geographical analysis

Geographical Analysis (2024) **56**, 79–96

Income Segregation Analysis in Limited-Data Contexts: A Methodology Based on Iterative Proportional Fitting

Gonzalo Peraza-Mues[1,2], Roberto Ponce-Lopez[1,2], Juan Antonio Muñoz Sanchez[3], Fernanda Cavazos Alanis[3], Grissel Olivera Martínez[3], and Carlos Brambila Paz[2]

[1]Center for the Future of Cities, Tecnologico de Monterrey, Monterrey, Mexico, [2]School of Government and Public Transformation, Tecnologico de Monterrey, Monterrey, Mexico, [3]School of Social Sciences and Government, Tecnologico de Monterrey, Monterrey, Mexico

Since the 1950s, researchers in Urban Geography have created multiple instruments for measuring income segregation. However, the computation of such indexes requires the availability of income data and population distribution for small areal units. This approach is problematic for countries and cities where a government's decennial census does not collect or report income data for small-enough areal units to capture income variability within a neighborhood. To address this gap, we use Iterative Proportional Fitting (IPF) to combine neighborhood-level census data with an individual-level income survey data and then estimate small area discrete and continuous income distributions for each small area. We show that it is possible to compute segregation indices based solely on estimated probability distributions without the need to generate a full synthetic population or to obtain integer population counts. We test our empirical method with the case of Mexican cities, for which global and local indexes of segregation are computed with bootstrapped confidence intervals. The major contributions of this article are twofold. First, it uses a method for income-data generation to measure income segregation. Secondly, it demonstrates a linkage between the computation of segregation measures based on probability distributions and the feasibility of computing them directly from the same IPF estimated distributions of income.

Introduction

Residential segregation by income can be defined as the degree of geographical separation among socioeconomic groups living in a given area (Reardon and O'Sullivan 2004). The goal of this research is to outline a methodology to estimate residential segregation based on income that can be used in cases when income data from a government census is unavailable.

Correspondence: Roberto Ponce-Lopez, Center for the Future of Cities, Tecnologico de Monterrey, Monterrey, Mexico.
e-mail: rpl@tec.mx

Submitted: January 6, 2023. Revised version accepted: May 23, 2023.

doi: 10.1111/gean.12367

Starting with the human-ecology-inspired segregation theory of the 1920s (Park et al., 1925), theories regarding the dynamics of urban life have associated segregation primarily with urban contexts. Consequently, researchers have typically addressed income segregation as exclusively a residential phenomenon (Musterd 2020). However, the topic of economic segregation is important and relevant because of the more recent focus on the potentially long-term consequences of segregation through so-called *neighborhood effects*. Specifically, research on this topic suggests that growing up in an area where poverty predominates negatively affects a resident's access to education and jobs, and thus to their adult earning potential (Van Ham et al. 2012).

Within the past century, a large body of the literature has proposed new indexes to measure economic segregation. The seminal work of Duncan and Duncan (1955) described and categorized classical indexes of segregation to measure the mixing of various groups living in the same area. In 1998, Massey and Denton proposed a typology to systematize such indexes along the spatial dimensions of evenness and the degree to which various groups are exposed to each other. Other researchers developed a new generation of indexes especially tailored to capture the ordinal characteristic of income groups (Reardon 2009; Reardon and Bischoff 2011), and to represent the spatial dimensions of income segregation (Wong 2002a, 2004a). More recently, Yao et al. (2019) summarized spatial arrangements as they related to mathematical formulations used in classical and newer versions of such indexes.

Despite the wealth of research on segregation available to date, there remains a gap in the literature. Although much has been written about new and better indexes of segregation, until now insufficient attention has been paid to the nature, quality and presentation of income data, especially at a spatial resolution approximating neighborhoods, has made it difficult to analyze segregation at an appropriate spatial scale, particularly in cities in the Global South. While income inequality measures, such as the Gini Index, can be estimated based on surveys done at the state level, segregation indexes, being spatial in nature, require a finer degree of spatial granularity.

Our research overcomes these limitations and provides such desired granularity by combining high resolution data from the census with a detailed income survey representative at the regional level, enabling researchers to represent the distribution of income groups at finer geographical scales. Specifically, the technique of Iterative Proportional Fitting (IPF) is used to estimate the income distribution in small areas. In geography and urban studies, there is a large body of literature dealing with methods used for small area estimation – being IPF one of them – and even some applications used to study social welfare and inequality, see for example Anderson (2013), Fabrizi and Trivisano (2016), Philips, Clarke, and Watling (2017), and Rose and Nagle (2017). In particular, the work by Anderson (2013) resembles our own in that IPF is also used to estimate small area income distributions. However, the approach taken by our article differs from this earlier work in that we also offer probabilistic arguments to justify the employment of IPF for income distribution and avoid the unneeded step of integerization. This is relevant as it serves as the foundation for the main contribution of the article: IPF can generate income data for small area units when not otherwise available. We show the estimation of small-area distributions from a global known distribution plus constraints. Moreover, certain indices of segregation can be calculated directly from the probability distributions estimated through IPF.

The remainder of the article is organized as follows: In Section 2, we describe the specificity of income data needed to measure segregation and the state of data collection and reporting within different countries. In Section 3, we first summarize the main indexes of segregation and their evolution over time. We then describe the use of IPF within the fields of geography, transportation and social sciences, where it has been employed mainly as a tool to build synthetic populations to compensate for scarce data. In Section 4, we explain how the IPF technique is used to estimate the income distribution and how to calculate the selected indexes of income segregation from an estimated distribution. In Section 6, we demonstrate an actual case study with results for the city of Monterrey, Mexico (results for other four cities in Mexico are available in the Data S1). Finally, in Section 7, we discuss the main methodological contributions of the article and suggest future areas of research.

