A Method for Studying Difference in Segregation Levels Across Time and Space

Benjamin Elbers*
Department of Sociology
Columbia University

This version: September 9, 2018

Abstract

An important topic in the study of segregation are comparisons across space and time. Theil's information index H is frequently used to study segregation. In its interpretation, H is sometimes treated a "margin-free" measure of segregation, which implies that the H index is not sensitive to marginal changes in the size of groups (for instance, racial groups) or organizational units (for instance, schools). This conclusion is only partially true, which complicates the understanding of differences in segregation levels across time and space. Regarding this issue, the paper makes three contributions. *First*, in line with arguments presented by Mora and Ruiz-Castillo (2009; 2011), it is shown that the closely related M index has some conceptual advantages over the H index. Additionally, the relationship between the M and H indices is further clarified. *Second*, by combining a method first introduced by Karmel and Maclachlan (1988) with the advantages of the M index, it is shown that a decomposition of changes in the M index into several components is possible: one component captures changes that are introduced due to the changing marginal distributions, and one component captures changes that are due to structural increases or decreases in segregation. Both of these can be further decomposed to study the precise sources of changing segregation. *Third*, the decomposition is further refined by taking into account the appearance or disappearance of new units and groups, and by distinguishing comparisons across time from those of across space. The paper concludes with an example.

The research in this article has been supported by National Science Foundation grant SES-1423828.

*be2239@columbia.edu

1 Introduction

Studies of segregation are concerned with a variety of substantive problems. Social scientists are interested in residential racial segregation, in the racial or class-based segregation of schools and workplaces, or the gender segregation of occupations. More generally, any study of the association between two categorical variables can be regarded as a segregation problem. Segregation is usually studied by applying a segregation index to a contingency table, which provides a one-number summary of the association between, for instance, gender and occupations.

Often, the interest in the study of a segregation problem lies not only in describing segregation at one point in time or in one place, but in comparing levels of segregation over time, across countries or cities, or between population groups. The contribution of this paper is to provide a general and practical method for the study of change in segregation across time, space, or population groups. Many of the ideas presented in this paper are discussed in the methodological literature on segregation, but are rarely or only partially considered in empirical studies of segregation. Given that the interest of many studies lies not so much in absolute levels of segregation but in the *changes* of segregation over time, a general debate about the comparability of segregation indices over time or across space seems necessary.

The main issue of comparability in many empirical studies is the margin-dependency of many segregation indices, which has been a long-standing issue in the occupational sex segregation literature. This literature is concerned with the extent to which men and women cluster in different occupations. Intuitively, occupational segregation is maximized when men and women work in completely different sets of occupations. On the opposite end, an intuitive definition of integration would be the notion that all occupations have the same gender distribution, which would then also be the overall gender distribution of the labor force. The empirical realities, of course, fall in between these two extremes, and this is where the segregation indices differ.

The issue of margin dependency implies that the level of segregation will generally change when the distribution of workers over occupations or female labor force participation increases. A margin-free index, by definition, does not change under either of these processes. This seems to call for margin-free measurement, but this is not an easy problem to solve. One of the best known approaches is the one by Charles and Grusky (1995), based on log-multiplicative modeling. Essentially, the index calculates the odds ratio of male and female employment within each occupation, and is then summarized by weighting all occupation-specific ratios equally. Thus, the index measures only the level of association, and is not influenced by changes in the marginal distribution of either occupations or genders. However, the index also has a number of disadvantages, namely issues of interpretation and decomposability. Especially if the sizes of occupations differ wildly, the index is problematic.[1] It thus seems even less applicable when school or residential segregation is studied.

Among the margin-dependent indices, Theil's Entropy Index H has become increasingly popular, although the much more problematic index of dissimilarity D is also still being used. The H has a number of attractive properties, some of which are presented below, and which make the H often an attractive measure of segregation. The margin dependency of the H index is often not explicitly considered in empirical studies, although this fact is known at

[1] See the exchange between Watts (1998) and Grusky and Charles (1998).

least since James and Taeuber (1985). It is also often not appreciated that the H is margin-dependent in both dimensions, i.e., both changes in the distributions of workers over genders and over occupations may influence H, even if there is no change in their association.

This paper argues that both margin-free and margin-dependent measurement are useful for the study of segregation. It is especially useful to consider a margin-dependent index as measuring the level of segregation that the average person in the population under study experiences, which makes it a natural measure in the cross section. Furthermore, entropy-based segregation indices such as the H and the M (introduced below) fulfill a variety of useful and desirable properties that allow a rigorous measurement of segregation (Reardon and Firebaugh, 2002; Frankel and Volij, 2011). However, when comparing segregation levels over time or across places, it is desirable to decompose changes that are due to changes in the marginal distribution from changes in the intrinsic association structure. To illustrate this point, consider that the two processes occur at the same time: the occupations that are more segregated grow at the expense of less-segregated occupations, while at the same time segregation *within* each occupation declines. The overall change in segregation will be positive if the first process leads to a greater change than the second process. If attention is only paid to the total difference, the conclusion will be that segregation has become "worse" (which is a warranted statement, at least for the average worker). However, the statement is also imprecise, because the segregation of each individual occupation has in fact decreased. The decomposition of change into the two components thus allows to pinpoint more clearly the sources of segregation change. Importantly, the prevalence and direction of the two trends may call for different policy responses.

Drawing on early work by Karmel and Maclachlan (1988), this paper presents a procedure that allows one to decompose change into several components. To do this, the M index, introduced by Mora and Ruiz-Castillo (2003; 2009; 2011) but not widely used in the literature yet, is used instead of the H index. The two indices are closely related, but the decompositions undertaken in this paper are only possible with the M index. The decomposition allows to decompose change in the levels of M into a "marginal" and a "structural" component. The M index allows a deeper understanding of both of these components to understand the precise sources of change. The M index also allows for a straightforward solution to the problem of "appearing" and "disappearing" units. For instance, if occupational segregation is measured at two points in time, some occupations may have disappeared and new ones have been established. The same problem arises in school or workplace segregation problems.

The paper proceeds as follows: First, the M and the H indices are introduced, and some of their mathematical properties are discussed. Second, it is shown that both the M and the H are margin-dependent in both dimensions. Third, possible solutions to the margin-dependency problems are discussed, among them the decomposition procedure introduced by Mora and Ruiz-Castillo (2009), and margin-free indices. None of these solutions are deemed satisfactory. The decomposition procedure is introduced and extended. Lastly, the method is applied to occupational gender segregation in the U.S.

2 The M index and the H index

To make the following more concrete, consider N organizational units, such as schools or occupations, and a number of population groups, G, such as racial groups or genders. For a school racial segregation problem, the number of students in each school-racial group combination can be cross-classified in a $N \times G$ contingency table. A segregation index summarizes the $N \times G$ contingency table to a single measure.

Recently, the H has become increasingly popular for the study of racial segregation, which is most likely due to two distinct advantages. First, the H allows for attractive decompositions, which make it possible to quantify the amount of segregation that is induced at different levels. For instance, by grouping schools into school districts, we might distinguish the amount of total segregation that is due to between-district segregation from within-district segregation. Second, the H allows for a natural treatment of the multigroup case, which has become increasingly important in the U.S., and is a natural requirement in other segregation problems. In their comprehensive overview of multigroup segregation indices, Reardon and Firebaugh (2002) conclude "that the information theory index H is the most conceptually and mathematically satisfactory index" (p. 33).

In a recent string of papers, Mora and Ruiz-Castillo (2003; 2009; 2011) pointed to an alternative, but closely related index, which they called the *Mutual Information (M)* index. Both the M and H were introduced by Theil (Theil, 1967; Theil and Finizza, 1971; Theil, 1971, 1972). Mora and Ruiz-Castillo, as well as Frankel and Volij (2011), outlined some of the advantages of the M over the H. Importantly, Mora and Ruiz-Castillo (2011) showed that the decomposition of an H index into between- and within-group terms (for instance, white/non-white) may be ambiguous, and they thus recommend the adoption of the M. Mora and Ruiz-Castillo's critique of the H still stands, but recent papers in sociology and economics continue to use the H index.[2]

To formally define H and M, assume that we observe the racial composition of N schools in a city. We define t_{gn} as the number of students of race g in school n, and the total number of students as t. From this contingency table, we define $p_{\cdot n} = \sum_{g=1}^{G} t_{gn}/t$ and $p_{g\cdot} = \sum_{n=1}^{N} t_{gn}/t$ as the marginal probabilities of schools and racial groups, respectively. The joint probability of being in school n and racial group g is $p_{gn} = t_{gn}/t$. We also write $p_{g|n}$ as the conditional probability of being in racial group g given school n (and $p_{n|g}$ likewise). Lower-case probabilities p refer to single probabilities, while P refers to the vector of probabilities defining the distribution.

