Measuring Social Environment Mobility*

Son Thierry Ly[†] Arnaud Riegert[‡]

March 2015

Document de travail du Département des études économiques

*Les auteurs souhaitent remercier Luc Behaghel, Sandra Black, Éric Dubois, Laurent Gobillon, Malik Koubi, Éric Maurin, Magne Mogstad, Fabrice Murat, Corinne Prost ainsi que les participants au lunch séminaire d'économie appliquée de l'École d'économie de Paris et au séminaire du département des études économiques de l'Insee pour leurs commentaires, questions et remarques qui ont permis de faire progresser ce document de travail.
†Paris School of Economics. son.thierry.ly@ens.fr
‡Insee – D2E / Paris School of Economics. arnaud.riegert@insee.fr

Measuring Social Environment Mobility

Abstract

Individuals experience a diversity of social environments throughout their lives. When measuring the degree to which different social groups are separated from each other, this fact is often overlooked: standard segregation indices always measure spatial separation at a given point in time. These segregation indices only tell one part of the story, just like income inequality indices do not take into account the fact that individuals are mobile across the income distribution throughout their lives. This paper introduces the notion of social environment mobility (SEM) and proposes tools and a methodology to analyze it. We show that unlike income mobility, SEM cannot erase segregation in the long run, and we derive an upper bound on SEM indices. We illustrate this concept using data on segregation in French middle schools. Our results show that SEM has a fairly high equalizing effect on within-school segregation but a low overall effect due to low mobility between schools.

Keywords: Mobility, Segregation

JEL codes: I24, D63, D85

Comment mesurer la mobilité entre environnements sociaux ?

Résumé

Au cours de leur vie, les individus fréquentent des environnements sociaux divers. Ce fait est généralement ignoré par les mesures classiques de ségrégation, c'est-à-dire les mesures de l'ampleur de la séparation entre des groupes sociaux différents. Tous les indices de ségrégation rencontrés dans la littérature ont en effet pour point commun de ne mesurer cette séparation spatiale qu'à un instant donné. Ainsi, ces indices ne racontent qu'une partie de l'histoire, tout comme les indices d'inégalités de revenus ne prennent pas en compte le fait que les individus se déplacent dans la distribution de revenus au cours de leur vie. Ce document de travail introduit une notion de mobilité en environnements sociaux et propose des outils et une méthodologie permettant de l'étudier. Nous montrons en particulier que contrairement à la mobilité de revenus, la mobilité en environnements sociaux ne permet pas d'effacer complètement la ségrégation sur le long terme. Nous calculons la borne supérieure des indices de mobilité en environnements sociaux que nous introduisons. Nous proposons enfin une illustration de cette notion sur la base de données concernant la ségrégation dans les collèges français. Nous montrons que la mobilité en environnements sociaux équilibre en partie la ségrégation intra-établissement mais a un effet très limité sur la ségrégation totale en raison d'une faible mobilité inter-établissement.

Mots-clés : Mobilité, Ségrégation

Classification JEL : I24, D63, D85

Introduction

School or residential segregation between social or ethnic groups has been a great concern for social scientists and policymakers for many decades. By depriving minorities from key resources in their environment, the spatial separation of individuals by social background or ethnicity may hamper their opportunities and thus reinforce inequalities (Crain & Strauss, 1985; Wells & Crain, 1994; Cutler & Glaeser, 1997). Segregation may also be viewed as a threat to democracy by developing uneasiness about interracial contacts, if not racial hostility (Braddock II & McPartland, 1989). This is suggested in particular in recent works on social networks, which show that geographical proximity could strongly reduce the homophilic feature of interactions between individuals (Currarini *et al.*, 2009; Marmaros & Sacerdote, 2006). For example, Camargo *et al.* (2010) report that different-race students are as likely to become friends as same-race students once they have been (randomly) assigned as roommates (also see Marmaros & Sacerdote, 2006). By depriving individuals from diversified interactions at an early stage of development, school segregation may therefore have long run implications on relationships between social or racial groups[1].

However, the high interest of scholars on segregation issues has stalled at the conceptual and technical challenges raised by its measurement. A rich literature has spawned from this issue, with recent papers (Echenique & Fryer, 2007; Frankel & Volij, 2011) that added new measures and concepts to the existing battery of segregation indices (Duncan & Duncan, 1955; Massey & Denton, 1988). Interestingly enough, all these papers share a common limited feature: they only consider segregation in a cross-sectional perspective. In other words, all segregation indices that have been developed only measure how much individuals from different groups are spatially separated from each other *at a given point of time*. Yet the scope of the conclusions that can be drawn from such indices is somewhat restricted: if individuals move a lot from an environment to another across years, segregation in a longitudinal perspective may be substantially lower than it appears from a cross-sectional point of view. For instance, cross-sectional indices may inform us about the extent to which students in a school are segregated between classes with mostly white students on the one hand, and classes with mostly black students on the other hand. They do not tell us whether each student stays every year in the same kind of classroom, or if instead, they tend to move between "rich" and "poor" classes every year. The implications for students' concrete experiences of social isolation is very different depending on the scenario.

This paper proposes to fill this gap in the literature by providing the first theoretical and empirical analysis

[1]In this direction, Camargo *et al.* (2010) also report interesting long run effects of diversity on white students. According to the authors, they are more likely to have black friends in the future when they are assigned to a black roommate in college.

of social environment mobility and its consequences on segregation. In doing so, our work closely relates to the strand of literature that embeds earnings mobility in the measurement of income inequality, following Shorrocks (1978b,a) in particular. By examining inequalities of permanent or lifetime income instead of "snapshot" income, this literature has clearly established the equalizing force of mobility (see *e.g.* Bonhomme & Robin, 2009, in the case of France). For example, Bowlus & Robin (2004) report that lifetime income inequality in the U.S. is 40% less than in instantaneous earnings. Obviously, considering income mobility also yields a different vision of the evolution of income inequality over time (Moffitt & Gottschalk, 2002) or of cross-country comparisons (Gottschalk & Spolaore, 2002). In the same spirit, *social environment mobility* (SEM throughout the paper) may have an equalizing effect on the distribution of environments across individuals.

The paper is organized as follows. In section 1, we review the literature on the measurement of income mobility and of segregation. We explain why considering income mobility leads to lower indices of income inequality and we present the main tools to measure income mobility. We also present the segregation indices that prevail in the literature on segregation. In section 2, we introduce the concept of social environment mobility. We first show how segregation indices are related to income inequality indices, in order to adapt the income mobility indices to the case of social environment mobility. We present our main result which is that unlike income mobility, social environment mobility can not reach 100 percent as long as there is some degree of segregation at one point in time. In other words, while lifetime income inequality may be reduced to zero even if instantaneous inequality is not null, there is a lower bound in the case of segregation: it is not possible that all individuals in the population experience the same average exposure to a reference group if that reference group is not evenly distributed geographically at all times. We provide tools that help using our social environment mobility index with real data and we conclude with a case study on French middle schools (section 3).

1 Literature review

There is an important and still active literature on the measurement of income mobility and the measurement of segregation. Although social environment mobility (SEM) is conceptually different from income mobility, its measurement uses similar mathematical tools. In this section, we present some of the tools we use as a basis to build our SEM indices.

1.1 Measuring income inequality

This section gives a quick overview of the most common tools to measure income inequality. We consider a population of N individuals $i \in \{1, \dots, N\}$. Individual i has income y_i. The income distribution has a cumulative distribution function $\mathcal{F}(y)$ and average $\overline{y}$.

Inequality indices are functions of the income distribution $I(\mathcal{F})$ that belong to the unit interval $[0, 1]$. They equal zero when all individuals have the same income and one when all the income is owned by one individual only. Different indices may have different variations between these two extreme cases depending on their axiomatic properties.

The *Lorenz curve*, introduced by Lorenz (1905), is one of the most popular tools to study income inequality. The curve is defined by the following function:

$$\mathcal{L}(u) = \frac{\displaystyle\int_0^u \mathcal{F}^{-1}(t)\,\mathrm{d}t}{\displaystyle\int_0^1 \mathcal{F}^{-1}(t)\,\mathrm{d}t} = \frac{1}{\overline{y}} \int_0^u \mathcal{F}^{-1}(t)\,\mathrm{d}t \qquad \forall u \in [0, 1] \tag{1}$$

Note that the joint knowledge of $\mathcal{L}$ and $\overline{y}$ is equivalent to knowing the full distribution of income (for instance, $\overline{y}\mathcal{L}'(u) = \mathcal{F}^{-1}(u)$).

The interpretation of the Lorenz curve is the following. $\mathcal{L}(20\,\%)$ is the share of all the income that is earned by the $20\,\%$ individuals with the lowest income.[2] It follows that $\mathcal{L}(u) \leq u$ for all u and the Lorenz curve is below the 45-degree line. In addition, $\mathcal{L}(0) = 0$ and $\mathcal{L}(1) = 1$, so the curve and the line cross at $(0, 0)$ and $(1, 1)$. If there is no inequality (everybody has the same income), $\mathcal{L}(u) = u$ for all u. At the other end, when all income is owned by one individual, $\mathcal{L}(u) = 0$ if $u < 1$ and $\mathcal{L}(1) = 1$.[3] Between these two extremes, the Lorenz curve lies between the x-axis and the 45-degree line. The closer it is to the 45-degree line, the lower inequality is.

The most popular inequality index – the Gini index – has a simple geometric relationship with the Lorenz curve. It is equal to the area between the Lorenz curve and the 45-degree line, divided by its highest possible value (1/2) so that it belongs to the unit interval. It can be written:

$$G = \frac{\displaystyle\int_0^1 [u - \mathcal{L}(u)]\,\mathrm{d}u}{1/2} = 1 - 2\int_0^1 \mathcal{L}(u)\,\mathrm{d}u \tag{2}$$

[2] Indeed, $\int_0^u \mathcal{F}^{-1}(t)\,\mathrm{d}t = \int_0^{\mathcal{F}^{-1}(u)} t f(t)\,\mathrm{d}t$ is the total income of the fraction u of the poorest individuals in the population, therefore $\mathcal{L}(u)$ is equal to that income divided by the total income of the whole population.

[3] In reality, $\mathcal{L}(u) = 0$ if $u \leq 1 - 1/N$, but we assume that $N \gg 1$.