Problems with scarcity of income data in countries from the global south

The lack of discussion regarding the spatial resolution of income data and its scarcity in segregation studies may originate in the fact that most research done in this area is carried out in developed countries, where data is collected and reported with a high degree of spatial granularity. For example, the Census Bureau of the United States uses tracts as approximations for neighborhoods, with associated residents within each tract described by income bracket.

Until 2000, the long questionnaire, part of the United States' decennial census, included income information. Since then, the American Community Survey (ACS) has functioned instead as the main source of income data for a 1-year and/or 5-year basis based on location. The Census Bureau reports income data from the questionnaires in two primary formats: (1) summarized as the median income by census tract and (2) aggregated by geography level for location and then reported as the number of individuals or households per income bracket. At the time of writing, the vast majority of the resulting segregation indexes have been based on households per income bracket, since diversity has been defined relative to the composition of a given household. The ACS includes 11 income brackets. We identify three main limitations with this method for measuring segregation by income.

First, countries from the Global South do not typically use instruments as complex as the ACS or the decennial census (CEPAL 2018) to report income. Consequently, economic segregation is not accurately measured in geographical areas with the highest rates of urban growth and the most serious problems of inequality. Second, even if income data are collected and reported, income thresholds differ widely across countries, which hinders comparisons of cities in different countries. Third, income thresholds used to define categories often change from one decennial census to the next, hampering comparisons across time-even for the same city and/or for countries with reliable income data.

Moreover, additional data from surveys can supplement and improve the accuracy of income measurements reported by the decennial Census. This is because people tend to provide more accurate information when they are also surveyed about other attributes closely related to income, such as access to private health care (Gaytán and Cantú 2014). Within economies shaped by less formal practices, such as the Mexican economy, individuals tend to not accurately report their income or material possessions, either because of the difficulty of identifying revenue from capital investment, or because informal income leads them to underestimate the monetary value of their work (Aguayo-Téllez 2018). Formal salary represents as little as 30% of all Mexican income, a small percent when international standards are considered (Gaytán and Cantú 2014). Estimates that include bias correction suggest that salary remunerations represent only 42–47% of total income (Gaytán and Cantú 2014). Likewise, see Logan et al. (2018) for a discussion about the reliability of income measurement in the ACS for the American context.

Literature review

This section reviews the most important work in two academic fields with regard to the main contribution of this article: measurement of income segregation and applications of IPF for public policy.

Indexes to measure income segregation

In the past 70 years, the literature of segregation studies has reported an evolution toward more complex and sophisticated techniques to measure the mixing of groups in space.

Indexes of group segregation can be global or local. Global indexes use income distribution from census tracts to derive a single indicator that measures the level of segregation for a specific spatial area. These indexes are especially helpful when comparing different cities or the same city over time. Global measures are the most common type of index in the literature and include the widely used Dissimilarity Index and the Information Entropy Index (see Massey and Denton 1988 for a discussion on classical indexes of segregation). More recent global indexes address the Modifiable Areal Unit Problem (MAUP) (Stewart Fotheringham and Wong 1991; Wong 2004b) and the spatial composition of group membership (Reardon and O'Sullivan 2004; Jargowsky and Kim 2009). In the 2000s, Reardon and Wong proposed global indexes that standardize the aerial unit and define the ordinal structure of income reported as the number of people within each category (Wong 2003, 2004a; Reardon 2009, 2011). In this work, we adopt the Rank Order Information Theory Index (Reardon and Bischoff 2011) as a global index of segregation.

In contrast with global indexes, local indexes compute a measurement of segregation for each census tract, relative to other tracts in the same region. They map segregation within a given city and help identify areas with high levels of segregation. Local measures of segregation fall into two groups: Local Indicators of Spatial Association (Anselin 1988) and Spatial Diversity Indexes (Wong 2002b). The Local Indicators of Spatial Association (LISA) quantify the local correlation between a continuous variable, such as median income or percentage of population in a certain income bracket, measured in a given unit of analysis and its spatial neighbors. The primary criticism of LISA is its vulnerability to the MAUP (Wong 2002b). A second limitation is that LISA does not account for income distribution by groups within the areal unit of analysis. In an attempt to solve these problems, Wong (2002b) developed the Spatial Diversity Index, a spatially modified version of the Entropy Index. Wong's local segregation index solved the MAUP and was able to accommodate multiple groups. The limitation of this index was that it did not address the ordinal structure of the groups, a shortcoming when specifying income segregation data type. In this article, we use the Local Centralization Index (Folch and Rey 2016), a spatial version of the Gini index, to measure the local segregation among income brackets or groups.

IPF for small area estimates

In the 1980s and early 1990s, researchers explored the use of IPF for small area estimation for geographical applications. Johnston, Hay, and Patty (Johnston and Hay 1982, 1983; Johnston and Pattie 1993) demonstrated the usage of entropy-maximizing methods based on the IPF procedure in order to estimate flows of voters across constituencies. Wong (1992) showed the utility of this technique in separating disaggregated spatial data from aggregated data.

Since the late 1990s, IPF has been broadly used within the computational social sciences for small-area estimates having applications for public policy when geographical referenced data is not available. Clarke et al. (1984) developed one of the first spatial micro-simulation models for the British Health district authorities. Called the Health Information and Planning System (HIPS), this model generated a synthetic population for each area contributing to urban and health planning systems.

