring a new angle to the conceptualization of multigroup segregation.

By following this method, we derive a family of multigroup segregation indices, Φ_α (with $\alpha > 0$), which meet desirable properties and remain valid even when some groups have zeros in some units, making them useful in empirical analyses. This family allows us to expand the range of available measures simply by changing the value of a parameter, making it easier to incorporate different value judgements into the measurement.

Our segregation measures can also be expressed as weighted averages of the unrepresentativeness of the units (with weights equal to their respective population shares),

---

[13] Note that the within-between decomposition allows us to determine how each cluster contributes to explain the within component, but not its contribution to the between component.

[14] The I_p index proposed by Silber (1992), which generalizes the Karmel and MacLachlan (1988) index to the multigroup case, also allows for this type of analysis, since it can be expressed as $I_p = \sum_g \frac{C^g}{T}\left(\frac{1}{2}\sum_j \left|\frac{c_j^g}{C^g} - \frac{t_j}{T}\right|\right) = \sum_j \frac{t_j}{T}\left(\frac{1}{2}\sum_g \left|\frac{c_j^g}{t_j} - \frac{C^g}{T}\right|\right)$. However, it should be noted that the I_p index cannot be expressed in terms of ϕ- and ψ-type functions since the terms in parentheses do not satisfy the properties of *sensitivity towards disequalizing movements between units (type I)* and *sensitivity towards disequalizing movements between groups (type I)*, respectively. Expression $\frac{1}{2}\sum_j \left|\frac{c_j^g}{C^g} - \frac{t_j}{T}\right|$ is actually the index proposed by Moir and Selby Smith (1979). This index can be derived applying the Schutz index, which does not satisfy the Pigou-Dalton transfers principle, to the “income” distribution described at the beginning of Section 2.2. Given the parallelism between this principle and the property of *sensitivity towards disequalizing movements between units (type I)*, the Moir and Selby-Smith index does not comply with this property. An analogous reasoning could be applied to $\frac{1}{2}\sum_g \left|\frac{c_j^g}{t_j} - \frac{C^g}{T}\right|$.

where this unrepresentativeness is quantified using indicators satisfying basic properties that we develop here. This characteristic involves not only the mutual information index, as already acknowledged in the literature (Frankel and Volij, 2011), but also the other indices Φ_α that we build based on other inequality views of the generalized entropy family.

It is important to bear in mind that the properties required for measuring the unevenness of a group or the unrepresentativeness of a unit are not the same to those required of overall segregation measures, although they have similarities since the three phenomena involve measuring inequalities. The fact that we think of overall segregation in terms of "local" indicators that measure either each group's unevenness or each unit's unrepresentativeness meeting some criteria brings a new angle to previous research that expressed indices *M*, *H*, and *C* in terms of local scores with no theoretical or conceptual interpretation (Reardon and Firebaugh, 2002; Elbers, 2023).

Our overall segregation indices are useful tools for empirical analyses. First, we can easily determine the contribution of each group to overall segregation, distinguishing between the group's relative size and its level of unevenness. Second, if we cluster the units according to some criterion (for example, by type of funding in the case of school segregation), we can determine the contribution of each cluster to overall segregation, something particularly pertinent if there are more than two clusters. This allows researchers moving beyond the popular within-between decomposition, which is useful to determine the importance of the criterion to explain overall segregation (reflected on the between component), but not to determine the role played by each cluster or unit.

References

Alonso-Villar, O. and Del Río, C. (2010): "Local Versus Overall Segregation Measures," *Mathematical Social Sciences* 60(1), 30−38.

Alonso-Villar, O. and Del Río, C. (2025): "From the Extent of Segregation to Its Consequences in Terms of Wellbeing," *Hacienda Pública Española / Review of Public Economics* 254, 5-43.

Boisso, D., Hayes, K., Hirschberg, J., and Silber, J. (1994): "Occupational Segregation in the Multidimensional Case. Decomposition and Tests of Significance," *Journal of Econometrics* 61(1), 161−171.

Del Río, C. and Alonso-Villar, O. (2010): "Gender Segregation in the Spanish Labor Market: An Alternative Approach." *Social Indicators Research* 98, 337–362.

Del Río, C. and Alonso-Villar, O. (2022): "On Measuring Segregation in a Multigroup Context: Standardized Versus Unstandardized Indices," *Social Indicators Research* 163, 633−659.