Intuitively, segregation is high when one school has many students of race A and few of race B, while another school has many students of race B and few race A. An absolute measure of segregation might consider schools segregated when the proportion of a single race crosses some threshold, say 80%. Besides the problematic arbitrary cut-off point, this raises the issue of the reference population: if the residential population surrounding each of the schools was 90% race A and 10% race B, it would be no surprise that many schools are majority race A as well. Intuitively, the measurement of segregation should thus be relative to the *expected* racial distribution in each school. This understanding of segregation is based on the unevenness of the racial distribution in different schools, but other definitions are possible (Massey and Denton, 1988).

The M index quantifies how strongly each school's racial group distribution deviates from

[2]Mora and Ruiz-Castillo (2011, p. 161) identify a small number of papers that prefer the M index over the H. Beyond those, DiPrete et al. (2017) and Forster and Bol (2018) have used the M index in the context of school-to-work linkages. However, most studies of segregation – be it in terms of residential segregation, school-racial segregation, workplace-racial segregation, or occupational-gender segregation – do adopt the H.

the overall (or expected) racial group distribution. This yields a segregation score for each school. The school segregation scores are then weighted by the size of the school, $p_{\cdot n}$. To measure the deviation, the logarithm of the ratio between actual and overall probabilities is used. As Theil (1972) has shown, the logarithm allows for the attractive additive decomposition properties. Thus,

$$M = \sum_n p_{\cdot n} \sum_g p_{g|n} \ln \frac{p_{g|n}}{p_{g\cdot}}. \tag{1}$$

Because the M is symmetric, it can also defined by summing proportion-weighted scores for each racial group, i.e.

$$M = \sum_g p_{g\cdot} \sum_n p_{n|g} \ln \frac{p_{n|g}}{p_{\cdot n}}. \tag{2}$$

Implicitly, equations 1 and 2 incorporate segregation scores for each school and each racial group, called local segregation scores. The local segregation for each school is

$$L(n) = \sum_g p_{g|n} \ln \frac{p_{g|n}}{p_{g\cdot}},$$

and equivalent for any racial group. The local segregation score for a school measures how strongly its racial distribution deviates from the overall racial distribution, while the local segregation score for each racial group measures how strongly its distribution across schools deviates from the overall distribution of students across schools.[3]

The M can also be motivated from an information-theoretic perspective, which is helpful to understand its basic properties. First, define the entropy $E(\cdot)$ of a distribution as

$$E(P) = -\sum_i p_i \ln p_i,$$

where P is a vector of probabilities that sums to 1. Entropy is a measure of *expected information* or *uncertainty* (Theil, 1972). Consider two events that occur with probabilities .99 and .01, respectively, and only one of the two events will occur eventually. When we are told that the first event occurred, we are not surprised, because it was virtually certain that this event occurred anyway. Hence, the expected information of this probability distribution is close to zero, i.e. $E(\{.99, .01\}) = .06$. On the other hand, for two events that will occur with a probability of ½ each, the outcome is uncertain, and the expected information is large, i.e. $E(\{.5, .5\}) = \ln 2 \approx .69$—the maximum entropy for a two-event probability distribution. More generally, the entropy is maximized at $\ln n$ when the probability of each event is $1/n$, where n is the number of events. Intuitively, the entropy is minimized at zero when it is

[3] The M can also be expressed in terms of the individual table cells. In this formulation, the expected value for each cell is calculated by multiplying the two marginal probabilities, p_g and p_n.

$$M = \sum_n \sum_g p_{gn} \ln \frac{p_{gn}}{p_g p_n} \tag{3}$$

This approach shows that the M is related to the likelihood-ratio test statistic (see Reardon et al., 2000, p. 362; Mora and Ruiz-Castillo, 2003, p. 154), and provides a natural way to relate the M index to other approaches that are concerned with the study of inequality in contingency tables.

certain that only one event will occur, i.e. $E(\{1,0\}) = 0$. In other words, uncertainty is low when entropy is low.

To define M from this perspective, we calculate the difference in expected information between a school's actual racial distribution and the overall racial group distribution. In information-theoretic terms: Given that we know the overall racial distribution, how much declines our uncertainty about a specific school's racial distribution once we know that school's racial distribution? Formally, this is the difference in entropies at the school level, weighted by the school's proportion:

$$M = \sum_n p_{\cdot n}[E(P_{g\cdot}) - E(P_{g|n})]. \tag{4}$$

Again, this can also be written from the racial group perspective:

$$M = \sum_g p_{g\cdot}[E(P_{\cdot n}) - E(P_{n|g})]. \tag{5}$$

From equation (4), it follows that M is minimized at zero when the racial group distribution of each school is identical to the overall racial group distribution. (Note that the entropy cannot be negative.) M is maximized at $\min(\{\ln N, \ln G\})$. To see this, assume that there are two racial groups, and the number of schools is greater than or equal to 2. Equation (4) is maximized when the entropy $E(P_{g\cdot})$ is maximized, and the entropy $E(P_{g|n})$ for each school is minimized. From above, we know that this is the case when each racial group in the city has the same proportion ($E(P_g) = \ln 2$, for the two-group case), and when each school contains only one racial group (for any n, $E(P_{g|n}) = 0$).

It may seem odd that a segregation index can only be maximized when all racial groups are the same size, but it is in line with information-theoretic expectations. This point will become clearer with an example. Consider two cities A and B, each with three schools and 200 students in total. For simplicity, we consider two racial groups, white and black. The cities differ in their proportion of white and black students. City A has 100 white and 100 black students, while city B has 20 white and 180 black students. The students are distributed across the two cities as follows, with the schools indexed by the rows of the matrix, and the racial groups indexed by the columns:

$$\text{A:}\begin{bmatrix} \textit{white} & \textit{black} \\ 100 & 0 \\ 0 & 50 \\ 0 & 50 \end{bmatrix} \qquad \text{B:}\begin{bmatrix} \textit{white} & \textit{black} \\ 20 & 0 \\ 0 & 90 \\ 0 & 90 \end{bmatrix}$$

In both cities, all three schools are completely segregated in the sense that there is no mixing within schools. If we apply any of equation (1)-(5) to these two matrices, we find that $M(A) = .69$ and $M(B) = .33$. The M index thus finds that segregation in city A is twice as high as in city B. This suggests to standardize the M index by the racial group entropy, which gives the H index:

$$H = \frac{M}{E(P_{g\cdot})}.$$

For the two cities, it follows that $H(A) = H(B) = 1$. The H is attractive because it is standardized between zero and one,[4] which facilitates comparisons between two cities with differing racial distributions. Nonetheless, there is an argument to be made for the M index as well. While the H index sees the amount of segregation as equal between the two cities, the M takes into account that it is much "harder" in city A to achieve complete segregation than it is in city B. Given that in city B 90% of the students are white, it is less surprising to find an all-white school in city B than it is in city A.

3 The problems and benefits of margin dependency

Studies of segregation are usually interested in differences of segregation levels over time or across space. In the school segregation literature, there is a debate about the resegregation of schools along racial lines (Reardon and Owens, 2014). The workplace segregation literature documented a decrease in workplace racial segregation levels, a finding that has recently been challenged (Ferguson and Koning, 2018). The gender-occupational literature is interested not only in comparing segregation over time, but also across regional or national economies.

The approach that is usually taken is to calculate the segregation index at several points in time. An increase in the segregation index is then taken as problematic, and a decrease as beneficial. While this approach sounds straightforward enough, it may obscure the underlying processes that are responsible for the changing levels of segregation. Consider, for instance, occupational segregation by gender. If measured using the M, overall segregation depends on the marginal distributions of occupations and genders. The M may thus change if the proportion of women in the labor force changes, or if the distribution of workers across occupations changes. The M is margin-dependent for both groups and units. Other indices, such as the index of dissimilarity D, are only margin-dependent in terms of the occupational distribution. The (apparent) problem of margin dependency has led to the development of alternative indices, such as the size-standardized index of dissimilarity SSD (which, while not margin-dependent on the occupational distribution, reintroduces a dependency on the gender distribution), or the log-linear index A, which is margin-free in both dimensions (Charles and Grusky, 1995).