Similar to the HIPS, Birkin and Clarke (1988) developed a numerical process that relied on IPF to generate a population with high spatial resolution. Known simply as *synthesis*, this methodology has been used in the past to evaluate the impact of public policies; in general (Avram and Dronkers 2012), with regard to applications in transport systems (e.g., Barthélemy and Toint 2015), and of initiatives to address specific health issues such as obesity (Edwards et al. 2011). See Smith, Pearce, and Harland (2011) for an application of the technique to predict smoking prevalence within small area units, Panori, Ballas, and Psycharis (2017) for the use of IPF to generate small-area income distributions, and Kavroudakis, Ballas, and Birkin (2013) for the use of IPF in analyzing likely education attainment. See Anderson (2013) for an application of IPF to estimate small area income deprivation in Wales. See Lin and Xiao (2023) for recent applications of small-area public synthetic microdata for public use.

IPF has been applied to numerous research areas that use contingency tables. This combines the method with statistical theory, which provides a solid foundation for quantitative and qualitative analysis. Some of the most important contributions in the development of the method have been those of: Kruithof (1937) for his double factor method; Stephan, Edwards Deming, and Hansen (1940) for the development of a new algorithm that allowed simultaneous enumerations and sampling; Fratar, Voorhees, and Raff (1954) for his method of successive approximations; Brown (1959) for his iterative scale procedure; Stone and Brown (1962) for the "R.A.S." algorithm; Darroch (1962) for the iterative scaling method; Birch (1963) for its method of maximum likelihood in three-way contingency tables; Furness (1965) for the Furness Iteration Procedure; Mosteller (1968) for his contribution to association and estimation in contingency tables; Bishop, Fienberg, and Holland (1975) for their contribution in discrete multivariate analysis.

In summary, a rich body of literature in segregation indexes has measured the various dimensions of this social phenomenon. The literature has evolved from simpler indexes to more complex and sophisticated measurements able to deal with MAUP, to make measurements of local segregation, and to handle ordinal categories rather than only discrete ones. However, these indexes are inflexible because if income data are not available as required, then segregation cannot be calculated. One limitation to increasing the usage of these indexes, specially in developing countries, is that input data with sufficient spatial granularity are not always available. Fortunately, IPF is a method with a long tradition in demography and geographical science that can solve this problem.

Methodology

Our methodology uses a two-step process. First, we estimate the local income distribution of individuals within each census tract from its joint distribution with a selection of socioeconomic variables. Then, we assess the degree of income segregation based on the estimated income distributions from the first step. The first step allows us to directly estimate the marginal distributions of discrete variables. To estimate the distribution of continuous variables, we then apply a weighting scheme to the survey population. To illustrate the use of these two approaches, using discrete or continuous variables, we tested the process in our empirical case using income categories as a discrete variable and individuals' income as a continuous variable. The following subsections describe the method.

Step 1: estimating local distributions

The first step combines two sources of data: (1) a survey of N individuals that collects and reports the target variable (e.g., income) but is not, due to the small sample size and sampling procedure, statistically representative of specific census tracts; and (2) a decennial census containing information at the desired level of aggregation (i.e., census tracts) but without information regarding the target variable. These two data sources each contain a set of K variables, $X_1, \dots, X_K$, being considered the linking or constraint variables, while the target variable appears only in the survey. The objective of this step is to estimate the statistical distribution of the target variable for each census tract. Our method allows for the possibility of a discrete target Y or a continuous target y. In the example provided here, the survey records income as a continuous variable: the actual reported income of individuals (y). To demonstrate the use of discrete variables, we break the target variable down into discrete categories using quantiles. The resulting discrete variable, having five categories, is denoted as Y.

The constraining variables X_k are social markers related to the target. For the variable income as the target, we use: sex, age range (reported as a population brackets), level of education, access to public health care, access to private health care, and home internet access. We tested different sets of constraining variables, and these variables were found to be best at reproducing an areal measure of welfare according to our local knowledge of the area. Since IPF works directly on contingency tables (as representations of a multinomial distribution), both the constraints and the target must be discrete variables. This condition requires a number of preliminary steps to recode and discretize the whole set of variables into a schema used in both the survey and the census. For example, encoding age in the survey into the age brackets reported in the census.

We begin by estimating a global joint distribution for the constraint and target variables by building a multiway contingency table C from the survey. The entries in the contingency table are the Maximum Likelihood Estimator of the implicit multinomial distribution of the variables. The table C is K-dimensional, with $M = \prod_{k=1}^{K} |X_k|$ entries, where $|X_k|$ is the number of possible categories of variable X_k. Each entry in the table is the estimated joint probability, $P(\vec{X}, Y)$, of possible variable combinations, with $\vec{X}$ denoting the vector of constraint variables.

Next, we localize the global distribution C onto each census tract j using the IPF algorithm, thereby creating a set of local contingency tables C_j, one for each census tract. Each C_j contains the full multinomial model of each tract, including the discrete target Y (income group). Specifically, the IPF algorithm solves the following optimization problem: to find the closest distribution C_j to C, as measured by the Kullback–Leibler (KL) divergence, subject to the constraints of the known population counts in each tract. These constraints, or population counts, are the observed marginal distributions of the linking variables $P(X_k|j)$ for each census tract (as given by the census), and must be preserved within the estimated local distributions. In other words, IPF minimizes the KL divergence between the global and the local distribution (Terrance Ireland and Kullback 1968), where the divergence is used as a measure of dissimilarity between probability distributions (Cover and Thomas 2006). The KL divergence also plays a role in calculating segregation indexes. A detailed description of the IPF algorithm, with historical remarks, can be found in Založnik (2011).

To calculate segregation coefficients, the marginal proportions of every income group in each tract are required. These proportions can be obtained from the C_j, marginalizing over the constraint variables to obtain the proportions, or marginal probabilities, $P(Y|j)$ in each tract j. This step finishes by estimating the discrete target variable Y.