Elbers, B. (2023): "A Method for Studying Differences in Segregation Across Time and Space," *Sociological Methods and Research* 52(1), 5−42.

Frankel, D. and Volij, O. (2011): "Measuring School Segregation," *Journal of Economic Theory* 146(1), 1−38.

Hutchens, R. (1991): "Segregation Curves, Lorenz Curves, and Inequality in the Distribution of People across Occupations," *Mathematical Social Sciences* 21, 31−51.

Hutchens, R. (2001): "Numerical Measures of Segregation: Desirable Properties and their Implications," *Mathematical Social Sciences* 42, 13−29.

Hutchens, R. (2004): "One Measure of Segregation," *International Economic Review* 45 (2), 555−578.

James, D. and Taeuber, K. (1985): "Measures of Segregation," *Sociological Methodology* 15, 1−32.

Karmel, T. and MacLachlan, M. (1988): "Occupational Sex Segregation—Increasing or Decreasing?," *The Economic Record* 64(3), 187−195.

Moir, H. and Selby Smith, J. (1979): "Industrial Segregation in the Australian Labour Market," *Journal of Industrial Relations* 21, 281−291.

Reardon, S. and Firebaugh, G. (2002): "Measures of Multigroup Segregation," *Sociological Methodology* 32, 33−76.

Silber, J. (1992): "Occupational Segregation Indices in the Multidimensional Case: A Note," *The Economic Record* 68, 276−277.

Theil, H. and Finizza, A. (1971): "A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts," *Journal of Mathematical Sociology* 1(2), 187−194.

Appendix

The Local Segregation Curve (Unevenness)

Alonso-Villar and Del Río (2010) define the local segregation curve for group g, denoted by S^g (Fig. 1). To build it, first, we must rank the units from low to high $\frac{c_j^g}{t_j}$ ratios. On the horizontal axis, we accumulate the proportion of individuals in these units ($\sum_{i\leq j}\frac{t_i}{T}$) and, on the vertical axis, we accumulate the proportion of individuals of group g in these same units ($\sum_{i\leq j}\frac{c_i^g}{C^g}$). Namely, if we denote by $\tau_j \equiv \sum_{i\leq j}\frac{t_i}{T}$ the cumulative proportion individuals in the first j units, ranked according to the above ratio, the segregation curve at that point is $S^g(\tau_j) = \frac{\sum_{i\leq j} c_i^g}{C^g}$. If group g is evenly distributed across units, S^g is equal to the 45º line. The further the curve is from this line, the greater is the unevenness of the group.

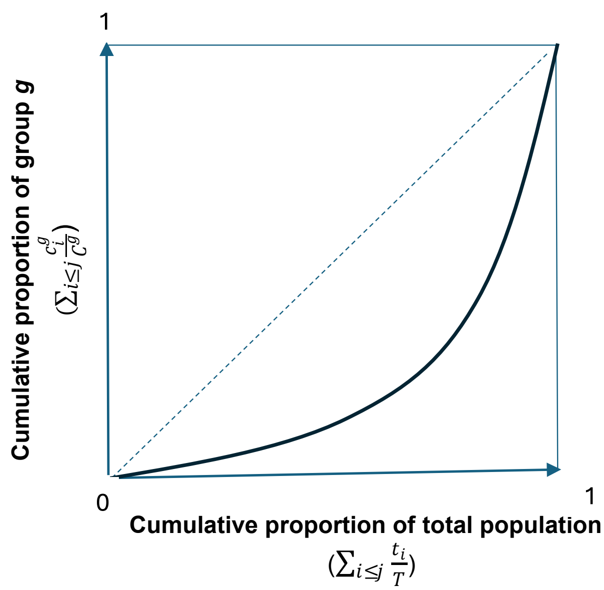

Figure 1. Local segregation curve of group g

A curve is said to dominate another if there is no point at which it lies below it and there is at least one point at which it lies above it.

The Effect of an Equalizing Movement on Φ_α (Principle of Transfers)

Below, we show how any index of this family changes when an individual of a group moves from one unit to another in which that group has a smaller representation. In other words, we calculate the derivative of these indices when there is an equalizing movement for that group (without replacement), as the principle of transfers requires. To do this, we follow steps analogous to those followed by Reardon and Firebaugh (2002) to prove that *H,* which is equal to *M* divided by its maximum, fulfills this property.