However, none of the alternative indices since proposed have the same properties as the entropy-based multi-group indices, which alone fulfill the necessary theoretical criteria for the measurement of segregation (Reardon and Firebaugh, 2002; Mora and Ruiz-Castillo, 2011;

[4] Two caveats: First, the standardization is only limited to the range from zero to one when $N \geq G$, which is the case in most segregation problems. Alternatively, Mora and Ruiz-Castillo (2011) also define the H^* index. This index is defined by standardizing the M by the school distribution entropy, i.e. $H^* = \frac{M}{E(P_{\cdot n})}$. This index is maximized when $E(P_{i|g}) = 0$ for every racial group, which is only true when all members of each racial group are concentrated at exactly *one* school. However, this is not possible with two racial groups and more than two schools with positive enrollment. It is thus not the case that the H^* index "reaches the unity whenever there is no racial mix within schools" (Mora and Ruiz-Castillo, 2011, p. 173). This can be easily verified, as in cities A and B there is no racial mix within schools, but $H^*(A) = 0.66 < 1$ and $H^*(B) = 0.34 < 1$. The H^* index is thus only appropriate when $G \geq N$, which for practical segregation problems is usually not the case. Second, the standardization only works when the size of the smallest group is larger than the smallest organizational unit. For instance, consider a city of 200 black and 700 white students, distributed across three schools of 300 students each. Even if the schools are maximally segregated (i.e. two all-white schools, and one school with 200 black and 100 white students), the indices reach their maxima at $M = .32$ and $H = .6$. Whether such constraints matter in practice depends on the concrete application.

Watts, 1998). In fact, the criterion of margin-independence conflicts with the criterion of organizational equivalence. Organizational equivalence implies that when two occupations with the same level of segregation are combined, segregation should be unchanged (James and Taeuber, 1985). This criterion is not fulfilled when occupations are equally weighted and the segregation level of the other, uncombined occupations differ from the two occupations that are combined. This shows that the discussion about the merits of margin-free versus margin-dependent indices cannot be resolved, because the two indices pursue goals that are not compatible. On the balance, however, it seems that the natural way of dealing with multiple groups and the attractive decomposition properties make entropy-based indices the more attractive choice for many studies of segregation.

While the issue of margin dependency has long been debated in the sociological literature (and especially in the occupational gender segregation literature), empirical studies often do not discuss this problem. This might be due to the fact that the H index is normalized, which could suggest that changes in the marginals are taken into account. For instance, An and Gamoran (2009, p. 20) write that they "use a measure [the H index] that is insensitive to changes in the U.S. school population, thereby concentrating solely on racial imbalance." A simple example shows that this is not the case. Consider three occupations and two genders. At time point 1, there are 55 men and 45 women, distributed across occupations in a way that the first occupation is integrated, while the other two are rather segregated. This matrix is shown at the left-hand sides of the arrows, with men in the first and women in the second column:

$$t_1 : \begin{bmatrix} 25 & 25 \\ 28 & 2 \\ 2 & 18 \end{bmatrix} \quad \rightarrow \quad t_2 : \begin{bmatrix} 25 & 25 \\ 28 & 2 \\ 4 & 36 \end{bmatrix}$$

$$t_1 : \begin{bmatrix} 25 & 25 \\ 28 & 2 \\ 2 & 18 \end{bmatrix} \quad \rightarrow \quad t_2' : \begin{bmatrix} 25 & 50 \\ 28 & 4 \\ 2 & 36 \end{bmatrix}$$

Consider then two counterfactual scenarios. In the first row, the size of the third occupation doubles, leaving all other values unchanged. Note that there are now more women, and that it is not possible to double the size of the third occupation *and* keeping the gender proportion constant without changing the internal proportions in the other two occupations. In the second row, the size of the female labor force doubles, with the numbers for men unchanged. The H and M indices, along with the margin-free A index (see appendix for formula) are then calculated for the different matrices (Table 1).

Measure	t_1	t_2	t_2'
M	0.203	0.233	0.197
H	0.295	0.337	0.297
A	7.222	7.222	7.222

Table 1: Changes in M and H due to changes in margins

Both the M and the H react to changes in the occupational distribution, as shown in the column for t_2. M increases by 22%, and H by 23%. This result is expected and confirms that both the M and the H are unit-margin dependent: All else equal, when the size of a segregated occupation increases, overall segregation increases. In the second case, the result is a bit more ambiguous. Both M and H register the doubling of the female labor force as a small decline in segregation. The M declines by 3%, the H by less than 1%. Clearly, the doubling of the female labor force has affected the entropy of the gender distribution, and because the H is standardized by this entropy, it is less affected by the change in the marginals. However, larger marginal changes will also affect the H more than shown here. Thus, both the M and the H are also group-margin dependent (for a formal proof, see Mora and Ruiz-Castillo, 2011). In comparison, the A index indicates that the level of segregation is unchanged in the two alternative scenarios.

How are these changes—introduced by shifting the marginal distributions—different from structural changes of segregation, which would influence the A index? Structural change can be described by changes in the odds ratios that define the contingency table. Two local odds ratios are sufficient to describe the association structure of a 3×2 table (as in the example above), which are $\theta_{1,1} = \frac{t_{1,1}t_{2,2}}{t_{1,2}t_{2,1}}$ and $\theta_{2,1} = \frac{t_{2,1}t_{3,2}}{t_{2,2}t_{3,1}}$ Agresti (2013, p. 54). At all three time points, these odds ratios are the same, which is to say that the association structure between occupations and gender does not change from t_1 to t_2 or from t_1 to t_2'.

Given the problems of margin-free indices, but realizing that margin-dependency can be problematic for the study of changes in segregation levels, Mora and Ruiz-Castillo (2009) presented two formulas that supposedly quantify structural and compositional change between two M indices. With a slightly adapted notation, the difference between two M indices, defined by the matrices t_1 and t_2, is decomposed as follows:

$$\begin{aligned} M(t_2) - M(t_1) &= \Delta\mathrm{N}(\Pi^n) + \Delta\mathrm{G}^n + \Delta\mathrm{U}(\Pi^n) \\ &= \Delta\mathrm{N}(\Pi_g) + \Delta\mathrm{U}_g + \Delta\mathrm{G}(\Pi_g), \end{aligned} \tag{6}$$

where the ΔU and ΔG capture changes in the marginals of unit and group proportions, respectively, and ΔN captures "composition-invariant" changes, which, however, are not the same as structural changes defined through the change in odds ratios. As Mora and Ruiz-Castillo themselves write, the interpretation of these terms hinges on crucial assumptions which are rarely met in practice (Watts, n.d.). For reasons of brevity, these problems are not explicated fully here. Instead, another problematic aspect of these decompositions is highlighted, and that is that there are two answers for each of the three components, which might furthermore be contradictory. The decompositions on the first and the second line will only in exceptional circumstances give the same numerical results. This is easily seen by applying equation (6) to the difference between t_1 and t_2 from the example in the previous section:

$$\begin{aligned} M(t_2) - M(t_1) &= 0 + 0.00376 + 0.0267 \\ &= -0.0209 + 0.0479 + 0.00346 = .03 \end{aligned}$$

The first decomposition implies that structural change is zero, and further suggests that the marginal change in the occupational distribution is largely responsible for the increase in segregation, which aligns with our expectations. However, the second line gives a contradictory answer, implying that structural segregation decreased. Furthermore, the size of the marginal components is not the same in the two decompositions. Even if the assumptions that underlie these decompositions were justified in practice (which is questionable), the fact that the two decompositions give two – and two possibly contradictory – answers is unsatisfactory and poses practical problems of interpretation.

4 Decomposition of change

Instead of attempting the margin-free measurement of segregation at each point in time, the approach outlined here follows the idea that *changes* in segregation indices can be decomposed into marginal and structural changes (Watts, 1998, n.d.; Mora and Ruiz-Castillo, 2009). This method was proposed by Theil himself Theil (1972, p. 131ff.), and was extended by Karmel and Maclachlan (1988) in the context of sex segregation. Karmel and Maclachlan used another segregation index, but the approach is applicable whenever a margin-free comparison of two contingency tables is desired.