For the continuous variable y, we estimate the cumulative distribution function $F(y|j)$ for each census tract j, derived from the y values in the survey and the local distributions C_j (estimated using IPF), as

$$\begin{aligned} F(y|j) &= \int_{y' \le y} f(y'|j) dy' \\ &= \int_{y' \le y} \left(\sum_m f(y'|\vec{X}_m) P(\vec{X}_m|j) \right) dy' \\ &\approx \sum_{y' \le y} \sum_m \frac{N_{y'|m}}{N_m} P(\vec{X}_m|j), \end{aligned} \tag{1}$$

where $P(\vec{X}_m|j)$ are the M entries of C_j indexed by m, each giving the joint probability of a particular combination of constraint variables, collected in the vector $\vec{X}_m$. The set of cumulative distribution functions use the conditional distributions obtained from IPF in the previous step, and are thus local estimates for each census tract. We approximate the integral over y with a discrete sum over all values of y in the survey, and the conditional density function as $f(y'|\vec{X}_m)dy' \approx \frac{N_{y'|m}}{N_m}$, where N_m is the number of individuals with constraint vector $\vec{X}_m$ in the survey, and $N_{y'|m}$ is the number of individuals with income y' and constraint vector $\vec{X}_m$.

The quantity $w_{ij} = \frac{P(\vec{X}_m|j)}{N_m}$ is referred to as the weight of individual i in census tract j and measures the degree of statistical representativeness of each individual i in tract j (Lovelace et al. 2015). When multiplied by the population of the tract, N_j, the weights yield the number of times an individual must be replicated at each tract. After weighting and replication, the final output is a synthetic population. If, for example, a synthetic population is needed to perform an agent-based simulation, this weight needs to be converted into an integer (see Lovelace and Ballas 2013 for a discussion of different integerization strategies). As shown in this article, for the calculation of segregation indexes only the estimated distributions are needed and integerization is not performed. This is important, as integerization introduces biases in the estimated distributions.

Step 2: Computing measures of segregation

The method described above enables us to estimate the distribution of any discrete or continuous variable for smaller areal units such as census tracts. In this subsection, we describe how to use these estimated distributions to calculate two indexes of income segregation, a local and a global one.

A local index of segregation

Consider the discrete variable Y first, with Y indicating different income groups obtained by discretizing income values in the survey using quantiles. We calculate the degree of segregation among the income groups using the Local Centralization Index proposed by Folch and Rey (2016). This index is a local measure of segregation that considers segregation among a particular binary grouping of the categories of Y. We focus on measuring segregation for the two most extreme binary groupings, the top/bottom income group as compared to the rest of the groups taken as one.

For each census tract, the Local Centrality Index is defined as the spatial Gini index between two population groups in a region adjacent to each census tract. Let the population of individuals with $Y = q$ in tract j be $n_{q,j}$. Consider the list of K nearest neighbors to tract j, indexed by $k = 1, \ldots, K$, with k increasing with distance from j, and $k = 0$ denoting tract j itself. This list of neighbors defines a local region around j, R_K, with total regional population for group q given by $N_{q,K} = \sum_{k=0}^{K} n_{q,k}$, and subregional populations up to neighbor k given by $N_{q,k} = \sum_{k'=0}^{k} n_{q,k'}$. The local centrality index for region R_K between two groups q and l, as given in Folch and Rey (2016), is

$$CI_K = \sum_{k=1}^{K} \left(\frac{N_{q,k-1}}{N_{q,K}} \frac{N_{l,k}}{N_{l,K}} - \frac{N_{q,k}}{N_{q,K}} \frac{N_{l,k-1}}{N_{l,K}} \right). \tag{2}$$

The local counts can be approximated from our estimated distributions as $n_{q,j} \approx P(Y = q|j)N_j$, where N_j is the total population of census tract j. Again, there is no need to transform the counts into integers, since the index is calculated using proportions. This can be stated explicitly by noting that the ratios in equation (2) can be expressed in terms of probabilities as

$$\frac{N_{q,k-1}}{N_{q,K}} = \frac{N_{q,k-1}/N}{N_{q,K}/N} = \frac{\sum_{j \in R_k} P(Y_q|j)P(j)}{\sum_{j \in R_K} P(Y_q|j)P(j)}, \tag{3}$$

where $P(j) = N_j/N$ is the probability that an individual is randomly chosen from tract j. Equation (3) shows that the centrality index can be obtained directly from the entries of the estimated contingency tables obtained from IPF.

The index is bounded within a range from -1 to 1. Negative values indicate the centralization of group l, while positive values indicate the centralization of group q. The extreme value of 1 means group q is included in the central tract with the whole population of group l in the surrounding tracts, and vice versa for the opposite extreme at -1.

A global index of segregation

We use the estimated distribution function for the continuous income $F(y)$ to calculate the Rank Order Information Theory Index (Reardon and Bischoff 2011), a global segregation index appropriate for the distribution of continuous variables such as income. The index ranges from 0 to 1, with 0 indicating no segregation, and 1 indicating full segregation. The index is calculated by considering all possible partitions of the population into two groups, a low income group with $y \leq y_p$ and a high income group with $y > y_p$, with the quantile y_p defined by the probabilities $P(y \leq y_p) = p$ and $P(y > y_p) = 1 - p$. This partition defines a binary random variable Y_p, that indicates the income group of an individual chosen from the population at random. The index is calculated according to the expression:

$$H^R = 2 \ln 2 \int_0^1 E(Y_p)H(Y_p)dp, \tag{4}$$

where $E(Y_p) = -\left[p\log_2 p + (1-p)\log_2(1-p)\right]$ is the binary entropy of the resulting global partition, and $H(Y_p)$ is the pairwise Entropy Index or Theil Index (Reardon and O'Sullivan 2004). H^R is a weighted average of the index $H(Y_p)$ for all possible binary partitions of the global population, the weights being the entropy $E(Y_p)$.