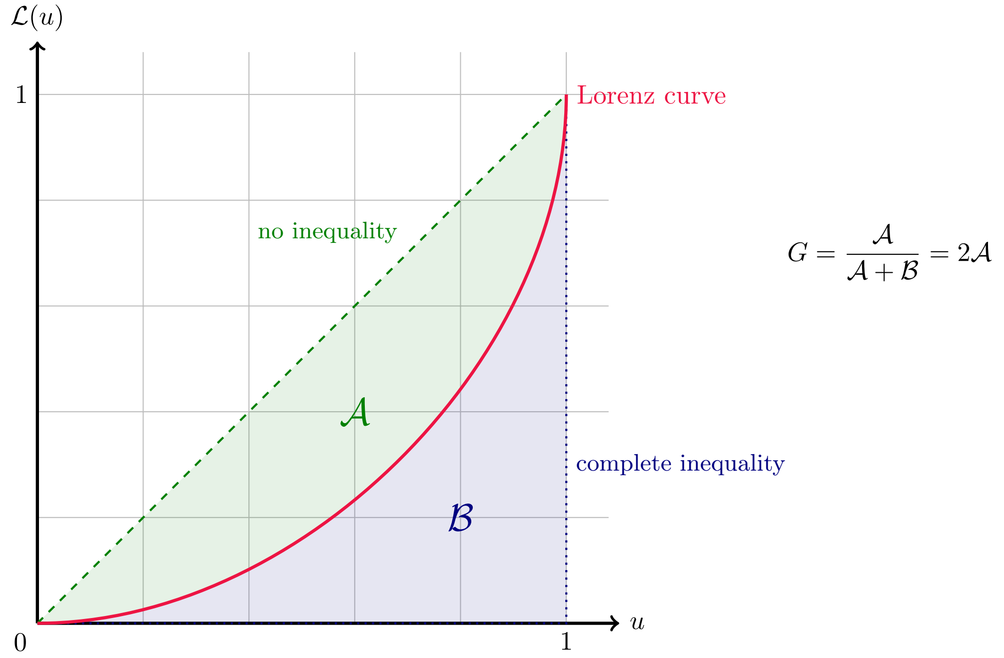

Figure 1.1: The Lorenz curve for income inequality

Note that the Gini index is also equal to half the average absolute difference between any two individuals' incomes, *i.e.*

$$G = \frac{1}{2N^2}\frac{1}{\bar{y}}\sum_{i=1}^{N}\sum_{j=1}^{N}|y_i - y_j| \tag{3}$$

See appendix A.1 for a proof of the equivalence of definitions (2) and (3).

Remark: The Gini index belongs to a broader class of inequality measures which have the following form (Yaari, 1988; Aaberge, 2001):

$$J_p(\mathcal{L}) = 1 - \int_0^1 p(u)\,\mathrm{d}\mathcal{L}(u) \tag{4}$$

where the weighting function p is a non-increasing function with unit integral and $p(1) = 0$. The choice of p belongs to the social planner and depends on the desired axiomatic properties of the inequality index. The Gini index is associated with the weighting function $p_G(u) = 2(1-u)$. Indeed, if we integrate by parts:

$$\begin{aligned} J_{p_G}(\mathcal{L}) &= 1 - \int_0^1 2(1-u)\,\mathrm{d}\mathcal{L}(u) \\ &= 1 - [2(1-u)\mathcal{L}(u)]_0^1 + \int_0^1 (-2)\mathcal{L}(u)\,\mathrm{d}u \\ &= 1 - 2\int_0^1 \mathcal{L}(u)\,\mathrm{d}u \\ &= G \end{aligned}$$

1.2 The equalizing force of income mobility

The purpose of income mobility indices is to show that instantaneous income inequality indices do not tell the full story of the income distribution. Income inequality is the combination of two very different phenomena at work. First, people earn different incomes throughout their lives (for instance their incomes increase with seniority or decrease when they retire), which generates inequality when looking at the whole population at one point in time. Second, within age groups, a number of personal characteristics affect income (skills, education, etc.). These two contributions should be analyzed separately: different societies may have different standards as to what level of inequality of each type is acceptable.

Income mobility indices measure to what extent people are likely to move in the income distribution. There are essentially two approaches to measuring this phenomenon. The first approach consists in measuring correlations between current and past rankings in the income distribution. The second approach consists in comparing the distribution of instantaneous incomes to the distribution of lifetime incomes.

Consider the following example with two individuals in two periods of time, and two scenarios. In scenario A, the income distribution is $(y_1 = 1,000; y_2 = 2,000)$ in both periods. In scenario B, the income distribution is $(y_1 = 1,000; y_2 = 2,000)$ in period 1 and incomes are swapped in period 2: $(y_1 = 2,000; y_2 = 1,000)$. In both scenarios and in both periods, any instantaneous inequality index would have the exact same value. For instance, the Gini index would equal $G = 1/6$. These two scenarios could thus not be distinguished by analyzing instantaneous inequality indices solely. Yet scenario B exhibits more income mobility than scenario A, as both individuals exchange places in the income distribution: therefore, the correlation between current and past rankings is lower, and the inequality in lifetime incomes is also lower (zero if we consider that the interest rate is zero).

High income mobility is a signal of equal opportunities: if people are likely to shift in the distribution, this means that people at the bottom of the distribution may make their way up to the top, and conversely people who are initially in the top of the distribution do not automatically keep their privileges.

If there is no income mobility at all, *i.e.* if all individuals keep the same (relative) income throughout their life, individuals also keep their positions in the income distribution, and the lifetime income is proportional to any instantaneous income. Therefore, the *perfect income rigidity* scenario leads to perfect correlation between present and past positions in the distribution, and identical measures of inequality based on instantaneous of lifetime income.

At the opposite end, defining *perfect income mobility* is more complicated. One possible definition would be to consider the case where past incomes have no influence on the current income (Prais, 1955). In this case, the correlation between past and current positions in the distribution would be equal to zero, and an index of inequality based on lifetime income would tend towards zero as the number of time periods would grow to infinity.

However, this definition does not describe the highest possible mobility. For instance, we could imagine a case where poor people have better odds being rich in the next period, while rich people will more likely get poor. In this case, the correlation between past and current ranks in the distribution would be negative, and the inequality index based on lifetime income may go to zero very rapidly.

In the rest of this section, we present income mobility indices that are based on the two classes of methods introduced above, *i.e.* measuring the correlation between past and current rankings, and comparing instantaneous and permanent income distributions. In order to compare present and past positions in the income distribution, a simple possibility is to use autocorrelation coefficients of the series of incomes. However, the literature on social mobility uses transition matrices, from which mobility indices can be derived. We also show two indices that are based on the comparison of instantaneous and lifetime income.

All the indices introduced here take values between zero and one. Mobility is equal to zero in the perfect rigidity scenario. The remainder of the axiomatic properties may vary from one index to another and can be found in the referenced papers.

1.2.1 Transition matrices

Transition matrices are a popular tool to measure social mobility, *i.e.* to observe how individuals are likely to move between categories of the population at different points in time. These matrices are especially useful when the population is partitioned into categories such as the socioeconomic status, but they may also be used to study mobility within the distribution of a continuous variable such as the income, by using income quantiles as categories.

Consider a population partitioned into n categories and two points in time ($t \in \{1; 2\}$). Denote $c_i^t \in \{1; \ldots; n\}$ the category of individual i at time t. The transition matrix is $T = (t_{jk})_{1 \leq j,k \leq n}$ where $t_{jk} = \mathbb{P}(c_i^2 = k | c_i^1 = j)$

is the share of the population that was in category j in the first period ending up in category k in the second period. A transition matrix should be stochastic, *i.e.* that the coefficients in each row must add up to one:

$$\sum_{k=1}^{n} t_{jk} = \sum_{k=1}^{n} \mathbb{P}(c_i^2 = k | c_i^1 = j) = 1 \tag{5}$$

The perfect rigidity scenario's transition matrix is the identity matrix. In the perfect mobility scenario, the initial category has no influence on the final category: all rows are identical. Shorrocks (1978b) argued that the perfect mobility scenario should be associated with a mobility index value of one. However, this axiom associated with a monotonicity axiom would lead to values higher than one when mobility is more than perfect. Shorrocks (1978b)'s recommendation is to limit the value of the index to one in those cases, that he considers very unlikely in the context of income mobility.

An example of such an index is the following:

$$M_T = \min\left(\frac{n - \text{Trace}(T)}{n-1}; 1\right) \tag{6}$$

Indeed, in the case of perfect rigidity, T is the identity matrix and has trace n, therefore $M_T = 0$. If there is perfect mobility, property (5) and the fact that the rows are identical leads to the trace being equal to one, therefore $M_T = 1$.

Another difficulty with indices based on transition matrices is that they greatly depend upon the definition of the categories. For instance, using income percentiles rather than deciles would automatically increase the value of the income mobility index, as the diagonal elements would be lower in proportion. The following sections introduce indices that adress these two limitations.

1.2.2 Comparing instantaneous and lifetime inequality

A second class of income mobility indices has been developed by Shorrocks (1978a), published just three months after Shorrocks (1978b). The intuition is that when income mobility is high, inequality indices based on lifetime incomes should be lower than those based on instantaneous incomes. Calling $\mathcal{F}_t$ the income cumulative distribution function (cdf) in period t, $\mathcal{F}^*$ the cdf of lifetime income, and $I(\mathcal{F})$ an income inequality index (such as the Gini index), an income mobility index is defined by:

$$M = 1 - \frac{I\left(\mathcal{F}^*\right)}{\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} I(\mathcal{F}_t)} \tag{7}$$

where I is a strictly convex function of the relative income distribution. This restriction on I ensures that M belongs to the unit interval (Shorrocks, 1978a). Consider the two extreme cases leading to $M = 0$ and $M = 1$.

In the perfect rigidity scenario, the instantaneous and lifetime income distributions are identical (up to a multiplicative constant in income), therefore the inequality indices – that are functions of *relative* incomes only – are equal, and M is equal to zero.

M equals one if and only if inequality in lifetime income $I(\mathcal{F}^*)$ is equal to zero. This may happen in only three scenarios: (i) if there is no inequality at any point in time, (ii) if income mobility is perfect and the number of periods is infinite or (iii) if income mobility is more than perfect so that inequality in a given period are compensated for in a different period. The first scenario is trivial and the second scenario has no empirical application. The third scenario is what Shorrocks (1978a) calls *complete income mobility*, as opposed to perfect income mobility.

1.2.3 Income mobility curves

A third method to measure income mobility has recently been introduced by Aaberge & Mogstad (2014), that also relies on the comparison of instantaneous and lifetime income. The authors extend the concept of Lorenz curve and introduce an income mobility curve which is defined as the difference between the Lorenz curve of the actual lifetime income distribution $\mathcal{F}^*$ and the curve based on a hypothetical, reference lifetime income distribution in perfect rigidity scenario, $\mathcal{F}_R^*$. In the reference distribution $\mathcal{F}_R^*$, "the rank of each individual is the same in every period; this distribution can be formed by assigning the lowest income in every period to the poorest individual in the first period, the second lowest to the second poorest, and so on.". Note that this definition does not give a particular role to the distribution in the first period: keeping the rankings constant and equal to the rankings in any period would lead to the exact same distribution.