The basic idea is to adjust the contingency table at time point t_1 so that the marginal changes between time points t_1 and t_2 are taken into account. Consider a city with black and white students distributed across three schools. We observe the city at two points in time. Between these two time points, the black population has grown and the white population has declined. At the same time school enrollment has changed in all three schools, with especially strong declines in the third school. The question is: if there are changes in segregation, how much of these changes can be attributed to changes in the distribution of school and racial group marginals alone, and how much of the change can be attributed to structural segregation change?

At the two time points, the students are distributed across schools as follows:

$$t_1: \begin{bmatrix} \textit{white} & \textit{black} \\ 20 & 100 \\ 180 & 50 \\ 600 & 50 \end{bmatrix} \qquad t_2: \begin{bmatrix} \textit{white} & \textit{black} \\ 10 & 170 \\ 80 & 60 \\ 240 & 40 \end{bmatrix}$$

Both the M and the H register large changes in segregation: the M increases by over 80% between t_1 and t_2, while the H increases by 33%. To identify how much of this change is due to marginal changes in the distribution of students across schools and racial groups, the matrix at t_1 is transformed to have the same margins as the city at t_2, while leaving the association structure (i.e., the odds ratios) intact. This can be achieved using Iterative Proportional Fitting (IPF): First, the cells of t_1 are scaled to achieve the overall *racial* marginal distribution of t_2. The adjusted cell counts are then scaled to achieve the marginal *school* distribution of t_2. This process is repeated until the margins of the adjusted table are within a small percentage of t_2. The first steps of the procedure are shown here:

$$
\begin{aligned}
&\begin{bmatrix} 20 & 100 \\ 180 & 50 \\ 600 & 50 \end{bmatrix} && \Longrightarrow \begin{bmatrix} 20 \times 330/800 & 100 \times 270/200 \\ 180 \times 330/800 & 50 \times 270/200 \\ 600 \times 330/800 & 50 \times 270/200 \end{bmatrix} \\
= &\begin{bmatrix} 8.3 & 135 \\ 74.3 & 67.5 \\ 248 & 67.5 \end{bmatrix} && \Longrightarrow \begin{bmatrix} 8.3 \times 180/144.4 & 135 \times 180/144.3 \\ 74.3 \times 140/141.8 & 67.5 \times 140/141.8 \\ 248 \times 280/315.5 & 67.5 \times 280/315.5 \end{bmatrix} \\
= &\begin{bmatrix} 10.3 & 168.4 \\ 73.4 & 66.6 \\ 220.1 & 59.9 \end{bmatrix} && \Longrightarrow \begin{bmatrix} 10.3 \times 330/303.8 & 168.4 \times 270/294.9 \\ 73.4 \times 330/303.8 & 66.6 \times 270/294.9 \\ 220.1 \times 330/303.8 & 59.9 \times 270/294.9 \end{bmatrix} \\
= &\begin{bmatrix} 11.2 & 154.2 \\ 79.7 & 61 \\ 239.1 & 54.8 \end{bmatrix} && \Longrightarrow \begin{bmatrix} 11.2 \times 180/165.4 & 154.2 \times 180/165.4 \\ 79.7 \times 140/140.8 & 61 \times 140/140.8 \\ 239.1 \times 280/293.9 & 54.8 \times 280/293.9 \end{bmatrix} \\
= &\begin{bmatrix} 12.2 & 167.8 \\ 79.2 & 60.7 \\ 227.8 & 52.2 \end{bmatrix} && \Longrightarrow \text{... (10 steps omitted)} \\
= &\begin{bmatrix} 13.7 & 166 \\ 83.5 & 56.5 \\ 233 & 47.3 \end{bmatrix} = t_1'
\end{aligned}
$$

The transformations at rows one and three adjust the racial group marginals; while the transformations at rows two and four adjust the school marginals. It is unimportant whether the procedure starts with the group or the unit marginals; it will always converge (for details on IPF, see Deming and Stephan, 1940; Agresti, 2013). After four steps, the racial and occupational marginals are already within 3-4% of the desired marginals. After 14 steps, the procedure yields the matrix at the last row, where the marginals are within 0.1% of the desired marginals. The resulting matrix is a counterfactual version of the t_1 matrix, where only the marginals changed in the direction empirically observed in t_2, but the odds ratios are the same as in t_1. This allows a decomposition of overall change in segregation levels as follows:

$$
\begin{aligned}
M(t_2) - M(t_1) &= \overbrace{M(t_1') - M(t_1)}^{\text{marginal}} + \overbrace{M(t_2) - M(t_1')}^{\text{structural}} \qquad (7) \\
&= (.238 - .150) + (.273 - .238) \\
&= .088 + .035 = .123
\end{aligned}
$$

The "marginal" component quantifies how much we would expect segregation to change given that the marginals changed towards those of t_2. The "structural" component quantifies any additional amount of segregation that is unexplained by marginal changes. In this case, 72% of the increase in segregation can be attributed to marginal changes, and 28% to structural changes.

To understand the behaviour of the decomposition, it is useful to consider the two extreme cases of "structural change only" and "marginal change only". Considering t_1, it is possible to construct an alternative matrix that redistributes the number of students across schools in such a way that the marginals will stay the same (e.g., by distributing 50 students from school 1 to the other two schools, and moving the same number of white students to school 1.) A change decomposition of these two matrices will find that marginal change is zero, because the IPF procedure converges immediately without changing any cell counts. Thus, the marginal term of Equation (11) would compare identical matrices. Similarly, it is also possible to construct a matrix where simply the number of black students doubled. In this case, the IPF procedure scales the margins in exactly this way, which means that the structural term of Equation (7) compares identical matrices.

This aggregate view of segregation differences can be further decomposed. The key property that is exploited here is that in the marginal component, the odds ratios are the same, and that in the structural component, the marginal distributions of units and groups are the same.

Decomposing marginal changes

The marginal change can be further subdivided into three components: one component quantifies the contribution of changing unit marginals, one quantifies the contribution of changing group marginals, and one is an interaction effect. The first two marginal effects are equivalent to just scaling the marginals of *either* units or groups of t_1 to t_2.

$$
\begin{aligned}
M(t_1') - M(t_1) &= (M(\begin{bmatrix} 20 \times 330/800 & 100 \times 270/200 \\ 180 \times 330/800 & 50 \times 270/200 \\ 600 \times 330/800 & 50 \times 270/200 \end{bmatrix}) - M(\begin{bmatrix} 20 & 100 \\ 180 & 50 \\ 600 & 50 \end{bmatrix})) \\
&+ (M(\begin{bmatrix} 20 \times 180/120 & 100 \times 180/120 \\ 180 \times 140/230 & 50 \times 140/230 \\ 600 \times 280/650 & 50 \times 280/650 \end{bmatrix}) - M(\begin{bmatrix} 20 & 100 \\ 180 & 50 \\ 600 & 50 \end{bmatrix})) + \text{Interaction term} \\
&= \underbrace{.05}_{\text{race}} + \underbrace{.11}_{\text{schools}} + \underbrace{(-.07)}_{\text{interaction}}
\end{aligned}
$$

In the first term, the racial distribution is scaled towards that of t_2 (as in the first step of the IPF procedure), in the second term, the school distribution is scaled towards that of t_2. Because the change in marginals is not additive, an interaction term is needed to explain increases in segregation that are not explained by the separate scaling of the two dimensions. In this case, the effect of the change in the distribution of students over schools is estimated to have twice the impact on raising segregation than the changes in the racial distribution. To simplify the interpretation of the interaction term, it can also be distributed proportionally over the two remaining terms (similar to an Oaxaca-Blinder decomposition):

$$
\begin{aligned}
.05 + .11 + (-.07) &= (.05 - .07\frac{.05}{.05 + .11}) + (0.05 - .07\frac{.11}{.05 + .11}) \\
&= .03 + .06
\end{aligned}
\tag{8}
$$

In this case, the marginal change for both units and groups is positive, and the interaction effect is negative, but this does not have to be the case.