The index $H(Y_p)$ is a binary segregation index, which can be understood as a normalized weighted average of the deviations of each tract's entropy $E(Y_p|J = j)$ from the global entropy

$E(Y_p)$, where J is the random variable referring to the probability that a randomly chosen individual belongs to tract j.

$$H(Y_p) = \sum_j \frac{N_j}{N} \frac{E(Y_p) - E(Y_p|J = j)}{E(Y_p)}. \tag{5}$$

The index can also be expressed in terms of the mutual information between the random variables Y_p and J, $I(Y_p, J)$, as

$$\begin{aligned} H(Y_p) &= \frac{1}{E(Y_p)} \left(E(Y_p) - \sum_j \frac{N_j}{N} E(Y_p|J = j) \right) \\ &= \frac{1}{E(Y_p)} \left(E(Y_p) - E(Y_p|J) \right) = \frac{I(Y_p, J)}{E(Y_p)}, \end{aligned} \tag{6}$$

where $E(Y_p|J)$ is the conditional entropy. Finally, the shared information is related to the expected value of the KL divergence by $I(Y_p, J) = E\left[D_{\mathrm{KL}}\left(Y_p|J \| Y_p\right)\right]$ (Cover and Thomas 2006), allows us to write H^R as a double average of KL divergence: an average based on census tracts, and an average over pair-wise partitions of the income distribution,

$$H^R = 2\ln 2 \int_0^1 E\left[D_{\mathrm{KL}}\left(Y_p|J \| Y_p\right)\right] dp, \tag{7}$$

where $2\ln 2$ is a normalization constant.

As mentioned above, the KL divergence measures the degree of dissimilarity of two distributions: in this case, the global distribution of the whole region (estimated from the survey) and the local distribution (based on each census tract). The resulting dissimilarity is interpreted as a measurement of segregation for the specific tract. Notice that equation (7) explicitly relates segregation to these differences between the local (for each tract) and global (city-wide) distributions of income (through the KL divergence). However, IPF minimizes this same divergence since IPF assumes similar distributions everywhere, thus potentially minimizing local hotspots. This introduces a potential bias, which can be expected with any estimation that tries to match global and local distributions, such as IPF. This effect pushes the segregation index toward smaller values, but it also causes segregation values that are statistically significant to match the areas with the highest degree of segregation. In other words, the combination of IPF and the Segregation Index change the scale of magnitude for segregation, making these values smaller. The values are still comparable across cities and useful to identify the most segregated regions.

Equation (7) also suggests a possible modification in how the segregation is calculated. Since our method creates an estimate of the full income distribution, rather than an average of all possible binary partitions of income, we could instead use the KL divergence for continuous distributions and calculate a modified index of segregation as:

$$H^R_{cont} = A\, E\left[D_{\mathrm{KL}}\left(y|J \| y\right)\right], \tag{8}$$

with A a suitable normalization constant to keep the index in the 0–1 range. We also leave the exploration of the properties of H^R_{cont} for future work.

To compute the integral in equation (4), the reference implementation fits a polynomial to calculated values $H_i(Y_p)$ at a set of reported cut points y_i of income, where y_i defines the intervals of the census income brackets. Here, since we have an estimate of the full income distribution, we can avoid the fitting step and instead approximate the integral directly from the estimated distribution, thereby preventing potential biases caused by imposing a functional form to $H(Y_p)$. For all reported values of income in the survey, y_i, we obtain the global probabilities $p_i = F(y_i)$ and the local probabilities $p_{ij} = F(y_i|j)$ for all tracts. With these probabilities, we then obtain discrete samples of the index, namely values $H_i(Y_p)$. If the survey is large enough, this produces a good approximation of the continuous function $H(Y_p)$, and we can then evaluate the integral in equation (4) numerically, for example, using Simpson's rule, as demonstrated in section 6.

Finally, for purposes of validation and statistical significance, any test appropriate for spatial information can be used, such as bootstrap resampling (DiCiccio and Efron 1996) or a permutation test (e.g., Rey, Murray, and Anselin 2011). In section 6, we use bootstrap resampling on the survey to calculate confidence intervals for segregation indexes, and reject those indexes that contain zero as statistically insignificant.

In summary, our methodology combines an income survey with census data to estimate local distributions using a technique based on IPF. The survey's finely grained distribution of income enables us to compute local and global segregation indexes for countries for which data on aggregated income levels are not available. Moreover, the Rank Index of segregation shares a statistical commonality with IPF. In section 5, to show how the method works in practice, we apply the process described in this section to five Mexican cities.

Empirical case

Data

The data used in our study is available from Mexican National Institute of Statistics and Geography (designated by the acronym INEGI in Spanish). We use two of its data sources: (1) The National Survey of Household Income and Expenditure (designated by the acronym ENIGH in Spanish); and (2) aggregated census data providing statistics for the constraint variables at the level of census tracts (AGEB in Spanish).

Income survey

The ENIGH is the most reliable instrument in Mexico for representing income data, since it reports both formal work and from other sources. Because income information from the 2020 survey was limited in the COVID-19 pandemic study, in this study we instead used information from the 2018 survey, which consists of 269,206 questionnaires from 73,354 households, representing all of Mexico's 32 states. Although the ENIGH sample size is too small to be seen as representative at the census tract level required for analysis of income segregation, the survey collects other relevant sociodemographic variables, which, unlike income, are also collected in the decennial census at the tract level, allowing them to be used as linking variables in our study.