Call $\mathcal{L}^*$ the Lorenz curve of the lifetime income distribution $\mathcal{F}^*$ and $\mathcal{L}_R^*$ the Lorenz curve of the reference distribution $\mathcal{F}_R^*$. The income mobility curve is defined as the difference between the two Lorenz curves:

$$\mathcal{M}(u) = \mathcal{L}^*(u) - \mathcal{L}_R^*(u) \tag{8}$$

Aaberge & Mogstad (2014) show that $\mathcal{M}(u) \geq 0$ for all $u \in [0;1]$, *i.e.* that the Lorenz curve of the reference distribution is below the Lorenz curve of actual lifetime income (which means that it is more unequal).

Mobility curves can be synthesized into income mobility indices in the same way Lorenz curves can be synthesized into income inequality indices. For instance, an income mobility index can be obtained by integrating a weighting function along the income mobility curve:

$$\Lambda_p(\mathcal{M}) = \int_0^1 p(u)\,\mathrm{d}\mathcal{M}(u) = \int_0^1 p(u)\,\mathrm{d}(\mathcal{L}^* - \mathcal{L}_R^*)(u) = J_p(\mathcal{L}_R^*) - J_p\left(\mathcal{L}^*\right) \tag{9}$$

If we consider the same weighting function as for the Gini index ($p(u) = 2(1-u)$), $\Lambda_p(\mathcal{M})$ is simply the difference between the Gini indices of both distribution. Aaberge & Mogstad (2014) argue that taking the difference rather than the ratio is preferrable in that a ratio whose denominator is very small will be very sensitive to small changes in the numerator. It will be equal to zero if the actual distribution of lifetime income is the same as the reference distribution that corresponds to the perfect rigidity scenario. It will be equal to one if the actual lifetime income distribution is perfectly equal *and* if the reference distribution is completely unequal, *i.e.* if all income belongs to one individual only.

1.3 Measuring segregation

Segregation is the fact that people with different characteristics (*e.g.* race, gender, socioeconomic status) are isolated from each other. Its measurement raises a number of conceptual and technical issues (Massey & Denton, 1988). In our case, we choose to focus on segregation indices that measure the extent to which the population is evenly distributed across units (*e.g.* cities or schools) depending on their characteristics. This definition is associated with a number of popular segregation indices such as the dissimilarity index, the Gini index, the normalized exposure and the entropy index. Frankel & Volij (2011) provide an excellent review of the main segregation indices, including their definitions and axiomatic properties.

We can see segregation indices as measures on an undirected graph where nodes are individuals and edges are social relations. We assume that each individual belongs to one of two groups denoted A and B.[4] Group A will be called the *reference group* (*e.g.* the rich, the black, the women, etc.) and group B will contain everyone else in the population. For instance, if group A denotes black individuals, group B will contain all non-black individuals (not only whites). The reference group may be the majority group or a minority, indifferently: we call $p \in [0;1]$ its share in the population. Each individual $i \in \{1,\ldots,N\}$ belongs either to group A ($U_i = 1$) or group B ($U_i = 0$). The *social environment* of individual i is the set of individuals that are connected to i through an edge, *including individual i herself.*[5] We denote μ_i the share of i's social environment that belongs to the reference group: it is called the *exposure to the reference group* of individual i.

In the two-group case, segregation indices are essentially measures of the inequality of the distribution of μ_i, *i.e.* they are indices of social environment inequality. However, there is one important difference between income inequality and social environment inequality, which is that individuals *are* each other's social environment. Therefore, the distribution of μ_i is subject to constraints. In particular, a widespead assumption is that the graph consists of a set of isolated, complete subgraphs, *i.e.* that individuals interact in clusters, as shown in Figure 1.2. With this hypothesis, the clusters are denoted $k \in \{1,\ldots,K\}$, each cluster contains N_k individuals and has a share p_k of individuals from the reference group. Therefore, if individual i belongs to cluster k, $\mu_i = p_k$. The average of μ_i over the whole population is then equal to p:

$$\mathbb{E}(\mu_i) = \frac{1}{N}\sum_{i=1}^{N}\mu_i = \frac{1}{N}\sum_{k=1}^{K}N_k p_k = p$$

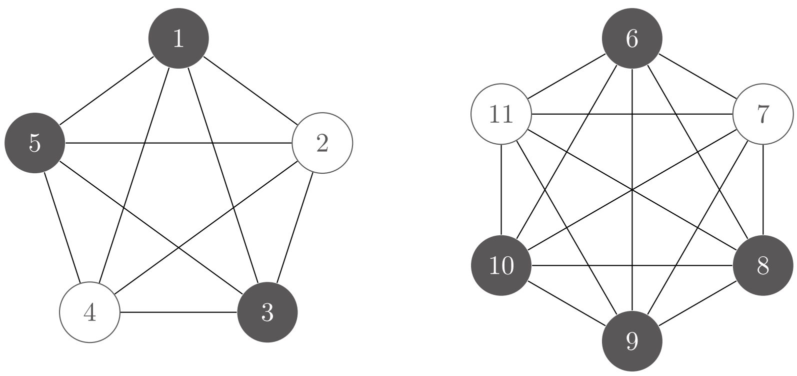

Figure 1.2: A social network graph with interactions in clusters. Filled circles denote individuals from group A and plain circles denotes individuals from group B. The average exposure to the reference group is 60 percent in the left cluster, 67 percent in the right cluster.

This hypothesis is very common in the literature on segregation, and most segregation indices are defined

[4] Segregation indices can be defined with more than two groups, as detailed in Frankel & Volij (2011). We focus on the two-group case for simplicity's sake; adding more groups has no fundamental impact on the analysis.

[5] The choice of including individual i in her own social environment or not is quite arbitrary: both options have advantages and drawbacks and they have their own justifications. The numerical difference is small when individuals have a large number of social relations. We choose to include individual i in her own social environment in order to simplify calculations.

based on these clusters. Below are the definitions of four popular segregation indices: the dissimilarity index (D), the Gini index (G), the normalized exposure (P), and the entropy index (H).

$$D = \frac{1}{2p(1-p)} \sum_{k=1}^{K} \frac{N_k}{N} |p_k - p| \tag{10}$$

$$G = \frac{1}{2p(1-p)} \sum_{k=1}^{K} \sum_{k'=1}^{K} \frac{N_k}{N} \frac{N_{k'}}{N} |p_k - p_{k'}| \tag{11}$$

$$P = \frac{1}{p(1-p)} \sum_{k=1}^{K} \frac{N_k}{N} (p_k - p)^2 \tag{12}$$

$$H = 1 - \frac{1}{h_2(p)} \sum_{k=1}^{K} \frac{N_k}{N} h_2(p_k) \tag{13}$$

where h_2 is the 2-category entropy function, defined by:

$$h_2(u) = u \log_2 \left(\frac{1}{u}\right) + (1-u) \log_2 \left(\frac{1}{1-u}\right) \tag{14}$$

Note that the definition of the Gini index is similar to the Gini index on income distribution defined in equation 3.

All these indices equal zero when $p_k \equiv p$ in all clusters, *i.e.* if every cluster has the same share of individuals from the reference group: in that case, there is no segregation at all. Conversely, they all equal one when $p_k \in \{0, 1\}$ in all clusters, *i.e.* if every cluster contains either only individuals from the reference group, or no individual from that group: there is no mixity in any of the clusters, segregation is complete.

The indices take different values for intermediate situations, they have different interpretations and different axiomatic properties, as detailed in Frankel & Volij (2011). For instance, the dissimilarity index D is equal to the minimal share of the population that would have to move to a different cluster in order to achieve full desegregation, normalized by its maximum value; the normalized exposure index P has three possible interpretations:

- First, it can be seen as the variance of p_k across all clusters, normalized by its highest possible value $p(1-p)$ which is reached if each cluster contains either only individuals from group A or only individuals from group B.

- Second, P is the share of between-clusters variance of the binary variable U_i (belong to the reference group), since $p_k = \mathbb{E}(U_i | i \in k)$. It is the R^2 of the regression of that variable on a cluster-level fixed effect.

- Third, calling μ^A the average exposure to group A *of the individuals in that group* and μ^B the average exposure to group A *of the individuals from group B*, the value of the index is equal to:

$$P = 1 - \frac{\mu^B}{p} = \mu^A - \mu^B \tag{15}$$

i.e. P is the exposure gap to the reference group between the two groups (proof in appendix A.3).

Note that since $P \geq 0$, these equalities yield the following double inequality:

$$\mu^B \leq p \leq \mu^A \tag{16}$$

The reference group always has a greater exposure to itself than average, while the rest of the population has a lower exposure to the reference group.

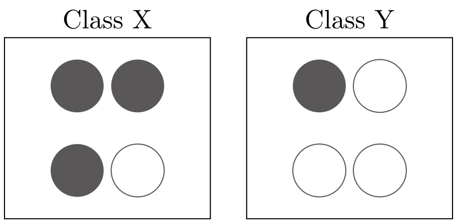

Figure 1.3: Segregation between upper class (full circles) and middle or lower class (plain circles) students.

In Figure 1.3, we consider the case of 8 students ($N = 8$) attending one of two classes ($K = 2$). Half of them are in the reference group A (full circles) and the other half are in the group B (plain circles): $p = 0.5$. The students in class X have an exposure to the reference group $p_X = 0.75$ while the students in class Y have an exposure $p_Y = 0.25$. The segregation indices equal $D = 0.5$, $G = 0.5$, $P = 0.25$ and $H = (3/4)\log_2(3) - 1 \simeq 0.19$.

2 Social environment mobility

The tools presented in the previous section lay out the grounds for building social environment mobility (SEM) indices. As seen in section 1.3, segregation indices are not straightforward adaptations of income inequality indices. Similarly, income mobility indices cannot be transposed directly to the case of SEM. In this section, we detail the fundamental differences and the adaptations they imply on the definition of SEM indices. For the sake of simplicity, we take the example of social segregation between upper class individuals (group A) and middle or lower class individuals (group B). We sometimes refer to the group B as lower class individuals, omitting the fact that group B contains both middle class and lower class individuals.

2.1 Breaking out of the clusters

While the definitions of the segregation indices defined in section 1.3 are convenient, they are not appropriate when studying SEM. Their main drawback is that they rely on the fact that individuals interact in clusters, which makes sense when looking at segregation at one point in time, but not when looking at SEM. For instance, in the case of school segregation, clusters may represent classes in a school, and those are not constant across school years. Individuals interact in clusters in each school year, but not overall: if student A interacted with B and B interacted with C, A may not have interacted with C.