School	Proportion	Local linkage		Weighted difference
n	p_n	$L(t_2;n)$	$L(t_1';n)$	$p_n(L(t_2;n) - L(t_1';n))$
1	0.300	0.573	0.514	0.018
2	0.233	0.001	0.004	-0.001
3	0.467	0.216	0.178	0.018

Table 2: Decomposition of structural changes: $M(t_2) - M(t_1')$

Decomposing structural changes

Usually, structural change is of greater interest than marginal change. The term for the structural component admits two straightforward decompositions based on local segregation scores. These decompositions were not exploited by Karmel and Maclachlan (1988) or others, because their index did not admit disaggregation by local segregation scores. The key property that these decompositions exploit is that $p_n^{t_2} = p_n^{t_1'}$ and $p_g^{t_2} = p_g^{t_1'}$, i.e. the equivalence of the margins. We can thus write:

$$\begin{aligned} M(t_2) - M(t_1') &= \sum_{n=1}^{N} p_n^{t_2}(L(t_2;n) - L(t_1';n)) \\ &= \sum_{n=1}^{N} p_n^{t_2} \sum_{g=1}^{G} p_{g|n}^{t_2} \ln \frac{p_{g|n}^{t_2}}{p_g^{t_2}} - p_{g|n}^{t_1'} \ln \frac{p_{g|n}^{t_1'}}{p_g^{t_2}} \end{aligned}$$

where $L(X;n)$ is meant to mean the local segregation score for unit n in matrix X. The difference in structural segregation can thus be attributed solely to differences in the conditional probabilities, holding the marginals constant. Table (2) shows the results for the decomposition of local segregation scores for schools. For schools one and three, structural segregation increased by 12% and 21%, respectively. For school two, local segregation is low and almost unchanged. By weighting the difference in local scores, we can conclude that schools one and three are equally responsible for an increase in structural segregation. In more realistic settings with a greater number of units and groups, the local segregation scores could now also be grouped by school district or another characteristic (public/private, charter), if individual schools are not of much interest. The sources of an increase or decrease in structural segregation can thus be precisely understood, and this is a key advantage of the M index over other segregation indices.

Comparisons across space or between subgroups

Until now, we have been concerned with changes in segregation over time by taking the difference $M(t_2) - M(t_1)$. However, segregation levels are often also compared across space or between population groups. For instance, we might compare the occupational sex segregation of older and younger cohorts. If the M or H indices are used, the comparison is likely tainted both by the difference in the occupational distribution between younger and older workers, and by differences in female labor force participation. The same is true for the comparison of sex segregation between two countries. As there is no obvious reference group in these cases (although one might be arbitrarily chosen), equation 7 can be adapted to achieve a symmetric comparison. Let A and B be two contingency tables that are compared, then:

$$M(B) - M(A) = \overbrace{\frac{1}{2}(M(A') - M(A)) + \frac{1}{2}(M(B) - M(B')}^{\text{marginal}} \\ + \underbrace{\frac{1}{2}(M(B) - M(A')) + \frac{1}{2}(M(B') - M(A))}_{\text{structural}}. \tag{9}$$

Note that this requires two separate IPF procedures, one which transform A into A' by adjusting towards the margins of B, and one which transforms B into B' by adjusting the margins towards A. The more detailed decompositions can be adjusted in the same way, by weighting the forward and backward change equally.

In many cases, the comparison between places will not be defined by the same units and/or groups. For instance, when occupational segregation is compared between two cities using the same data source or classification scheme, the occupational scheme and the gender groups will be the same in both cities, and equation (9) can be applied. But when comparing racial residential or school segregation between two cities, the units (e..g, schools, tracts, or blocks) will not be comparable, and the IPF procedure cannot be applied. Of course, the M and the H values between the two cities can be compared, but whether this comparison is meaningful depends on the specific case and the theoretical comparability of the units. It might be worthwhile in these cases to adjust just the racial group margin to achieve higher comparability.

Appearance and disappearance of units

Until now, we assumed that at both points in time, all units and groups have non-zero counts. However, this assumption is often not met in practice. In the case of school segregation, schools may have closed down and new schools may have opened. In the case of occupational segregation, some occupations may have vanished and now occupations have become established. Capitalizing on the decomposition properties of the M, the approach used here can be extended to account for the effects of adding or removing units and groups.

Assume the simple case that in a city of five schools and two racial groups, two schools close down:

$$t_1 : \begin{bmatrix} 5 & 15 \\ 15 & 5 \\ 10 & 10 \\ 5 & 15 \\ 15 & 5 \end{bmatrix} \qquad \rightarrow \qquad t_2 : \begin{bmatrix} 8 & 23 \\ 23 & 8 \\ 19 & 19 \\ & \\ & \end{bmatrix}$$

In this scenario, the students from the closed schools were distributed across the remaining schools, so that there are still 50 students in each racial group. Between t_1 and t_2, the M

declines from .105 to .076. Is this purely an effect of there being fewer units now? Or were the schools that closed more segregated than the schools that stayed open?

To answer this question, define the set $S = \{1, 2, 3\}$ for the three schools that stay open, and $D = \{4, 5\}$ for the schools that close. The sets S and D define "super-units" that are composed of individual units, and the share p_D is the proportion of students in set D at t_1. The goal is to decompose $M(t_1)$ into the contribution of the schools that close down and those that continue to operate, which can be done using the general form of the between-within decomposition of M (Mora and Ruiz-Castillo, 2011). Total segregation thus equals the between-super-unit M plus the weighted M within the two matrices defined by the two super-units, i.e.

$$M(t_1) = \underbrace{M\begin{pmatrix} 30 & 30 \\ 20 & 20 \end{pmatrix}}_{\text{between open/closed}} + p_D \underbrace{M\begin{pmatrix} 5 & 15 \\ 15 & 5 \end{pmatrix}}_{\text{within closed}} + (1 - p_D) \underbrace{M\begin{pmatrix} 5 & 15 \\ 15 & 5 \\ 10 & 10 \end{pmatrix}}_{\text{within open}}$$

Then solve for the last M term, which we call M^*:

$$\begin{aligned} M^*(t_1) = M\begin{pmatrix} 5 & 15 \\ 15 & 5 \\ 10 & 10 \end{pmatrix} &= \frac{1}{1 - p_D}\left[M(t_1) - M\begin{pmatrix} 30 & 30 \\ 20 & 20 \end{pmatrix} - p_D M\begin{pmatrix} 5 & 15 \\ 15 & 5 \end{pmatrix}\right] \\ = .087 &= \frac{1}{.6}\left[.105 - 0 - .4 \times .131\right] \end{aligned}$$

This expression summarizes the mechanical effect of dropping schools on the M index. To arrive at the "reduced M" on the left-hand side, we subtract from M all the sources of segregation that are due to the dropped schools only, which consists of a "between" and a "within" term. The between term summarizes how strongly the racial group composition of the closed school deviates from the schools that stay open, in total, while the within term summarizes how much segregation there is *within* the population of schools that close. The division by $1 - p_D$ has the effect of scaling the other schools' proportions upward.[5]

$M^*(t_1)$ will be larger than $M(t_1)$ when the schools that closed were less segregated compared to the remaining schools, and will be smaller in the opposite case. In this case, removing schools 4 and 5 from t_1 reduces the M from $M(t_1) = .105$ to $M^*(t_1) = .087$. The "reduced M" can now be compared to the situation at t_2 using the regular IPF method. The approach outlined here thus amounts simply to a comparison of only those units that overlap across time points. However, an advantage of the M is that there is an intuitive interpretation for the "missing" units.

Applying the IPF procedure to the example above, the total decomposition is the following:

$$\begin{aligned} M(t_2) - M(t_1) &= \text{removals} + \text{marginal changes} + \text{structural changes} \\ 0.076 - 0.105 &= -0.017 + -0.006 + -0.006 = -0.029 \end{aligned}$$

[5] This can be clearly seen by assuming that we drop one unsegregated school only (school 3 from t_1). Then, the expression simplifies to $M^{\text{dropped}} = \frac{M(t_1)}{1-p_D}$. This shows that the mechanical consequence on M when a unsegregated school closes depends only on the size of the closed school, p_D.

In total, about 60% of the decline in segregation can be attributed to the effect of removing schools 4 and 5. The remaining decline is equally due to changes in the marginals and to structural changes.

For simplicity, the example was only concerned with the removal of units, but additional units, such as newly opened schools, can be handled in exactly the same way.

Summary of decomposition approach

The total decomposition of change between two segregation indices is thus:

$$\begin{aligned} M(t_2) - M(t_1) &= \text{additions} \\ &+ \text{removals} \\ &+ \text{marginal changes} \\ &+ \text{structural changes} \end{aligned} \tag{10}$$

Depending on the segregation problem, all of these four terms can be decomposed further to study the precise sources of change. For most segregation problems, equation 10 is the minimum that is required to robustly understand changes in segregation, because the possible sources of change may point in opposite direction. Large changes in the marginals may hide worsening segregation at the structural level, or improvements in structural segregation might be overwhelmed by changes in the marginals.