If we denote by $r_{gj} \equiv \frac{\frac{c_j^g}{t_j}}{\frac{C^g}{T}}$, we can express Φ_α as:

$$\Phi_\alpha = \sum_g \frac{C^g}{T} \sum_j \frac{t_j}{T} f_\alpha(r_{gj}),$$

where $f_\alpha(r_{gj}) = \frac{1}{\alpha(\alpha-1)}\left(r_{gj}^\alpha - 1\right)$ if $\alpha > 0$ and $\alpha \neq 1$ and $f_\alpha(r_{gj}) = r_{gj} \ln(r_{gj})$ if $\alpha = 1$.

Let us assume, for simplicity, that the transfer involves group 1, so that *x* individuals of this group move from unit *i* to unit *h*, where $\frac{c_i^1}{t_i} > \frac{c_h^1}{t_h}$ (i.e., this is an equalizing movement for group 1). In other words, $r_{1i} > r_{1h}$. This means that after the transfer, the size of group 1 in unit *i* is $c_i^1 - x$ and in *h* is $c_h^1 + x$, whilst the population in *i* is $t_i - x$ and in *h* is $t_h + x$. To determine how this transfer affect overall segregation, we need to explore the next derivative (where $\pi_{gj} \equiv \frac{c_j^g}{t_j}$):

$$\frac{d\Phi_\alpha}{dx} = \sum_g \sum_j \frac{C^g}{T} \frac{t_j}{T} \frac{df_\alpha}{dr_{gj}} \frac{dr_{gj}}{d\pi_{gj}} \frac{d\pi_{gj}}{dx} + \sum_g \sum_j \frac{C^g}{T} \frac{1}{T} \frac{dt_j}{dx}\, f_\alpha(r_{gj}).$$

Note that $\frac{dr_{gj}}{d\pi_{gj}} = \frac{T}{C^g}$, $\frac{dt_i}{dx} = -1$, $\frac{dt_h}{dx} = 1$, and $\frac{dt_j}{dx} = 0$ for $j \neq i, h$. Also note that $\frac{d\pi_{1i}}{dx} = \frac{c_i^1 - t_i}{{t_i}^2}$, $\frac{d\pi_{1h}}{dx} = \frac{t_h - c_h^1}{{t_h}^2}$ and $\frac{d\pi_{1j}}{dx} = 0$ for $j \neq i, h$, whilst for the other groups $\frac{d\pi_{gi}}{dx} = \frac{c_i^g}{{t_i}^2}$, $\frac{d\pi_{gh}}{dx} = \frac{-c_h^g}{{t_h}^2}$ and $\frac{d\pi_{gj}}{dx} = 0$ for $j \neq i, h$. Taking all of this into account, it is easy to prove that:

$$\frac{d\Phi_\alpha}{dx}=\frac{1}{T}\left[\sum_g \frac{df_\alpha}{dr_{gi}}\frac{c_i^g}{t_i}-\sum_g \frac{df_\alpha}{dr_{gh}}\frac{c_h^g}{t_h}+\frac{df_\alpha}{dr_{1h}}-\frac{df_\alpha}{dr_{1i}}+\sum_g \frac{C^g}{T}\Big(f_\alpha(r_{gh})-f_\alpha(r_{gi})\Big)\right].$$

When $\alpha = 1$, given that $\frac{df_1}{dr_{gj}} = 1 + \ln(r_{gj})$ for any g and j, we get:

$$\frac{d\Phi_1}{dx}=\frac{1}{T}\ln\left(\frac{r_{1h}}{r_{1i}}\right),$$

which is negative, given that $\frac{r_{1h}}{r_{1i}} < 1$. In other words, the M index decreases after the transfer and, therefore, it fulfills the transfer principle (as shown in Reardon and Firebaugh, 2002).

When $\alpha > 0$ and $\alpha \neq 1$, given that $\frac{df_\alpha}{dr_{gj}} = \frac{r_{gj}{}^{\alpha-1}}{\alpha-1}$ for any g and j, we get:

$$\frac{d\Phi_\alpha}{dx}=\frac{1}{T}\left\{\frac{1}{\alpha}\sum_g\left(\frac{C^g}{T}\right)^{1-\alpha}\left[\left(\frac{c_i^g}{t_i}\right)^{\alpha}-\left(\frac{c_h^g}{t_h}\right)^{\alpha}\right]+\frac{1}{\alpha-1}\left(\frac{C^1}{T}\right)^{1-\alpha}\left[\left(\frac{c_h^1}{t_h}\right)^{\alpha-1}-\left(\frac{c_i^1}{t_i}\right)^{\alpha-1}\right]\right\},$$

and we cannot guarantee that this derivative is always negative after the transfer.