In Figure 2.1, students 1 to 8 are assigned to classes X and Y in the first year, and classes X' and Y' in the second year. In case 1, classes X' and Y' are identical to classes X and Y, but they are not in case 2. If we want to provide some information on the variety of social environments students have experienced, the clusters only define the students' social environments *at one point in time*. But belonging to different clusters over time is an indication of social environment mobility. Just like income mobility indices, SEM indices provide more additional information on the degree of segregation over time. Although segregation indices have the same values in case 1 and in case 2, case 2 leads to less segregation over time: for instance, student 4 was the only lower class student in her class two years in a row in case 1 but not in case 2.

Therefore, we need to define measures of segregation that do not rely on clusters. Fortunately, the indices introduced in the previous section can be written as a function of the individual exposures to the reference group μ_i. As we introduce a time variable $t \in \{1, \ldots, T\}$, we will now denote μ_{it} the social environment of individual i at time t (which is still computed using the cluster individual i belongs to at time t). For now, we assume that the population is stable over time and that the individual characteristics that define the reference

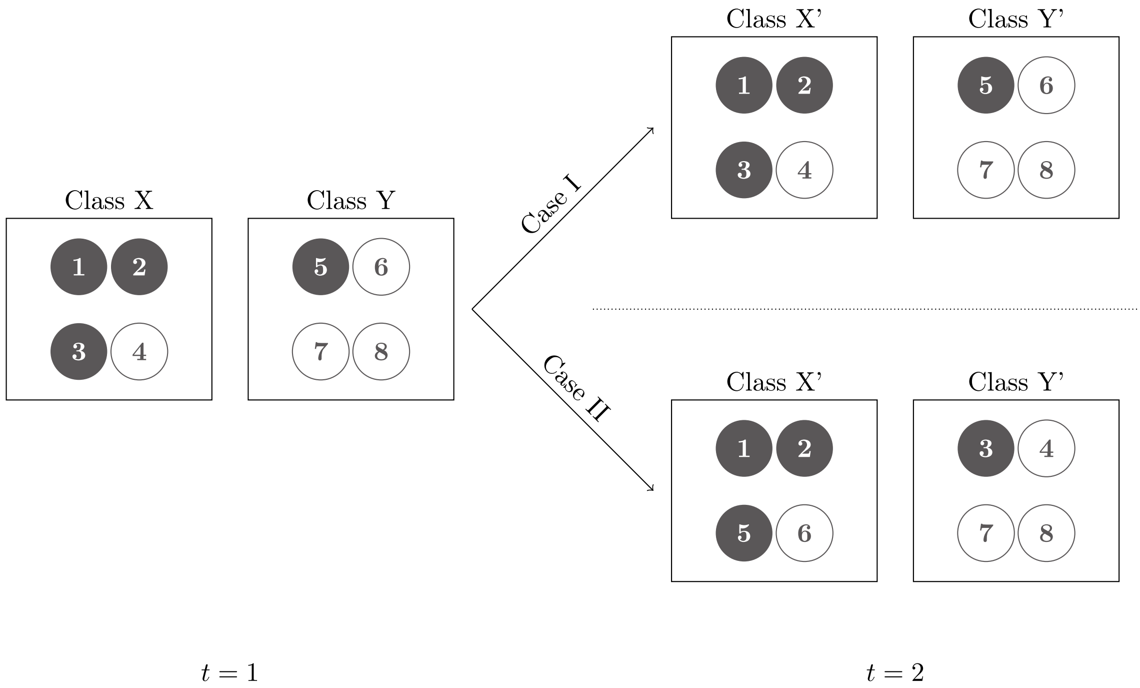

Figure 2.1: Social environment mobility in the case of school segregation: students 1, 2, 3 and 5 are the group A (upper class students) and students 4, 6, 7 and 8 are the group B (lower class students). In the first case, the two classes are identical in both time periods while they are reshuffled in the second case, leading to a higher SEM.

group are stable as well: therefore, N *and* p *are constants.*

$$\begin{aligned} D_t &= \frac{1}{2p(1-p)} \frac{1}{N} \sum_{i=1}^{N} |\mu_{it} - p| \\ G_t &= \frac{1}{p(1-p)} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'=1}^{N} |\mu_{it} - \mu_{i't}| \\ P_t &= \frac{1}{p(1-p)} \frac{1}{N} \sum_{i=1}^{N} (\mu_{it} - p)^2 \\ H_t &= \frac{1}{h(p)} \frac{1}{N} \sum_{i=1}^{N} h(\mu_{it}) \end{aligned}$$

With these expressions, segregation indices are a measure of inequality on the distribution of the social environments μ_{it}. In section 1.1, we introduced income inequality indices that were functions of the Lorenz curve. Yet the literature on segregation has not been using the Lorenz curve as such, but rather an adaptation of it introduced by Duncan & Duncan (1955) and called the *segregation curve*. However, the definition of this curve requires that individuals interact in clusters. The segregation curves is a continuous, piecewise linear function (*i.e.* a spline function of degree one) which goes through the points $(X_k, Y_k)_{k \in \{0,\ldots,K\}}$ where $(X_0, Y_0) = (0, 0)$; X_k is the share of the group A that belongs to the k first clusters with the smallest share of individuals from that group; Y_k is the share of the group B that belongs to those same k clusters. It follows that $(X_K, Y_K) = (1, 1)$.

The segregation curve has interesting properties, in particular it has geometric relations to some segregation indices, notably the dissimilarity index and the Gini index. The dissimilarity index is the maximal vertical distance between the curve and the 45-degree line (whose highest possible value is one). The Gini index is the area between the curve and the 45-degree line, divided by its highest possible value 1/2: this is the same relation as the Gini index and the Lorenz curve in the case of income inequality.

Since we cannot consider clusters when studying SEM, we choose to go back to the initial Lorenz curve. To that end, we introduce the cumulative distribution function of μ_{it}, denoted $\mathcal{F}_t$. Our four segregation indices can be written in integral form using $\mathcal{F}_t$:

$$D_t = \frac{1}{2p(1-p)} \int_0^1 \left|\mathcal{F}_t^{-1}(u) - p\right| \,\mathrm{d}u \tag{17}$$

$$G_t = \frac{1}{2p(1-p)} \int_0^1 \int_0^1 \left|\mathcal{F}_t^{-1}(u) - \mathcal{F}_t^{-1}(u')\right| \,\mathrm{d}u' \,\mathrm{d}u \tag{18}$$

$$P_t = \frac{1}{p(1-p)} \int_0^1 \left(\mathcal{F}_t^{-1}(u) - p\right)^2 \,\mathrm{d}u \tag{19}$$

$$H_t = \frac{1}{h(p)} \int_0^1 h\left(\mathcal{F}_t^{-1}(u)\right) \,\mathrm{d}u \tag{20}$$

The Lorenz curve for social environments at time t, denoted $\mathcal{L}_t$ has the same definition as the Lorenz curve for the income distribution. When there is no segregation, the Lorenz curve coincides with the 45-degree line. However, when segregation is maximal, $\mathcal{L}_t(u) = 0$ only if $u \leq 1-p$. Recall that when all the income is earned by only one individual, $\mathcal{L}_t(u) = 0$ if $u < 1$. In the case of maximal segregation, the group A (which represents a share p of the population) is exposed to itself and to itself only, therefore they have $\mu_{it} = 1$, while the group B (whose share is $1-p$) has $\mu_{it} = 0$. As a consequence, the Lorenz curve for social environments has a lower envelope which is equal to zero on $[0, 1-p]$ and is linear on $[1-p, 1]$, with $\mathcal{L}_t(1) = 1$, as shown on Figure 2.2 (blue curve).

With this limitation, the highest possible vertical distance between the Lorenz curve and the 45-degree line is $1-p$ and the highest possible area is $(1-p)/2$. Interestingly, the dissimilarity index defined above is equal to the maximum vertical distance between the Lorenz curve and the 45-degree line divided by the highest possible value $1-p$. Similarly, the Gini index is equal to the area between the Lorenz curve and the 45-degree line divided by the highest possible value $(1-p)/2$. Therefore, the Lorenz curve remains useful when studying segregation as long as we keep in mind that it has a lower envelope that affects the normalization of the indices.

In order to adapt income mobility measures that were introduced in section 1.2, we also need to introduce

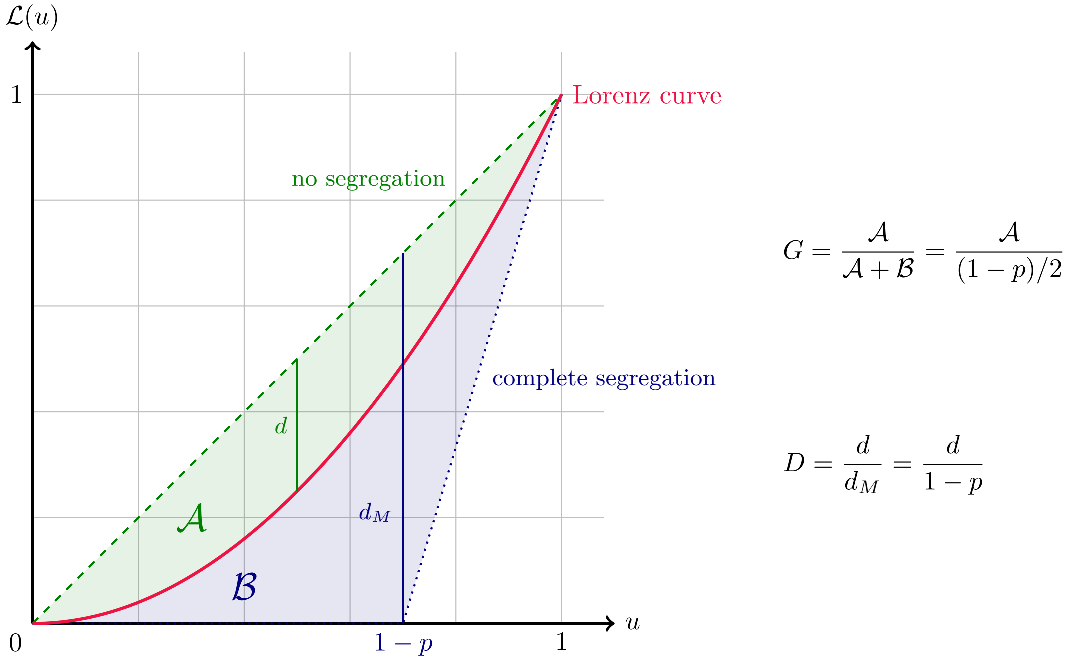

Figure 2.2: The Lorenz curve for social environments

the distributions of *lifetime exposure*, *i.e.* of the variable $\overline{\mu}_i = (1/T)\sum_{t=1}^{T}\mu_{it}$. We call $\mathcal{F}^*$ the corresponding cumulative distribution function and $\mathcal{L}^*$ the Lorenz curve. We can then define the *lifetime segregation indices* S^* (where S is any of the four segregation indices introduced in section 1.3) by replacing the instantaneous distribution $\mathcal{F}_t$ with the lifetime distribution $\mathcal{F}^*$ in the formulae (17), (18), (19) or (20). Note that unlike instantaneous segregation indices, lifetime segregation indices generally do not have simple interpretations, except for the lifetime normalized exposure $\overline{P}$ which is the gap in average exposure to the reference group between the two groups A and B: $\overline{P} = \overline{\mu^A} - \overline{\mu^B}$.