Note also that this procedure can be used to decompose any M index. Because the many possible decompositions of an M index again yield M indices, their change can also be studied over time. For instance, when studying school segregation in a metropolitan area, one might be interested in the change not only in the total M, but also for the partial M indices that define segregation between the central city and the suburbs, within the central city, and within suburbs. It is well known that the total M can be decomposed as follows:

$$M = M^{\text{between}} + p_{\text{city}} M^{\text{city}} + p_{\text{suburbs}} M^{\text{suburbs}}, \tag{11}$$

where p_{city} and p_{suburbs} define the proportion of the student population living in cities and suburbs, respectively. When change is observed over time, the three M indices defined in this decomposition can then be studied using the IPF procedure outlined here.

5 Example

To consider the practical value of the above, we study occupational sex segregation in the U.S. between 1990 and 2016. IPUMS provides harmonized occupational codings based on the 1990 Census occupational codes for this period (Ruggles et al., 2018). The sample has been selected to comprise the employed population aged 16-66 with non-missing occupations. All estimates reported below are weighted. The occupational codings for 1990 can be grouped into seven major groups. Because the “military” major group contains only one occupation, we drop it. The “Managerial and Professional” and the “Technical, Sales, Admin” major groups are very large—accounting for a quarter to a third of the total occupational distribution—, so they are further subdivided according to the official Census groupings. The occupations are thus grouped into nine major groups (see Table 3).

When comparing occupations over time, two problems arise. First, the degree to which fine-grained occupations are recorded changes over time, and this is often a problem induced by the harmonization efforts. For instance, "Sociologists" are not coded separately in 2000-2016, but are available as a separate code in 1990. This is because the schemes originally used to record occupations in later years do not identify sociologists separately. Second, occupations may vanish or new occupations may appear. Occupations such as "Stenographers" or "Railroad brake, coupler, and switch operators," for instance, are no longer coded in later years, and this is probably because they no longer exist as recognizable occupations. In many cases, it is hard to distinguish whether the problem is one of harmonization or one of disappearing occupations. For the purpose of this example, we will make the simplifying assumption that the harmonized occupations that are coded in each year represent recognizable, established occupations.

Descriptive statistics and total segregation

Table 3 contains descriptive statistics by year. Panel A shows the number of unique occupations that are available in each year, along with the number of categories that appear and disappear in each year. Panels B and C document a well-known pattern of occupational change, both in terms of female labor force participation and in terms of a changing occupational distribution. Panel D shows that there is considerable heterogeneity in terms of female labor force participation across occupational groups, and a heterogeneous pattern of change. In most occupational groups, female labor force participation increased, while in the Administrative and Operators/Laborers major groups, the share of women declined.

The full contingency table that is analyzed here contains as rows the detailed occupational codings, and as columns the number of male and female workers in each occupation. We calculate the M and the H for the total labor force, as well as separately for each major occupational group. This is based on the decomposition of the M into between and within-cluster terms, as in Equation (11). In this case, the between group term measures the segregation that is induces by the major occupational groups alone, while the within terms measure the segregation of detailed occupations within each major group.

The results are shown in Figure 1. Overall gender segregation, shown in the top panel, declined by 15% from 1990 to 2016 for the H and the M.[6] In 1990, the H was at 31%, and declined to 26% by 2016. The between term also declined, which means that major occupational groupings are becoming less informative about gender composition over time. Note, however, that major occupational groupings explain a large amount of overall gender segregation: for the M, 45% of total segregation in 1990 and 42% in 2016 are explained by the gender segregation of major groups alone.

While overall segregation declined, the within-terms reveal some heterogeneity. In most major groups, gender segregation declined. In others, notably Farming and Forestry as well as

[6]Because the M is sensitive to the number of categories, one might suspect that the higher gender segregation in 1990 is an artifact of measurement. The normalization of the H index corrects for the changing number of categories, as shown above. Either way, in this case the variation in the number of occupations is too small to matter: if we restrict the calculation to the 317 occupations that are available at all five points in time, the M and H values are within 1% of the values presented in Figure 1.

	1990	2000	2010	2016
Sample size (in 1000)	5917	6542	1443	1441
A. Number of occupations				
Number of occupations	369	336	330	319
Appearing occupations		0	0	0
Disappearing occupations		33	6	11
B. Labor force participation (%)				
Female	46	47	48	48
C. Distribution of occupational major groups (%)				
Managerial	12	12	13	14
Professional	13	16	17	18
Technical	4	4	4	4
Sales	12	11	11	11
Adminstrative	16	16	14	13
Service	13	14	17	17
Farming, Forestry	2	2	2	2
Production, Craft	11	11	9	9
Operators, Laborers	16	14	12	12
D. Female labor force by major groups (%)				
Managerial	43	44	45	46
Professional	54	57	59	59
Technical	46	48	49	48
Sales	49	50	51	51
Adminstrative	78	74	72	70
Service	57	59	60	59
Farming, Forestry	17	18	17	19
Production, Craft	8	10	10	11
Operators, Laborers	27	25	20	20

Table 3: Descriptive statistics

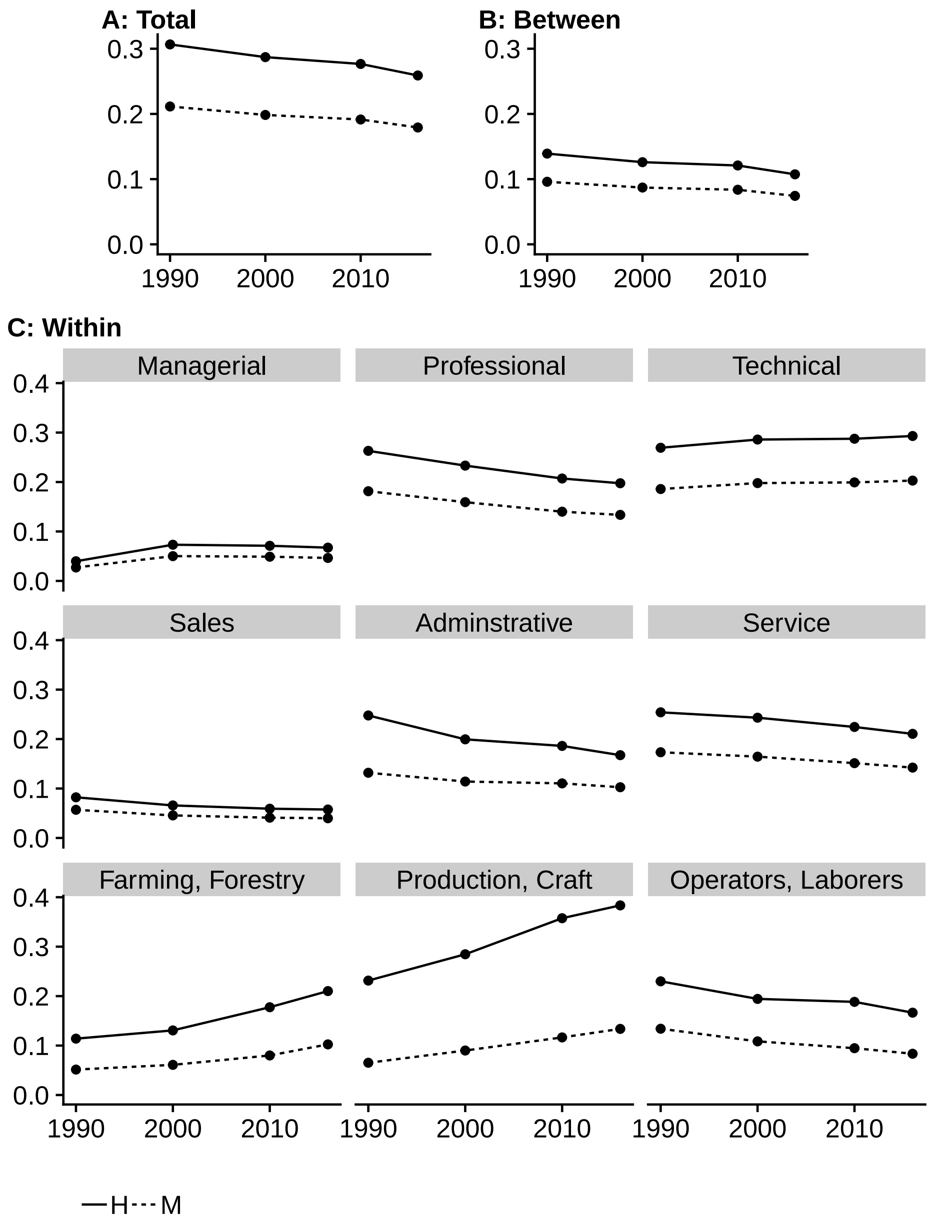

Figure 1: Occupational gender segregation, 1990-2016.
Panel A shows total segregation by gender and detailed occupations. Panel B shows segregation between gender and major occupational groups. Panel C shows within-major-group gender segregation by detailed occupations.