Census data

INEGI performs decennial censuses at the geographical levels of state, municipality, locality, and tract. Since 2010, they have included individual city blocks, which provides sociodemographic information about the individuals, attributes of the dwelling units and household characteristics.

Until 2000, there was a single, long questionnaire that collected income information with statistical representation of the tracts or AGEBs. Income was reported the same way as in the United States, but with *thresholds based on minimum wage standards*: no minimum wage, less than 1 minimum wage, 1–2 minimum wages, 2–5, 5–10, and 10-or-more minimum wages. It was possible to measure segregation from the recorded income brackets, though the *times the minimum wage* standard hindered comparisons between locations within Mexico and the United States (Monkkonen, 2011). However, since 2010, the census has used a long and short version of the questionnaire: the comparison of segregation has no longer been possible using census data alone since income information is collected through the long questionnaire in only a sample of households presumed to be statistically representative of their municipality.

Among 222 variables from the census and 770 variables from the ENIGH, we selected a set of six variables that were both simultaneously available in both data sources and that correlated highly with income: sex, age, education level, access to public health care, access to private health care, and internet access. It is noted that, the Mexican National Commission of Social Policy uses these same six variables to create indexes of well-being and poverty (CONEVAL 2015). The set of linking variables is intentionally small, as a large set of linking variables creates a prior contingency matrix C of high dimensionality, which demands larger sample sizes to accurately estimate its entries. Other variable selection schemes are possible, such as searching for a set of variables maximizing their mutual information with income while minimizing the mutual information among them (Peng, Long, and Ding 2005), or that maximize the variance on the census data (Lu et al. 2007), among others (Li et al. 2017). Evaluating the schemes for selecting the best features has been deferred to a future study.

Results

We tested the method in five Mexican cities representing various regions and urban morphology in the country. Figures in the section refer to the city of Monterrey, similar figures for other cities can be found in the supplementary material.

In Figure 1, we show the estimated distribution of income as a continuous variable y for Monterrey. The inset shows, in black, the ECDF for income, $F(y)$, obtained from the survey. In gray, we show the ECDFs for each tract, $F(y|j)$, obtained from equation (2).

The main axis of Figure 1 shows the economic-segregation profile $H(p)$, estimated from sample points $H_i(Y_p)$ for all possible binary splits at income quantile p_i in the survey. Larger values of H are found for large values of p, indicating a high level of segregation between higher earners and the rest of the population, a trend that had previously been observed in the United States (Logan et al. 2020). Also in Figure 1, in gray, we plot the expected KL divergence $E[D_{\mathrm{KL}}(Y_p|J||Y_p)]$, the integrand in equation (7), which is used to calculate the global index H^R. It is worth noting that the noise at the tails of the segregation profile $H(p)$, caused by small-size groups with extreme values of p, is attenuated in $E[D_{\mathrm{KL}}(Y_p|J||Y_p)]$. The maximum of $E[D_{\mathrm{KL}}(Y_p|J||Y_p)]$ is shifted to the right due to the more pronounced segregation within the top quantiles. Based on the estimated distributions, any statistic, such as the mean income for each census tract, can be calculated. A map showing mean income per tract is included in the supplementary material.

Using 10,000 bootstrap samples from the income generated data, we computed both global and local indexes of segregation with 95% bootstrap confidence intervals. The global index of segregation for Monterrey is 0.01634 with a confidence interval of (0.01284–0.02012). See Data S1 for a comparison of the segregation index with that of other four cities.

Figure 2 shows the local centrality index of each tract for the lowest ($Q = 1$) and highest ($Q = 5$) income quantiles at two-neighborhood scales of $K = 5$ and $K = 100$ neighbors. Tracts

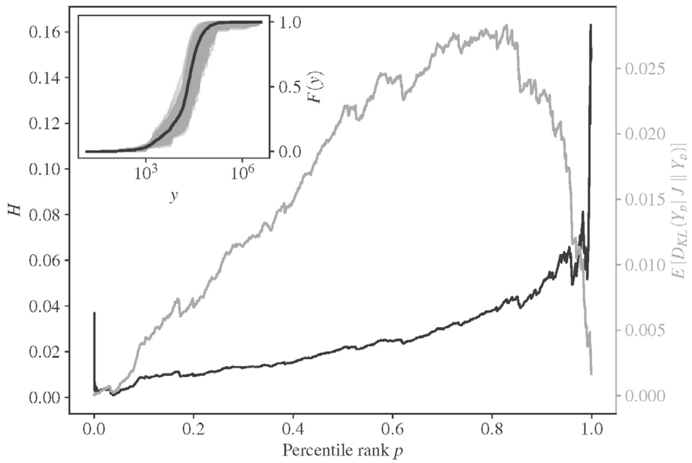

FIGURE 1. Functions involved in the estimation of continues distributions. The binary entropy index H (black) and the expected Kullback–Leibler divergence (gray) in the main axis. The inset shows the Empirical Cumulative Distribution Function (ECDF) $F(y)$ from the survey (black) as well as each tract's estimated ECDF (gray).

in gray do not have an estimated income distribution because the census does not report data for small populations due to either privacy concerns or because the tract includes only industrial or commercial uses. Therefore, such tracts do not possess the marginal constraints required for IPF and are excluded from the analysis. While the distance between tracts varies due to the unique size and shape of each geographical unit, these two scales roughly correspond to the area of a local neighborhood ($K = 5$) and to a municipal area ($K = 100$). For instance, the average distance to the fifth and the hundredth neighbor tracts in Monterrey are approximately 1 and 6 km, respectively. In general, any value of K can be used, allowing the exploration of segregation at different scales. The geography of the terrain determines the urban shape of Monterrey. For example, the Huajuco Canyon and the Mitras Mountain Range create the elongated shape of the built area in the south and cause the gaps in contiguity for this functional area, as can be observed in the hillshade of the map.