We then have all the tools to adapt the income mobility measures to the case of SEM:

- Following Shorrocks (1978a), we can define SEM indices that compare the lifetime segregation index with the average instantaneous segregation index:

$$\mathcal{M}^S = 1 - \frac{S^*}{\overline{S}} \tag{21}$$

  where $\overline{S}$ is the average instantaneous segregation, *i.e.* the average value of S_t over the T periods. Note that for any segregation index, $S^* \leq \overline{S}$, so that $\mathcal{M}^S$ belongs to the unit interval. This is true for the same reasons why the income mobility index defined in section 1.2.2 belongs to the unit interval, since the mathematical definition is identical (see Shorrocks, 1978a, p. 386).

- Aaberge & Mogstad (2014)'s income mobility curves can also be imported to social environment mobility using the exact same definition:

$$\mathcal{M}(u) = \mathcal{L}^*(u) - \mathcal{L}_R^*(u) \tag{22}$$

  where the Lorenz curve $\mathcal{L}_R^*$ of the "reference distribution" that describes the perfect rigidity scenario has the same definition as in the income mobility case. Both the lifetime segregation Lorenz curve and the reference curve have the same lower envelope, therefore the difference has no lower envelope other than the zero function.

Although the definitions are identical, the SEM indices have one important difference with income mobility indices, which is that even when SEM is very high, lifetime segregation can never vanish to zero as long as there is some degree of instantaneous segregation. We examine this point in more details in the next section.

2.2 The limited equalizing force of social environment mobility

As seen in section 1.2, income mobility gives a different perspective on income inequality, as a significant fraction of income inequality may vanish when comparing lifetime incomes instead of instantaneous incomes. In fact, theoretically, there could exist a society in which instantaneous inequality is always high but income mobility may be so high that inequality in lifetime incomes equals zero. This would be the case, for instance, if the income was a direct function of the individual's age and all individuals have the same lifespan.

Yet in the case of SEM, it is not possible to have no lifetime segregation if we have instantaneous one. As stated in equation (16), the group A will always have a greater exposure to itself than the group B ($\mu_t^A \geq \mu_t^B$), unless there is no segregation in which case the exposures are equal. Suppose that in a given period, there is some segregation: then, for that period, group A will have a greater exposure to itself than group B. If we want lifetime segregation to be zero, *i.e.* if we want all individuals to have the same lifetime exposure to the group A, then in average the group B must have a greater exposure to the group A than the group A itself in other periods: this cannot happen. Therefore, any level of instantaneous segregation has irreversible consequences on the minimum level of lifetime segregation that may be observed after several time periods. An implication is that the Lorenz curve of lifetime exposures $\overline{\mathcal{L}}(u)$ cannot coincide with the 45-degree line: instead, it has an upper envelope $\overline{\mathcal{L}}^*(u)$. Our goal in this section is to determine this upper bound, *considering instantaneous segregation as given.* In other words, we are interested in finding the upper envelope $\overline{\mathcal{L}}^*(u)$ for a given series of the instantaneous distributions $(\mathcal{F}_t)$.

The intuition is the following. As explained in section 1.3, as long as there is some level of segregation at any time, white people and non-white people cannot have the same exposure to white people. The values of μ_t^{A} and μ_t^{B} are determined by the level of instantaneous segregation, which we consider as given. More precisely, using the normalized exposure index P_t, we have the relationships:

$$P_t = 1 - \frac{\mu_t^{\mathrm{B}}}{p} = \mu_t^{\mathrm{A}} - \mu_t^{\mathrm{B}} \tag{23}$$

Since these relationships are linear, we can average over time:

$$\overline{P} = 1 - \frac{\overline{\mu^{\mathrm{B}}}}{p} = \overline{\mu^{\mathrm{A}}} - \overline{\mu^{\mathrm{B}}} \tag{24}$$

Now, in the best case scenario, there is no difference in $\overline{\mu_i}$ *between individuals from the reference group* and *between individuals from the group B*, *i.e.* $\overline{\mu_i} = \overline{\mu^g}$ if $i \in g$ where g denotes group A or B: if SEM is optimal, it will eliminate differences within each group, but it cannot eliminate differences between the two groups.

In that case – this is our main result – the Lorenz curve of lifetime social environments is a spline function of degree one with a knot at $u = 1 - p$ with $\mathcal{L}^*(1-p) = (1-p)\left(1 - \overline{P}\right)$, as plotted in Figure 2.3 (see proof in appendix A.4).

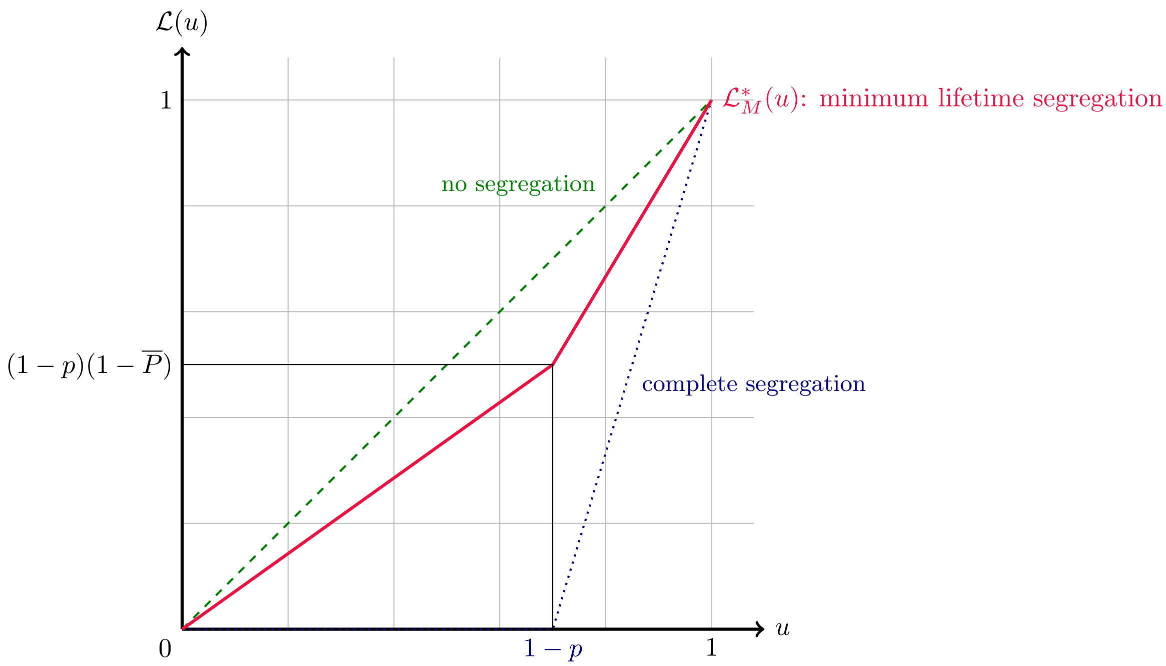

Figure 2.3: The upper envelope of lifetime social environment Lorenz curves

This upper envelope translates into lower bounds for the four lifetime segregation indices:

$$D^*_m = \overline{P} \tag{25}$$

$$G^*_m = \overline{P} \tag{26}$$

$$P^*_m = \overline{P}^2 \tag{27}$$

$$H^*_m = \frac{(1-p)\cdot h\left(p\left(1-\overline{P}\right)\right) + p\cdot h\left(p\left(1-\overline{P}\right)+\overline{P}\right)}{h(p)} \tag{28}$$

and into upper bounds for the SEM indices adapted from Shorrocks (1978a):

$$\mathcal{M}^D_M = 1 - \frac{\overline{P}}{\overline{D}} \tag{29}$$

$$\mathcal{M}^G_M = 1 - \frac{\overline{P}}{\overline{G}} \tag{30}$$

$$\mathcal{M}^P_M = 1 - \overline{P} \tag{31}$$

$$\mathcal{M}^H_M = 1 - \frac{H^*_m}{\overline{H}} \tag{32}$$

Note that this upper bound to SEM indices is a theoretical one. Reaching this upper bound exactly is not always possible if the number of time periods is not big enough. The SEM index is equal to its upper bound if and only if all individuals within each group have the exact same lifetime exposure to the reference group.

Consider the case described in Figure 1.3 and suppose that each year, there are two classes with identical group compositions but possibly different individual compositions (*i.e.* class X always contains three students from group A but they can be any of the four group A students). The value of the segregation index P_t will be 25 percent in each year, therefore $\overline{P}$ is also equal to 25 percent and the upper bound of the SEM index is 75 percent as per equation (31). The average exposure of the upper class students is equal to 62.5 percent in each period: if SEM is complete, the lifetime exposure of each upper class student should be equal to 62.5 percent. This is possible if and only if each student spends three time periods in class X for each time period spent in class Y: therefore, the number of periods must be a multiple of four (there is an identical constraint on lower class students).

Therefore, the number of time periods induces a bias on the upper bound to mobility. This bias decreases with the number of time periods: in our example, complete mobility could not be reached for $T = 10$ or $T = 14$ but the variance of lifetime exposure within *e.g.* the reference group would be smaller in the second case. The bias also gets smaller when the distribution of exposures at each time is peaked around its mean value: in that case, a large share of individuals will be close to the mean value at each time and the lifetime exposure can converge more rapidly towards this mean value.

Unfortunately, there is no simple formula for this bias: finding the optimal series of allocations of individuals across clusters at each time is a complicated algorithm. In practice however, if the allocation of individuals across clusters is not too peculiar, the distribution of exposures allows for a relatively rapid convergence towards a zero bias.

2.3 Dealing with attrition

Attrition is a classic issue when working with panel data. In the case of SEM, attrition entails an additional methodological difficulty. While we need individuals to be observed several times in order to compute a mobility index, individuals who are observed only once still affect the other individuals' social environments. Therefore, it is not possible to simply remove individuals who are not always observed out of the sample: at least, they must be taken into account when computing the exposures.