The same example offered by Reardon and Firebaugh (2002) to show that the *C* index does not satisfy the transfer principle (p. 63) can be used to prove that other members of this family with $\alpha > 1$ also do not satisfy it. On the other hand, simply by changing the null presence of group B in unit 3 in their example to a sufficiently large value (for example, 1000), we can show that indices with $0 < \alpha < 1$ also do not satisfy it.

The Local Segregation Curve (Unrepresentativeness)

Drawing on Alonso-Villar and Del Río (2010), we develop here a curve for measuring the unrepresentativeness of a unit *j* (Fig. 2). To build this curve, denoted by S_j, we must rank the groups in ascending order of the ratio $\frac{c_j^g}{C^g}$. Then, using that ranking, we plot the cumulative shares of the groups ($\sum_{i\leq g}\frac{C^i}{T}$) on the horizontal axis and the cumulative proportions of the groups in unit j ($\sum_{i\leq g}\frac{c_j^i}{t_j}$) on the vertical axis. Namely, if we denote by $\tau_g \equiv \sum_{i\leq g}\frac{C^i}{T}$ the population proportion of the first g groups, the curve of unrepresentativeness at point τ_i is given by $S_j(\tau_g) = \sum_{i\leq g}\frac{c_j^i}{t_j}$. If unit j were fully representative, this curve would be equal to the 45º line. The further the curve is from the that line, the greater the unrepresentativeness of unit j.

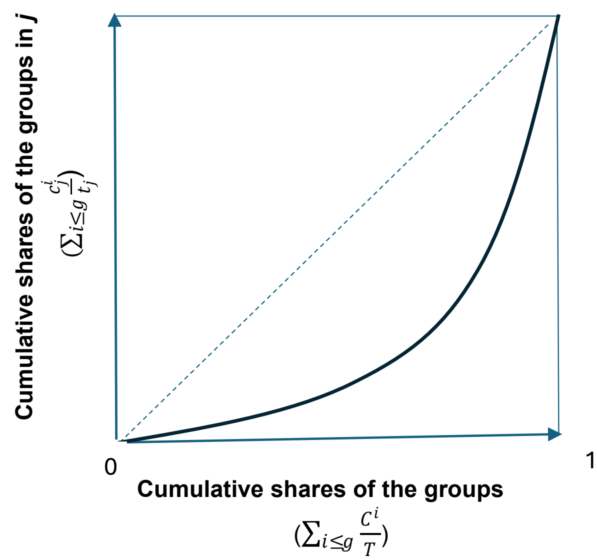

Figure 2. Local segregation curve of unit j

A curve is said to dominate another if there is no point at which it lies below it and there is at least one point at which it lies above it.

Sensitivity Towards Disequalizing Movements between Groups (type I): **Index ψ_1**

To prove that ψ_1 satisfies this property, note that when δ individuals from group 1 in unit j are replaced by δ individuals from group 2, where $c_j^1 < c_j^2$ and $C^1 = C^2$, the difference in the index between the final distribution and the initial one (without changing the size of group 1 and 2 in the total population) is equal to:

$$\Delta\psi_1 = \frac{c_j^1-\delta}{t_j} ln\left(\frac{\frac{c_j^1-\delta}{t_j}}{\frac{C^1}{T}}\right) - \frac{c_j^1}{t_j} ln\left(\frac{\frac{c_j^1}{t_j}}{\frac{C^1}{T}}\right) + \frac{c_j^2+\delta}{t_j} ln\left(\frac{\frac{c_j^2+\delta}{t_j}}{\frac{C^2}{T}}\right) - \frac{c_j^2}{t_j} ln\left(\frac{\frac{c_j^2}{t_j}}{\frac{C^2}{T}}\right),$$

which can also be written as:

$$\Delta\psi_1 = \frac{1}{t_j}\left[(c_j^1-\delta)ln(c_j^1-\delta) - c_j^1 ln(c_j^1) + (c_j^2+\delta)ln(c_j^2+\delta) - c_j^2 ln(c_j^2)\right].$$