Component	M 1990	M 2016	M Diff.	Disappearing occupations	Marginal Occupation	Marginal Gender	Structural
Total	0.211	0.179	−0.032	0.001	−0.014	0.001	−0.020
			(100%)	(−5%)	(45%)	(−3%)	(63%)
Between major groups	0.096	0.074	−0.022	0.000	−0.016	0.001	−0.006
			(100%)	(−0%)	(75%)	(−3%)	(28%)
Within major groups (sorted by Diff.)							
Operators, Laborers	0.134	0.083	−0.051	0.015	−0.019	−0.013	−0.033
			(100%)	(−29%)	(37%)	(26%)	(64%)
Professional	0.181	0.134	−0.048	0.000	−0.011	−0.004	−0.033
			(100%)	(−1%)	(23%)	(8%)	(70%)
Service	0.173	0.142	−0.031	−0.002	−0.002	−0.000	−0.026
			(100%)	(6%)	(7%)	(2%)	(83%)
Adminstrative	0.132	0.103	−0.029	0.004	−0.002	0.010	−0.037
			(100%)	(−12%)	(6%)	(−33%)	(126%)
Sales	0.057	0.040	−0.017	0.000	0.003	−0.000	−0.020
			(100%)	(−0%)	(−16%)	(1%)	(114%)
Technical	0.186	0.203	0.017	0.005	0.014	0.001	−0.004
			(100%)	(28%)	(84%)	(9%)	(−22%)
Managerial	0.027	0.046	0.019	0.000	0.001	0.000	0.017
			(100%)	(0%)	(8%)	(2%)	(90%)
Farming, Forestry	0.051	0.102	0.051	0.001	0.024	0.006	0.020
			(100%)	(3%)	(46%)	(12%)	(39%)
Production, Craft	0.065	0.134	0.069	−0.002	0.028	0.017	0.024
			(100%)	(−3%)	(41%)	(24%)	(35%)

Table 4: Decomposition of change

Production and Craft occupations, gender segregation increased strongly. This heterogeneity suggests that its worthwhile to study major groups separately.

Decomposition of change

To simplify the analysis of change, we focus in the following on the changes between 1990 and 2016. To decompose change into its marginal and structural components, we apply the method developed above to the eleven M values shown in Figure 1. As shown in equation (8), we distribute the interaction across the two main marginal terms to simplify interpretation. Because no new occupations appear over time in this example, there is also only a term for disappearing occupations. The total difference of any M term is thus decomposed into four components: the effect of those occupations that are removed, the effect of the changing occupation marginal distribution, the effect of the changing gender marginal distribution, and the structural component.

For the total M, the decline can be attributed to the changing occupational structure—i.e., the labor force has shifted to occupations that are less segregated—, and, for the most part, to structural decrease. The decline in structural segregation accounts for 62% of the total decline in segregation.

Segregation declined in five out of the nine major groups, and the share of the structural component was high in all five groups (between 65 and 117%). Within the major group of Operators and Laborers, the occupations that disappear were relatively less segregated than the ones that remain, which increased segregation. However, the large marginal and structural components offset this small increase.

Segregation increase for four major groups. Except for the Managerial group, structural increase plays less of a role for these groups. For Technical occupations, structural change was in fact negative, but the marginal changes, especially the effect of the changing occupational distribution, led to an increase in segregation. In Farming and Forestry and Production and Craft occupations, structural segregation increased, but the changes in the marginals had a larger effect on the increase in segregation than the structural change. For the Managerial occupations, the increase in segregation is almost entirely due to a structural increase in segregation. Overall, a rough pattern emerges: For those occupational major groups where segregation declined, it declined in large part because of a structural decrease in segregation. When segregation increased, it increased mostly because of changes in the marginal distributions—with the notable exception of Managerial occupations.

The increasing labor force participation of women accounts for only a minor part of the overall segregation difference: around 3% of the total change is explained by changing gender marginals. One might wonder why the sign of these effects does not correspond to the changing patterns of female labor force participation from Table 3. For instance, the female share of production and craft workers has increased from 8% to 11%, and this led to an expected *increase* in segregation. To understand why this is the case, consider the example of carpenters. In 1990, this occupation was 98.2% male, while the male share in the major group was 92.2%. This leads to a local segregation score for carpenters (within the major group) of $.982 * \ln(.982/.922) + .018 * \ln(.018/.078) = 0.036$. In 2016, the overall share of male workers is now 88.9%, which represents a reduction in the share of about 4% for men and an increase in the share of women of about 42%. After proportionally increasing the number of women and reducing the number of men, the share of carpenters that are men is now 97.4%. (Note that the marginal adjustment changes not only the gender proportion of the major group, but also each occupation's gender distribution.) This leads to a counterfactual segregation score of $.974 * \ln(.974/.889) + .026 * \ln(.026/.111) = 0.051$, which is higher than before. In this case, the expected effect of proportionally increasing the share of women within each occupation increases segregation, because it emphasizes existing patterns of segregation even more. This relationship, however, is not deterministic. For "Operators and Laborers" as well "Administrative" occupations, the share of women declined, but the marginal effect is negative in the first case and positive in the second. The effect of the changing patterns of female labor force participation thus depends on the existing association structure between occupations and gender. This shows that the marginal effects have to be interpreted with caution. They represent *expected* changes in segregation under ceteris paribus conditions.

To learn more about the changes in different occupations, a further analysis could decompose structural change for each occupation as done in Table 2. This allows the researcher to study whether structural change in each occupation is associated with other occupational characteristics, such as the gender pay gap or educational inequalities.

		Occupation	
		margin dependent	**margin free**
Gender	**margin dependent**	M H	SSD
	margin free	D	A adjusted M

Table 5: Margin-dependency of different indices
Adapted from Charles and Grusky (1995, p. 934)

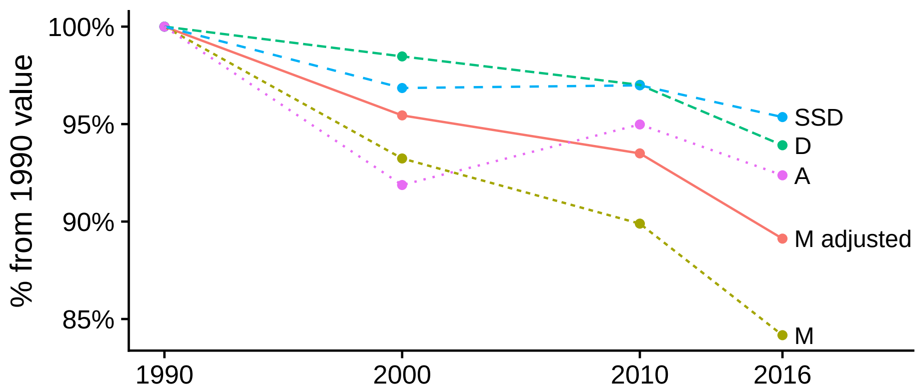

Figure 2: Comparison of M and margins-adjusted M indices with the index of dissimilarity D, the size-adjusted index of dissimilarity SSD, and the log-linear index A

Comparison with other indices

The decomposition makes it possible to create a time series of adjusted M indices that are not confounded by marginal changes. To do this, we choose 1990 as the reference year and adjust the other years (2000, 2010, 2016) towards the marginals of the year 1990. Alongside with the adjusted M index, we also calculate the observed M, and three other common indices: the index of dissimilarity D (Duncan and Duncan, 1955), the size-standardized index of dissimilarity SSD, and the closed-form of the log-linear index A (Charles and Grusky, 1995; Grusky and Charles, 1998) (see Appendix for formulae).