The centrality index (shown in Figure 2) reflects group concentrations for a particular tract relative to that of the tract's closest K neighbors. A positive index value (blue color) means that the specific income group being considered ($Q = 1$ for the lowest income group and $Q = 5$ for the highest income group) in a given tract is surrounded by people of different groups (others) in the K definition of neighborhood. A negative index value (red color) indicates that people of other income groups also live there, and that the tract is surrounded by people of the reference group. For $Q = 1$, blue identifies tracts having a large concentration of the lowest income group, as compared to their surrounding K neighbors; red tracts indicate a concentration of others (groups Q = 2,3 and 4) with predominantly group 1 in the neighborhood tracts. The case of $Q = 5$ designates a concentration of the highest income group. Blue indicates a tract with a large concentration of high-income residents, while a red a tract indicates that a high-income group is included in the neighborhood but absent from that particular tract. At the regional scale, it is easier to identify larger clusters of spatially correlated indexes for a neighborhood of $K = 100$. For $K = 5$, the centrality index reflects income concentration relative to that of each tract's

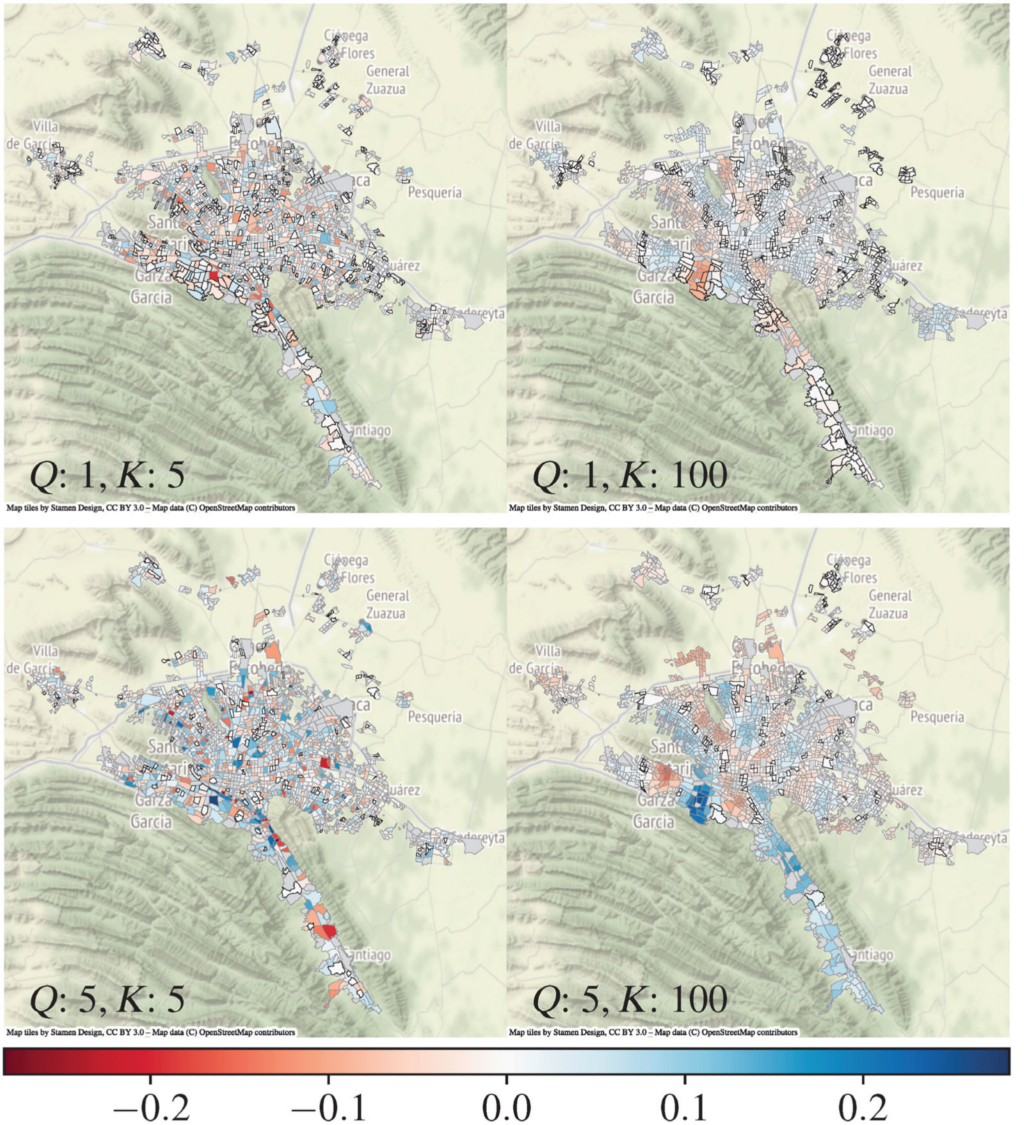

FIGURE 2. Local indexes of segregation for the city of Monterrey for lowest ($Q = 1$) and highest ($Q = 5$) income group using five ($K = 5$) and a hundred ($K = 100$) closest neighbors. Tracts with black borders do not have a statistically significant index using 95% bootstrapped confidence intervals. Tracts in gray do not have population counts included in the Census due to either privacy concerns or because they consist mainly of land with industrial/commercial uses and therefore are unpopulated.

immediate neighbors; the segregation measure is more local and is useful for identifying great income disparities among neighboring tracts.