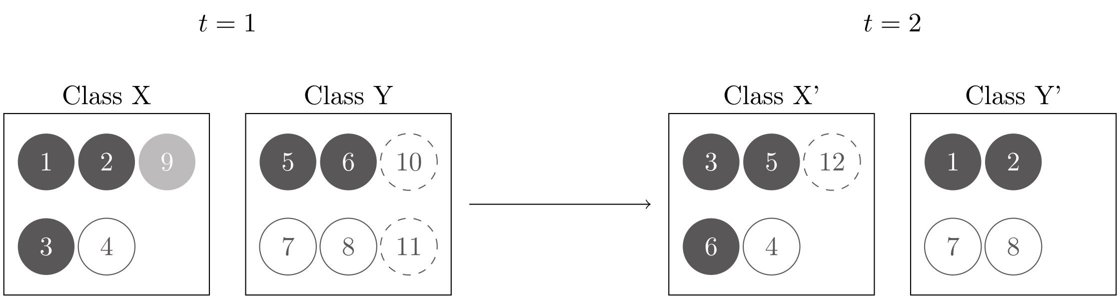

Figure 2.4: Attrition: students 9, 10, 11 and 12 are not present in both time periods but they cannot be ignored since they affect the social environment of students 1 to 8.

Call Ω_t the set of individuals observed in t and define:

$$\Omega = \bigcup_{t=1}^{T} \Omega_t \qquad \text{and} \qquad \omega = \bigcap_{t=1}^{T} \Omega_t$$

Ω is the set of individuals that we observe at least once and $\omega \subset \Omega$ is the *balanced panel*, *i.e.* the subset of individuals that we observe in each time period. The larger sample Ω is used to compute the exposures to the reference group: the social environment of individual i at time t is a subset of Ω (but maybe not of ω), and μ_{it} is the share of individuals from the reference group in that subset. For instance, Figure 2.4 represents a group of 12 students who interact in two classes at two points in time. The restricted sample is $\omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$

and the full sample is $\Omega = \omega + \{9, 10, 11, 12\}$. In the first period, the social environment of student A is the entire class X, which contains 5 students: 4 upper class students and one lower class student. Therefore, the student 1 has exposure $\mu_{1,1} = 4/5 = 0.8$.

We could then compute instantaneous segregation indices based on the distribution of μ_{it} on the restricted sample, using the formulae (17), (18), (19) and (20). However, these definitions may lead to segregation indices out of the unit interval. This also means that the SEM indices based on segregation indices cannot be used directly.

In this section, we focus on the SEM index $\mathcal{M}^P$, which is based on the normalized exposure segregation index and which can be easily adapted to a case with attrition – unlike the other SEM indices that cannot be adapted in a simple way.

As stated earlier, the normalized exposure index P_t is equal to the variance of the series of exposures $(\mu_{it})_{it \in \{1,\dots,N\} \times \{1,\dots,T\}}$, normalized by its highest possible value $p(1-p)$, which would be reached if each class contained either only individuals from group A or only individuals from group B. Similarly, P^* is the variance of the series $(\overline{\mu_i})_{i \in \{1,\dots,N\}}$, normalized by its highest possible value which is also $p(1-p)$. Since the normalization is the same for P_t and P^*, the ratio $P^*/\overline{P}$ is also equal to the ratio of the variances $V^* = p(1-p)P^*$ and $V = p(1-p)\overline{P}$. V^* is the variance of the lifetime exposures $\overline{\mu_i}$ while V is the total variance of μ_{it} (as long as p is constant, which is one of our hypotheses). Then,

$$\mathcal{M}^P = 1 - \frac{V^*}{V} = \frac{V - V^*}{V} \tag{33}$$

This new definition of the mobility index $\mathcal{M}^P$ has an interesting interpretation. Indeed, a variance decomposition shows that $V - V^*$ is equal to the average individual variance of exposure across time periods: if v_i is the variance of the series $(\mu_{it})_t$, $V - V^*$ is the average value of v_i. Therefore, $V - V^*$ captures the diversity of the social environments experienced by individuals. The SEM index $\mathcal{M}^P$ compares the average diversity of social environments experienced by a single individual to the diversity of all existing social environments in the setting. An equivalent approach is to look at the panel series (μ_{it}) and to regress it against an individual fixed effect:

$$\mu_{it} = \alpha_i + \epsilon_{it} \tag{34}$$

$\mathcal{M}^P$ is then equal to $1 - R^2$ where R^2 is the coefficient of determination of the above regression.

The definition can then be adapted to a scenario with attrition. Denoting V_ω^* the variance of the series $(\overline{\mu_i})_{i\in\omega}$ and V_ω the variance of the series $(\mu_{it})_{it\in\omega\times\{1,\dots,T\}}$, we can use this simple adaptation of the SEM index:

$$\mathcal{M}_\omega^P = 1 - \frac{V_\omega^*}{V_\omega} = \frac{V_\omega - V_\omega^*}{V_\omega} \tag{35}$$

i.e. we compare the average individual variance to the overall variance of the exposures *in the reference group*. The interpretation given above is still valid and $\mathcal{M}^P$ still belongs to the unit interval.

The formula of the upper bound defined in section 2.2 should also be adapted to this new definition. In order to compute this upper bound, we consider once again the scenario of optimal SEM. In this best case scenario, all individuals from the reference group *in the balanced panel* have the same lifetime exposure, and all individuals from the group B *in the balanced panel* has the same lifetime exposure too. These two values $\overline{\mu_\omega^{\mathrm{A}}}$ and $\overline{\mu_\omega^{\mathrm{B}}}$ can be computed directly from the actual distribution of μ_{it}:

$$\overline{\mu_\omega^g} = \mathbb{E}(\mu_{it} | i \in \omega \cap g) \tag{36}$$

Denoting p_ω the share of the reference group in the balanced panel and $\overline{\mu_\omega}$ the average exposure to the reference group in this sample,[6] the smallest possible variance of lifetime exposures is given by

$$V_{\omega,m}^* = p_\omega(\overline{\mu_\omega^{\mathrm{A}}} - \overline{\mu_\omega})^2 + (1-p_\omega)(\overline{\mu_\omega^{\mathrm{B}}} - \overline{\mu_\omega})^2 \tag{37}$$

Therefore, the upper bound of the SEM index is:

$$\mathcal{M}_{\omega,M}^P = 1 - \frac{p_\omega(\overline{\mu_\omega^{\mathrm{A}}} - \overline{\mu_\omega})^2 + (1-p_\omega)(\overline{\mu_\omega^{\mathrm{B}}} - \overline{\mu_\omega})^2}{V_\omega} \tag{38}$$

2.4 Nested levels of clustering

When studying segregation, it is not uncommon to find cases where there are nested levels of clustering. The typical example is school segregation, where individuals (students) interact mainly within classes but also within schools that contain several classes. The value of the segregation index varies depending on the choice of the level of clustering, *i.e.* on the definition of a student's social environment. In particular, if we take the classroom

[6] $\overline{\mu_\omega}$ is different from p_ω because the individuals from the balanced panel interact with individuals from Ω who are not in ω. If there was no attrition ($\omega = \Omega$), then we would have $\overline{\mu_\omega} = p_\omega$.

as the reference cluster, the segregation index will be larger than if the reference cluster is the whole school. In the case of the normalized exposure index P, there is a simple relationship between the indices P^{class} and P^{school}.

Call μ_i^{class} (resp. μ_i^{school}) the share of the reference group in student i's class (resp. school). Recall that U_i is a binary variable that equals 1 if the individual i belongs to the reference group. As stated earlier, the segregation index P^{class} is the share of the variance of U_i that is explained by classes, which is equal to the variance V^{class} of the series $(\mu_i^{\text{class}})_i$ divided by the variance of U_i, *i.e.* by $p(1-p)$. Similarly, the segregation index P^{school} is equal to the variance V^{school} divided by $p(1-p)$. Note that the series $(\mu_i^{\text{class}})_i$ (resp. $(\mu_i^{\text{school}})_i$) is identical to the series of the shares of the reference group in each class (resp. school), weighting each class (resp. school) by its size.

Therefore, the total variance $p(1-p)$ is the sum of three terms:

1. the variance of μ_i^{school}, which represents the *between-school variance*;
2. the average *within-school variance* of μ_i^{class};
3. the average within-class of the variable U_i, which is the remaining variation that is not explained by schools or classes.

The share of the first term in the total variance is the between-school segregation P^{school}. The share of the first two terms is the between-class segregation P^{class}. This reasoning can be extended to more clustering levels: city, region, etc.

Multi-level clustering is also interesting in the case of SEM. In section 2.1, we explained how the segregation indices had to be reworked in order to be independent of the notion of clusters, since SEM implies that individuals do not remain in the same cluster throughout their life. However, when there are several levels of clustering, we can imagine that the smallest-level clusters are indeed unstable over time, but that larger clusters do remain throughout all time periods. In the case of school segregation, we could imagine that classes are reshuffled every year but that students cannot change schools. In this scenario, the SEM index has to be adapted.

As detailed in section 2.3, the SEM index $\mathcal{M}^P$ can be seen as the ratio of the average individual variance of exposure to the variance of all exposures in the population and throughout all time periods. Yet if students cannot leave their school, it makes sense to compare the diversity of exposures experienced by an individual student to the distribution of exposures that are available in their school rather than in any existing school. There are two ways in which this can be taken into account in the SEM indices:

1. If students cannot change schools, the upper bound to mobility is actually lower than the standard value. The best case scenario under this constraint is that *within each school, every individual from the reference group has the same lifetime exposure and every individual from the other group has the same lifetime exposure*. The value of the upper bound of the SEM index can easily be computed under this condition.
2. It may also make sense to change the denominator of the SEM index given in equation (33) in order to compare the individual variances to the within-school variance rather than the total variance V. Note that in this case, the upper bound should also be computed using the method described in the previous point (optimal within-school mobility).

In practice however, clusters are rarely perfectly persistent over time: although most students remain in the same school, a small fraction of them is likely to change schools. The existence of such students has important consequences on the within-school SEM index's axiomatic properties: in theory, if some students may change schools, within-school mobility may exceed one. If these students are a small minority, we choose to treat them as if they left the balanced panel and we consider them as attrition: in the notations of section 2.3, they belong to $\Omega \setminus \omega$, *i.e.* they do affect the other students' environments, but they are not taken into account when computing the SEM index.