As Table 5 shows, the M and H indices are margin-dependent in both directions, while the D and SSD are margin-dependent in either one or the other direction. The A index and the adjusted M are margin-free in both directions—the A by weighting each occupation equally, the M by making not the index in the absolute margin-free, but in its relative comparison to other states. Of course, the adjusted M is also margin-dependent in the sense that occupations and genders are weighted, in this case, by their 1990 proportions, but as shown above, this can be a desirable property.

The results for the five indices are plotted in Figure 2. To ease comparison across the indices, the absolute numbers are transformed to be percentages of the 1990 values.[7] First, it should be noted that all indices register a decline in segregation (although the A and SSD indices increase between 2000 and 2010). The structural decline, as calculated by the adjusted

[7]The H index is not plotted because the line overlaps almost exactly with the line for the M.

M, amounts to 10 percentage points of the 1990 value. The observed M clearly overstates the decline—as seen in Table 4, this is because the change in the occupational margins contributed to the decrease in segregation. The other indices underestimate structural change compared to the adjusted M. The differences between the margin-free A and adjusted M are due to the different occupational weights. The A weights each occupation equally, which makes is susceptible to extreme values for small occupations that arise from sampling variability (Watts, 1998), which is a possible explanation for its more erratic movement compared to the other indices. The A thus quantifies the segregation of the average occupation, while the M quantifies the segregation of the average worker.[8]

6 Conclusion

Entropy-based segregation indices have many useful properties that allow a more precise understanding of segregation. This includes the decomposability of the M or H index into between-cluster and within-cluster terms and the decomposability into local segregation scores. They also handle the study of segregation with more than two groups, which is a natural requirement in many segregation problems. However, the M and H indices are both margin-dependent: the level of segregation may change when either margin changes. Margin-independent indices are not a viable alternative in many situations. They do not handle multiple groups, and they solve the margin-dependency problem by giving each unit the same weight. As shown above, this approach conflicts with an important principle that segregation indices should obey, the principle of organizational equivalence. The conflict between margin-dependent and margin-free indices thus cannot be resolved.

The approach taken in this paper instead seeks to decompose *changes* in segregation levels. It has been shown that the difference between two M indices can be decomposed into marginal and structural changes. This method is, in principle, applicable to any segregation index. However, the advantages of the M index became apparent when a further decomposition of the marginal and structural terms is desired. The decomposition of the structural term into the contribution of individual units or groups is especially useful. The change in structural segregation allows a more precise testing of hypotheses about the causes and effects of changing levels of segregation at the individual unit level, and this change will be uninfluenced by changes in the marginal distributions.

Lastly, it should be noted that changes or differences in the marginal distributions may often be the relevant social fact compared to differences in segregation levels. Consider the comparison of a city over time. At the beginning of the period, the city has a balanced racial population, and at the end of the period, one racial group has a 90% majority. One could compare the absolute level of segregation over time, but it seems that in such an extreme case, the relevant difference seems to be the starkly different demographic profiles.[9] When there is a lot of change in the marginal distributions (as in this case), it is also questionable if the IPF method provides a convincing approximation. The IPF method would adjust the balanced racial distribution at the first time point towards the extreme majority situation at

[8]This shows another advantage of margin-dependent indices: Even if the local segregation score for some small occupation is extreme because of sampling variability, its influence on total segregation will be small.

[9]The same argument could be made for the comparison of two countries with vastly different rates of female labor force participation.

the second time point, assuming that nothing else changed. However, if the changes have been as extreme as here, this assumption is likely not warranted. The deeper point here is that changing marginal distributions are an important part of segregative processes, and that the summary of a standardized segregation index should be used with caution when the margins are very different.

Appendix

Formulas for alternative segregation indices

For the alternative indices, the two-group versions are used. For some of these indices, multi-group versions have been developed as well. Let N be the number of occupations, T the total number of workers, M (W) the total number of male (female) workers, T_i the number of workers in the i-th occupation, and M_i (W_i) the number of male (female) workers in the i-th occupation. The notation follows Weeden (1998). D is the index of dissimilarity, SSD the standardized index of dissimilarity, and A the closed form of the log-linear index.

$$D = \frac{1}{2}\sum_{i=1}^{N} |\frac{W_i}{W} - \frac{M_i}{M}|$$

$$SSD = \frac{1}{2}\sum_{i=1}^{N} |\frac{W_i/T_i}{\sum_{i=1}^{N} W_i/T_i} - \frac{M_i/T_i}{\sum_{i=1}^{N} W_i/T_i}|$$

$$A = \exp\left(\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\ln\frac{W_i}{M_i} - \left(\frac{1}{N}\sum_{i=1}^{N}\ln\frac{W_i}{M_i}\right)\right]^2\right\}^{\frac{1}{2}}\right)$$

References

Agresti, Alan. 2013. *Categorical Data Analysis*. Hoboken, NJ: Wiley, third edition.

An, Brian P. and Adam Gamoran. 2009. "Trends in School Racial Composition in the Era of Unitary Status." In *From the Courtroom to the Classroom*, edited by Claire Smrekar and Ellen Goldring.

Charles, Maria and David B. Grusky. 1995. "Models for Describing the Underlying Structure of Sex Segregation." *American Journal of Sociology* 100:931–971.

Deming, W. Edwards and Frederick F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table." .

DiPrete, Thomas A., Thijs Bol, Christina Ciocca Eller, and Herman G. van de Werfhorst. 2017. "School-to-Work Linkages in the United States, Germany, and France." *American Journal of Sociology* 122:1869–1938.

Duncan, Otis Dudley and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20:210–217.

Ferguson, John-Paul and Rembrand Koning. 2018. "Firm Turnover and the Return of Racial Establishment Segregation." *American Sociological Review* 83:445–474.

Forster, Andrea G. and Thijs Bol. 2018. "Vocational Education and Employment over the Life Course Using a New Measure of Occupational Specificity." *Social Science Research* 70:176–197.

Frankel, David M. and Oscar Volij. 2011. "Measuring School Segregation." *Journal of Economic Theory* 146:1–38.

Grusky, David B. and Maria Charles. 1998. "The Past, Present, and Future of Sex Segregation Methodology." 35.

James, David R. and Karl E. Taeuber. 1985. "Measures of Segregation." *Sociological Methodology* 15:1–32.

Karmel, T. and M. Maclachlan. 1988. "Occupational Sex Segregation —Increasing or Decreasing?" *Economic Record* 64:187–195.

Massey, Douglas S. and Nancy A. Denton. 1988. "The Dimensions of Residential Segregation." *Social Forces* 67:281–315.

Mora, Ricardo and Javier Ruiz-Castillo. 2003. "Additively Decomposable Segregation Indexes. The Case of Gender Segregation by Occupations and Human Capital Levels in Spain." *The Journal of Economic Inequality* 1:147–179.

Mora, Ricardo and Javier Ruiz-Castillo. 2009. "The Invariance Properties of the Mutual Information Index of Multigroup Segregation." In *Research on Economic Inequality*, edited by Yves Flückiger, Sean F. Reardon, and Jacques Silber, volume 17, pp. 33–53. Emerald Group Publishing Limited.

Mora, Ricardo and Javier Ruiz-Castillo. 2011. "Entropy-Based Segregation Indices." *Sociological Methodology* 41:159–194.

Reardon, Sean F. and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32:33–67.

Reardon, Sean F. and Ann Owens. 2014. "60 Years After *Brown* : Trends and Consequences of School Segregation." *Annual Review of Sociology* 40:199–218.

Reardon, Sean F., John T. Yun, and Tamela McNulty Eitle. 2000. "The Changing Structure of School Segregation: Measurement and Evidence of Multiracial Metropolitan-Area School Segregation, 1989-1995." *Demography* 37:351–364.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2018. "IPUMS USA: Version 8.0." .

Theil, Henri. 1967. *Economics and Information Theory*. Amsterdam: North Holland.

Theil, Henri. 1971. *Principles of Econometrics*. New York: Wiley.

Theil, Henri. 1972. *Statistical Decomposition Analysis*. Amsterdam: North Holland.

Theil, Henri and Anthony J. Finizza. 1971. "A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts." *Journal of Mathematical Sociology* .

Watts, Martin. 1998. "The Analysis of Sex Segregation: When Is Index Measurement Not Index Measurement?" *Demography* 35:505–508.

Watts, Martin. n.d. "The Use and Abuse of Entropy-Based Segregation Indices."

Weeden, Kim A. 1998. "Revisiting Occupational Sex Segregation in the United States, 1910-1990: Results from a Log-Linear Approach." *Demography* 35:475–487.