Given that function $xlnx$ is convex, it is easy to prove that $\Delta\psi_1 > 0$. In other words, the level of unrepresentativeness of unit *j* is greater after a disequalizing movement. To prove this, we analyze three possible scenarios depending on the positions of $y \equiv c_j^1$ and $z \equiv c_j^2$ (depicted in Figures 3a-3c):

a) Consider the case where *y* and *z* are on the decreasing part of the function (Figure 3a), i.e., they are lower than e^{-1}. In this scenario, when *y* decreases by δ and *z* increases by the same amount (and after this change it is still lower than e^{-1}), $\Delta\psi_1 > 0$, given that $(y-\delta)\ln(y-\delta) - ylny$ is a positive number of a higher magnitude than $(z+\delta)\ln(z+\delta) - zlnz$, which is negative. On other hand, if $(z+\delta) > e^{-1}$, then $(z+\delta)\ln(z+\delta) - zlnz$ is either a negative number smaller than before or a positive number, which reinforces the fact that $\Delta\psi_1 > 0$.

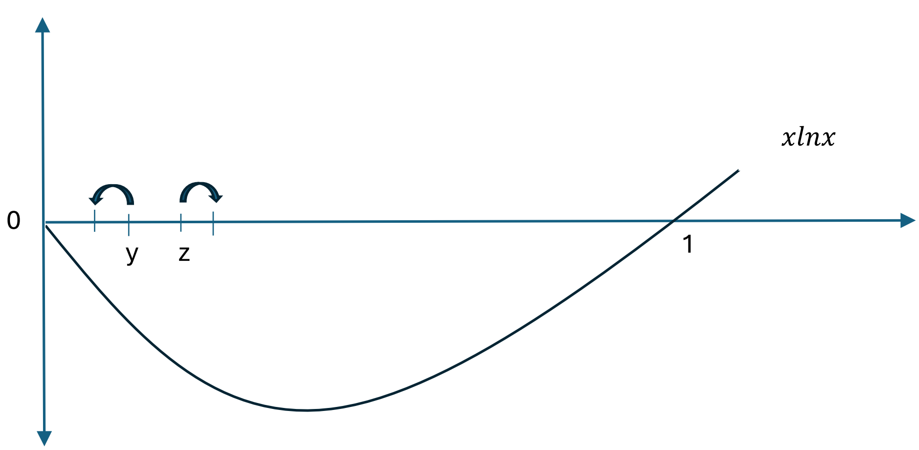

Figure 3a. Change in the index when *y* and *z* are on the decreasing part of $xlnx$

b) When y is on the decreasing part and z is on the increasing part (Figure 3b), both $(y-\delta)\ln(y-\delta) - ylny$ and $(z+\delta)\ln(z+\delta) - zlnz$ are positive numbers. This is the case not only when z is lower than 1, but also when it is higher than 1. Therefore, $\Delta\psi_1 > 0$.

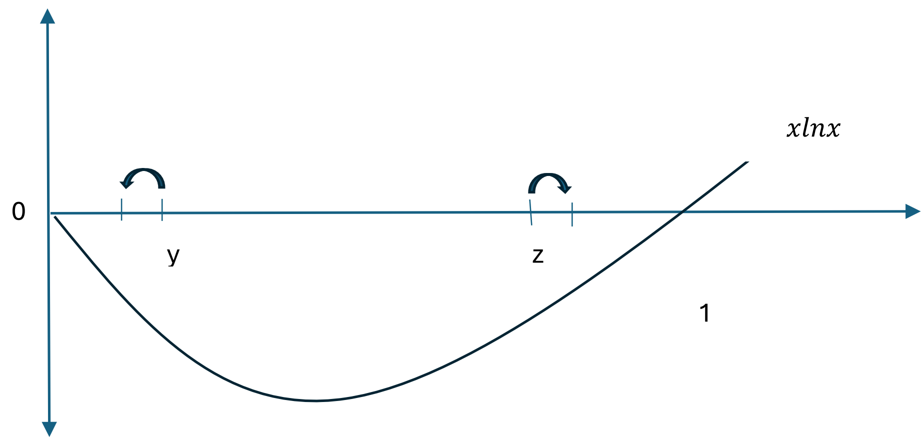

Figure 3b. Change in the index when y and z are on the decreasing and increasing part of $xlnx$, respectively

c) When y and z are on the increasing part of the function (Figure 3c), $(z+\delta)\ln(z+\delta) - zlnz$ is a positive number and $(y-\delta)\ln(y-\delta) - ylny$ is either a negative number of a lower magnitude or a positive number. Therefore, $\Delta\psi_1 > 0$.