Overall, if we consider two levels of clustering, there are three SEM indices to consider:

1. the standard SEM index where we compare the average individual variance to the overall variance in exposures;
2. the same standard SEM index but restricted to a balanced panel consisting of individuals who remain in the same school;
3. the SEM index obtained by dividing the average individual variance of the restricted balanced panel by the overall within-school variance.

The difference between the values of the first and the second index should be small and is only due to a selection effect. However, they have different upper bounds, since in the second case the relevant upper bound is based on the assumption that students cannot change schools. The third index will be mechanically higher than the second, since the only difference is to replace the total variance by the within-school variance in the denominator, in both the SEM index and the upper bound.

3 Case study: SEM in French middle schools

In this section, we apply the SEM analysis methodology that we laid out in section 2 to a concrete example. All segregation and SEM indices rely on the normalized exposure index P, which has several advantages:

- the lifetime segregation index P^* has an intuitive interpretation, unlike other lifetime segregation indices;
- the upper bound to the SEM index can be expressed as a function of the average instantaneous segregation $\overline{P}$ when there is no attrition;
- when there is attrition, we can use adapted versions of the SEM indices based on $\mathcal{M}^P$;
- we can use the additive decomposability property of the segregation index and compare SEM indices on different levels.

We take the example of social segregation and SEM in French middle schools. In 2009, a cohort of 797,894 students entered 6th grade. These students spend four years in middle school, from grade 6 to grade 9. The objective of this section is to provide figures on social segregation in French middle schools and on the diversity of social environments experienced by the middle school students during these four years.

Thanks to a unique database from the French Ministry of Education, we are able to follow 79 percent of the cohort throughout the four years of middle school (possibly in different schools at different points in time): these students are the balanced panel. We know about their socioeconomic status and the class they attended in each year. The 21 percent of students who are not observed all four years either repeated a grade (7 percent) or could not be matched from one dataset to another (14 percent).

Our reference group is the upper class students, which make up 38 percent of the sample ($p = 0.38$); while the rest of the sample contains both lower class and middle class students, we refer to them as the lower class students for the sake of simplicity.

Figure 3.1 gives the value of the instantaneous segregation index at each grade $t \in \{6, 7, 8, 9\}$. In average, the segregation index equals 19 percent. Given that the share of upper class students is 38 percent, this means

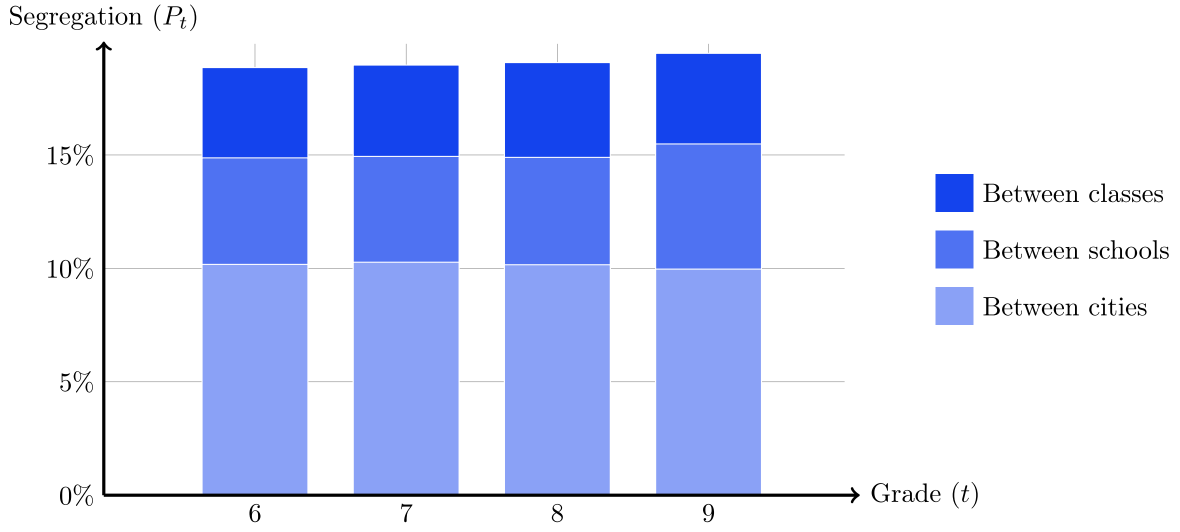

Figure 3.1: Normalized exposure in French middle schools, cohort 2009

that in average, *the class of an upper class student contains 50 percent of upper class students, while the class of a lower class student contains only 31 percent of upper class students*, as per equation (15). The figures also show the decomposition of the segregation across three levels of clustering: cities, schools and classes.

The between-cities segregation is equal to 10 percent, which means that if all schools and all classes were identical within each city, the total segregation would equal 10 percent instead of 19 percent: residential segregation accounts for 53 percent of the total segregation. The other half is driven by within-city, between-school segregation (5 points, 26 percent of the total) and within-school, between-class segregation (4 points, 21 percent of the total).

Because of the attrition in our sample, we need to use the SEM index defined in equation (35) and the upper bound defined in equation (38). We find that the value of the SEM index $\mathcal{M}_\omega^P$ is 15 percent and its upper bound is 81 percent. The numerical value of the upper bound is very close to $1 - \overline{P}$ which is the value in the case where there is no attrition.

A simple, yet approximate, interpretation of the SEM index is the following. For each student in the balanced panel, compute the average share of upper class students in their class throughout the four years of middle school. The gap between an upper class and a lower class student's lifetime exposure to upper class students, P^*, is about 15 percent less than the segregation index, *i.e.* 16 percent instead of 19 percent. This interpretation is not strictly correct since the numerator in equation (35) does not relate exactly to the exposure gap $\overline{\mu_\omega^1} - \overline{\mu_\omega^0}$ (it does only if there is no attrition). In this case however, this interpretation remains fairly accurate – which can be checked by computing the actual lifetime exposures.

A second interpretation is that the individual variance of exposures to upper class students is about 15 percent of the variance of exposure throughout all classes in middle school. In other words, each student only experiences about 15 percent of the diversity of social environments available in all French middle schools. Although the average student could not experience all this diversity because of the existing segregation at each time, the actual value is quite far from the upper bound (81 percent).

Among the 630,466 from the balanced panel, 545,250 (86 percent) remained in the same school all four years. This indicates that mobility across schools is very limited. Therefore, it makes sense to measure the within-school SEM index. To that end, we focus on the restricted balanced panel of the 545,250 students who remained in the same school all four years. The SEM index is slightly lower for this sample at 12 percent instead of 15 percent. If SEM within each school was optimal, *i.e.* if all upper class students and all lower class students within each school had the same lifetime exposure, this SEM index would reach 20 percent: as long as students do not change schools, the social segregation is such that middle school principals could not get their students to experience more than 20 percent of the variety of the social environments in the country. If we compare the individual variances to the within-school variance instead of the total variance, we obtain the within-school SEM index, which equals 58 percent, with an upper bound of 93 percent. Overall, within-school mobility is quite substantial: this indicates that a large share of students experiences a fair share of the social environments that exist in the school throughout their four years.

Future work will provide a more in-depth analysis of segregation and social environment mobility, which this case study is a preview of. We also plan to analyze social environment mobility in different contexts (*e.g.* residential mobility in France or in other countries).

References

Aaberge, Rolf. 2001. Axiomatic Characterization of the Gini Coefficient and Lorenz Curve Orderings. *Journal of Economic Theory*, **101**(1), 115–132.

Aaberge, Rolf, & Mogstad, Magne. 2014. *Income Mobility as an Equalizer of Permanent Income*. Discussion Paper 769. Statistics Norway Research department.

Bonhomme, Stéphane, & Robin, Jean-Marc. 2009. Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990–2000. *The Review of Economic Studies*, **76**(1), 63–92.

Bowlus, Audra J., & Robin, Jean-Marc. 2004. Twenty Years of Rising Inequality in U.S. Lifetime Labour Income Values. *The Review of Economic Studies*, **71**(3), 709–742.

Braddock II, Jomills Henry, & McPartland, James M. 1989. Social-Psychological Processes That Perpetuate Racial Segregation: The Relationship Between School and Employment Desegregation. *The Journal of Black Studies*, **19**(3), 267–289.

Camargo, Braz, Stinebrickner, Ralph, & Stinebrickner, Todd. 2010. Interracial Friendships in College. *Journal of Labor Economics*, **28**(4), 861–892.

Crain, Robert L., & Strauss, Jack. 1985. *School Desegregation and Black Occupational Attainments: Results from a Long-Term Experiment*. Tech. rept. 359. Center for the Social Organization of Schools, Baltimore.

Currarini, Sergio, Jackson, Matthew O., & Pin, Paolo. 2009. An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, **77**(4), 1003–1045.

Cutler, David M., & Glaeser, Edward L. 1997. Are Ghettos Good or Bad? *The Quarterly Journal of Economics*, **112**(3), 827–872.

Duncan, Otis Dudley, & Duncan, Beverly. 1955. A Methodological Analysis of Segregation Indexes. *American Sociological Review*, **20**(2), 210–217.

Echenique, Federico, & Fryer, Roland G. 2007. A Measure of Segregation Based on Social Interactions. *The Quarterly Journal of Economics*, **122**(2), 441–485.

Frankel, David M., & Volij, Oscar. 2011. Measuring School Segregation. *Journal of Economic Theory*, **146**(1), 1–38.

Gottschalk, Peter, & Spolaore, Enrico. 2002. On the Evaluation of Economic Mobility. *The Review of Economic Studies*, **69**(1), 191–208.

Lorenz, Max Otto. 1905. Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association*, **9**(70), 209–219.

Marmaros, David, & Sacerdote, Bruce. 2006. How Do Friendships Form? *The Quarterly Journal of Economics*, **121**(1), 79–119.

Massey, Douglas S., & Denton, Nancy A. 1988. The Dimensions of Residential Segregation. *Social Forces*, **67**(2), 281–315.

Moffitt, Robert A., & Gottschalk, Peter. 2002. Trends in the Transitory Variance of Earnings in the United States. *Economic Journal*, **112**(478), C68–C73.

Prais, S. J. 1955. Measuring Social Mobility. *Journal of the Royal Statistical Society. Series A (General)*, **118**(1), 56–66.

Shorrocks, Anthony. 1978a. Income Inequality and Income Mobility. *Journal of Economic Theory*, **19**(2), 376–393.

Shorrocks, Anthony. 1978b. The Measurement of Mobility. *Econometrica*, **46**(5), 1013–1024.

Wells, Amy Stuart, & Crain, Robert L. 1994. Perpetuation Theory and the Long-Term Effects of School Desegregation. *Review of Educational Research*, **64**(4), 531–555.

Yaari, Menahem E. 1988. A controversial proposal concerning inequality measurement. *Journal of Economic Theory*, **44**(2), 381–397.

A Mathematical proofs

A.1 Equivalence of the two definitions of the Gini index

We prove here that the definitions of the Gini index provided in equations (2) and (3) are equivalent, *i.e.*:

$$G = 1 - 2\int_0^1 \mathcal{L}(u)\,\mathrm{d}u = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|y_i - y_j| \tag{39}$$

Without losing generality, we can assume that $i < j$ implies $y_i < y_j$, *i.e.* that the incomes are sorted in ascending order. Under that assumption, the double sum from the right hand side can be rewritten:

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N}|y_i - y_j| &= \sum_{y_j \geq y_i}(y_j - y_i) + \sum_{y_j < y_i}(y_i - y_j) \\
&= 2\sum_{y_j \geq y_i}(y_j - y_i) \\
&= 2\sum_{i=1}^{N}\sum_{j=i}^{N}(y_j - y_i) \\
\frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|y_i - y_j| &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i}^{N}(y_j - y_i)
\end{aligned}$$

We now transform this double sum into a double integral, by noting that if $x = i/N$, $\mathcal{F}^{-1}(x) = y_i$:

$$\begin{aligned}
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i}^{N}(y_j - y_i) &= \int_0^1\int_x^1 \left(\mathcal{F}^{-1}(t) - \mathcal{F}^{-1}(x)\right)\,\mathrm{d}t\,\mathrm{d}x \\
&= \int_0^1\left[\int_x^1 \left(\mathcal{F}^{-1}(t) - \mathcal{F}^{-1}(x)\right)\,\mathrm{d}t\right]\,\mathrm{d}x \\
&= \int_0^1\left[\int_x^1 \mathcal{F}^{-1}(t)\,\mathrm{d}t - (1-x)\mathcal{F}^{-1}(x)\right]\,\mathrm{d}x
\end{aligned}$$

Note that

$$\int_0^1 \mathcal{F}^{-1}(x)\,\mathrm{d}x = \overline{y}$$

and

$$\int_x^1 \mathcal{F}^{-1}(t)\,\mathrm{d}t = \int_0^1 \mathcal{F}^{-1}(t)\,\mathrm{d}t - \int_0^x \mathcal{F}^{-1}(t)\,\mathrm{d}t = \overline{y} - \mathcal{L}(x)$$

Therefore

$$
\begin{aligned}
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i}^{N}(y_j - y_i) &= \int_0^1 \left[\int_x^1 \mathcal{F}^{-1}(t)\,\mathrm{d}t - (1-x)\mathcal{F}^{-1}(x)\right] \mathrm{d}x \\
&= \int_0^1 \left[\overline{y} - \mathcal{L}(x) - \mathcal{F}^{-1}(x) + x\mathcal{F}^{-1}(x)\right] \mathrm{d}x \\
&= \int_0^1 \left[x\mathcal{F}^{-1}(x) - \mathcal{L}(x)\right] \mathrm{d}x
\end{aligned}
$$

We integrate the first term under the integral by parts, noting that a primitive of $\mathcal{F}^{-1}$ is $\overline{y}\mathcal{L}$ according to definition (1):

$$
\begin{aligned}
\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i}^{N}(y_j - y_i) &= \int_0^1 \left[x\mathcal{F}^{-1}(x) - \mathcal{L}(x)\right] \mathrm{d}x \\
&= \int_0^1 x\mathcal{F}^{-1}(x)\,\mathrm{d}x - \int_0^1 \mathcal{L}(x)\,\mathrm{d}x \\
&= \left[x \cdot \overline{y}\mathcal{L}(x)\right]_0^1 - 2\int_0^1 \mathcal{L}(x)\,\mathrm{d}x \\
&= 1 - 2\int_0^1 \mathcal{L}(x)\,\mathrm{d}x
\end{aligned}
$$

which proves equality (39).

A.2 Normalization of the exposure index

The normalized exposure index is defined by

$$
P = \frac{1}{p(1-p)}\sum_{k=1}^{K}\frac{N_k}{N}(p_k - p)^2 \tag{40}
$$

We are interested in the extreme values of

$$
f(\mathbf{p}) = \sum_{k=1}^{K} N_k(p_k - p)^2 \tag{41}
$$

subject to

$$
\sum_{k=1}^{K} N_k p_k = p \tag{42}
$$

where $\mathbf{p} = (p_k)_{k\in\{1,\dots,K\}} \in [0;1]^K$.

We introduce the Lagrangian:

$$\mathcal{L}(\mathbf{p},\lambda) = \sum_{k=1}^{K} N_k(p_k - p)^2 - \lambda\left(\sum_{k=1}^{K} N_k p_k - p\right) \tag{43}$$

and we differentiate with respect to p_k:

$$\frac{\partial \mathcal{L}}{\partial p_k}(\mathbf{p},\lambda) = 2N_k(p_k - p) - \lambda N_k \tag{44}$$

The derivative equals 0 if and only if $p_k = p + \lambda/2$. Plugging this equality in the constraint (42) shows that λ must equal zero, therefore $p_k \equiv p$ for all k. This case corresponds to the minimal value of the exposure index ($P = 0$). Since the Lagrangian has no other extremum, the maximum values are reached on the corners of the domain $[0;1]^K$, *i.e.* if $p_k \in \{0;1\}$ for all k. In that case, $p_k^2 = p_k$, therefore with the constraint (42):

$$f(\mathbf{p}) = \sum_{k=1}^{K} N_k(p_k - p)^2 = \sum_{k=1}^{K} N_k(p_k - 2\cdot p\cdot p_k + p^2) = N\cdot p - 2\cdot N\cdot p^2 + N\cdot p^2 = N\cdot p(1-p) \tag{45}$$

Hence the normalization in equation (40).

A.3 Normalized exposure is equal to the exposure gap

The normalized exposure index is defined by

$$P = \frac{1}{p(1-p)}\sum_{k=1}^{K}\frac{N_k}{N}(p_k - p)^2$$

We first want to show that

$$P = 1 - \frac{\mu^B}{p}$$

where μ^B is the average exposure of group B to group A.

Denote $N^B = N(1-p)$ the number of individuals in group B and $N_k^B = N_k(1-p_k)$ the number of individuals from group B in cluster k. Then, the definition of μ_B is:

$$\mu^B = \frac{1}{N^B}\sum_{k=1}^{K} N_k^B p_k = \frac{1}{N(1-p)}\sum_{k=1}^{K} N_k(1-p_k)p_k = \sum_{k=1}^{K}\frac{N_k}{N}\cdot\frac{p_k(1-p_k)}{1-p}$$

Multiply the first two equations by $p(1-p)$:

$$\begin{aligned} p(1-p)P &= \sum_{k=1}^{K}\frac{N_k}{N}(p_k-p)^2 \\ p(1-p)P &= p(1-p)-(1-p)\mu^B = p(1-p)-\sum_{k=1}^{K}\frac{N_k}{N}p_k(1-p_k) \end{aligned}$$

We need to show that

$$\sum_{k=1}^{K}\frac{N_k}{N}(p_k-p)^2 = p(1-p)-\sum_{k=1}^{K}\frac{N_k}{N}p_k(1-p_k)$$

i.e.

$$p(1-p) = \sum_{k=1}^{K}\frac{N_k}{N}(p_k-p)^2+\sum_{k=1}^{K}\frac{N_k}{N}p_k(1-p_k)$$

We start from the right hand side of this last equality:

$$\begin{aligned} \sum_{k=1}^{K}\frac{N_k}{N}(p_k-p)^2+\sum_{k=1}^{K}\frac{N_k}{N}p_k(1-p_k) &= \sum_{k=1}^{K}\frac{N_k}{N}\left[(p_k-p)^2+p_k(1-p_k)\right] \\ &= \sum_{k=1}^{K}\frac{N_k}{N}\left[p_k^2-2p\cdot p_k+p^2+p_k-p_k^2\right] \\ &= p^2+(1-2p)\sum_{k=1}^{K}\frac{N_k}{N}p_k \\ &= p^2+(1-2p)\cdot p \\ &= p(1-p) \end{aligned}$$

because the sum of N_k/N is one and the sum of $(N_k/N)p_k$ is the weighted average of p_k, which is equal to p. Hence the intermediary result:

$$P = 1-\frac{\mu^B}{p}$$

Introduce μ^A, the average exposure of the reference group to itself. The average exposure of the population to the reference group is equal to p. It is also equal to the weighted average of the exposures of groups A and

B, *i.e.*:

$$p = p \cdot \mu^A + (1-p) \cdot \mu^B$$

Therefore, if $p \neq 0$

$$\mu^A = 1 - \frac{1-p}{p}\mu^B$$

which yields the final result:

$$\mu^A - \mu^B = 1 - \frac{1-p}{p}\mu^B - \mu^B = 1 - \frac{\mu^B}{p} = P$$

A.4 Upper envelope of the Lorenz curve of lifetime social environments

As stated in section 2.2, "in the best case scenario, there is no difference in $\overline{\mu_i}$ *between individuals from the reference group* and *between individuals from the group B*, *i.e.* $\overline{\mu_i} = \overline{\mu^g}$ if $i \in g$ where g denotes group A or B: if SEM is optimal, it will eliminate differences within each group, but it cannot eliminate differences between the two groups."

The values of the exposures in each group can be derived from equation (24):

$$\begin{aligned} \overline{\mu^A} &= p(1-\overline{P}) + \overline{P} \\ \overline{\mu^B} &= p(1-\overline{P}) \end{aligned}$$

Since group B has the lowest average exposure to the reference group and it represents a share $1-p$ of the population, the inverse of the cumulative distribution function is given by

$$\mathcal{F}^{*-1}(u) = \begin{cases} \mu^B = p(1-\overline{P}) & \text{if} \quad u < 1-p \\ \mu^A = p(1-\overline{P}) + \overline{P} & \text{otherwise} \end{cases} \tag{46}$$

The cdf is piecewise constant, therefore the Lorenz curve, which is defined by

$$\mathcal{L}^*(u) = \frac{1}{p}\int_0^u \mathcal{F}^{*-1}(t)\,\mathrm{d}t \tag{47}$$

is continuous and piecewise linear: it is a spline of degree one, with a knot at $u = 1 - p$ (at the level of the discontinuity in $\mathcal{F}^{*-1}$). We always have $\mathcal{L}^*(0) = 0$ and $\mathcal{L}^*(1) = 1$, therefore we only need to determine $\mathcal{L}^*(1-p)$ in order to draw the curve in Figure 2.3:

$$\mathcal{L}^*(1-p) = \frac{1}{p} \int_0^{1-p} p(1-\overline{P})\,\mathrm{d}t = (1-p)(1-\overline{P})$$