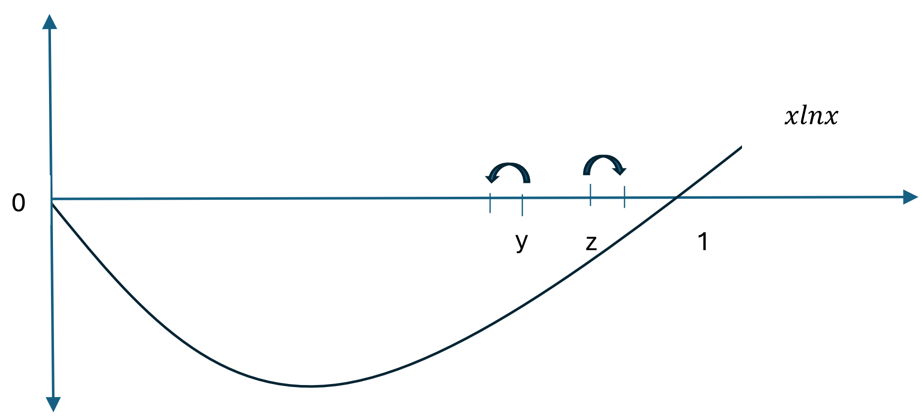

Figure 3c. Change in the index when y and z are on the increasing part of $xlnx$

Sensitivity Towards Disequalizing Movements between Groups (type I): **Index ψ_α**

To prove that ψ_α (with $\alpha > 0$ and $\alpha \neq 1$) satisfies this property, note that when δ individuals from group 1 in unit *j* are replaced by δ individuals from group 2 (without changing the size of group 1 and 2 in the total population), where $c_j^1 < c_j^2$ and $C^1 = C^2$, the difference in the index between the final distribution and the initial one is:

$$\Delta\psi_\alpha = \frac{1}{\alpha(\alpha-1)}\left\{\frac{C^1}{T}\left[\left(\frac{\frac{c_j^1-\delta}{t_j}}{\frac{C^1}{T}}\right)^\alpha - 1\right] - \frac{C^1}{T}\left[\left(\frac{\frac{c_j^1}{t_j}}{\frac{C^1}{T}}\right)^\alpha - 1\right] + \frac{C^2}{T}\left[\left(\frac{\frac{c_j^2+\delta}{t_j}}{\frac{C^2}{T}}\right)^\alpha - 1\right] - \frac{C^2}{T}\left[\left(\frac{\frac{c_j^2}{t_j}}{\frac{C^2}{T}}\right)^\alpha - 1\right]\right\} =$$

$$= \frac{1}{\alpha(\alpha-1)}\left(\frac{C^1}{T}\right)^{1-\alpha}\left(\frac{1}{t_j}\right)^\alpha\left[\left(c_j^1-\delta\right)^\alpha - \left(c_j^1\right)^\alpha + \left(c_j^2+\delta\right)^\alpha - \left(c_j^2\right)^\alpha\right].$$

When $0 < \alpha < 1$, function x^α is positive, increasing, and concave (for positive *x* values as in our case) and, therefore, $\left(c_j^1-\delta\right)^\alpha - \left(c_j^1\right)^\alpha + \left(c_j^2+\delta\right)^\alpha - \left(c_j^2\right)^\alpha < 0$. Taking into account that $\frac{1}{\alpha(\alpha-1)} < 0$, hence $\Delta\psi_\alpha > 0$.

When $\alpha > 1$, function x^α is positive, increasing, and convex, and therefore, $\left(c_j^1-\delta\right)^\alpha - \left(c_j^1\right)^\alpha + \left(c_j^2+\delta\right)^\alpha - \left(c_j^2\right)^\alpha > 0$. Taking into account that in this case $\frac{1}{\alpha(\alpha-1)} > 0$, it follows that $\Delta\psi_\alpha > 0$.