In Figure 2, the tracts delineated in black are those whose centrality index is not statistically significant, that is, those where the bootstrap confidence interval contains zero. The entire set of indexes and confidence intervals for all cases shown in Figure 2 are plotted in Figure 3, with

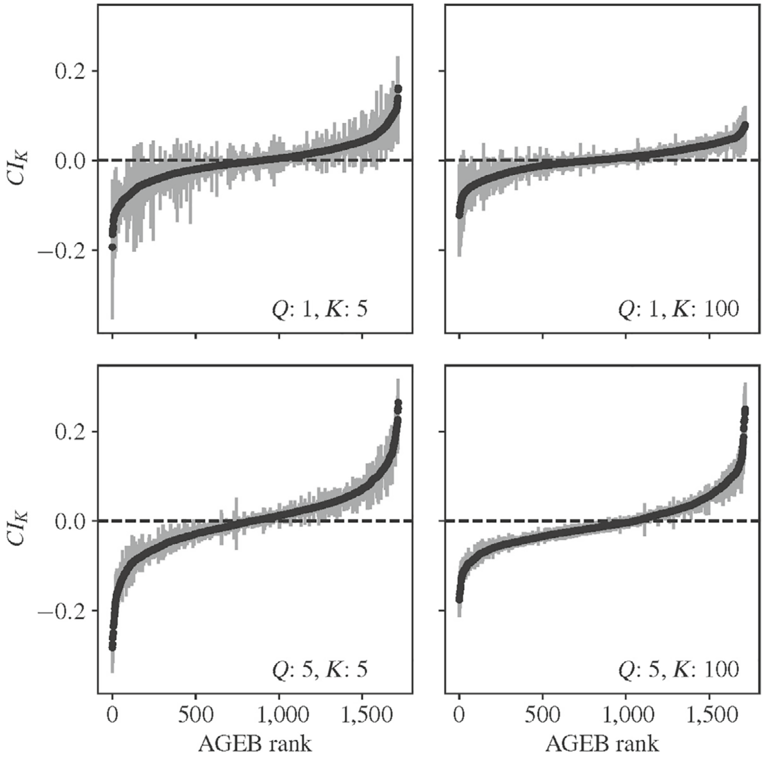

FIGURE 3. Confidence intervals for local indexes for lowest ($Q = 1$) and highest ($Q = 5$) income group using five ($K = 5$) and a hundred ($K = 100$) closest neighbors. AGEBs are sorted by index value. Note that the confidence intervals are shorter when computing the indexes for the *wealthiest* group.

tracts sorted by their index value. Indexes are generally larger for the fifth quantile ($Q = 5$), indicating greater segregation for the highest income group. Also, confidence intervals are wider for the first quantile ($Q = 1$), with a larger proportion of nonsignificant indexes.

Conclusion and future work

In this article, we proposed using an adaptation of the IPF method that combines an income survey with census data to estimate local distributions of discrete or continuous variables of interest.

Although the segregation indexes computed from the IPF estimates may be biased toward lower values, the main advantage of our technique is that the resulting indices can be compared and are useful for: identifying the most segregated regions, grouping the population into low and high income categories, and capturing differences at the local level (by tract) and global level (by city). To empirically test the method, we estimated local income distributions and calculated global and local segregation indexes for five major cities in Mexico and then presented the results for Monterrey. We estimated the distribution of the variable income by tract and then computed both local and global residential income segregation indexes. We estimated confidence intervals using bootstrap to validate the statistical significance for segregation.

Mexico was chosen for this case study because, unlike most developed countries, it lacks reliable statistics regarding residential segregation. This is partly because the information on income level that is regularly reported in the income and expenditure survey lacks the geographic specificity required to calculate appropriate segregation metrics by tract. Moreover, the census does not report income categories by population composition at the tract level.

The main empirical results in our study of Monterrey indicate that there are more high-income groups distributed throughout the city than lower-income groups. The spatial distribution of each group is heterogeneous, and the concentrations of similar groups become more visible as the number of close neighbors increases. In other words, we find that segregation is more pronounced in the higher income quantiles. These findings are consistent with the estimates from the centrality index of each tract for the first and fifth income quantiles, which, when considering estimates for 5 and 100 nearest neighbors, again show a greater level of segregation in the higher income groups. Segregation measurements for the other four Mexican cities are available in the Data S1.

The article also discusses that the estimation of distributions introduces a bias into segregation studies, a relationship made visible through the KL divergence. This bias does not occur solely with the use of IPF, but rather it is common to all methods that attempt to match local distributions to a global distribution, thus, reducing potentially important differences between small area estimates. For future work, we envisage improving our application by using different prior distributions for different regions of the zone of interest, the regions being chosen by either local experts who have fieldwork experience in the region or different sources of information reflecting the local distribution of income, such as real estate data. Although the IPF method has limitations, our proposed methodology is generally applicable and could be used to estimate distributions for variables other than income. The estimated distributions could then be used to calculate any desired statistic, not only segregation indexes. We believe that this method and its possible adaptations will prove useful in complementing regional analysis, especially in countries from the Global South where data is rarely reported at high spatial resolution.

Code and data used to generate the results presented in section 6, including a complete implementation of the methodology, are available online at https://github.com/CentroFuturoCiudades/income-segregation.

Acknowledgements

This research was funded by a FEMSA Foundation grant to Tecnologico de Monterrey to support the research project entitled "A Metropolitan Information System for Monterrey." This research was also supported by Tecnologico de Monterrey through the grant E025-ECSG-GI04-B-T4-E to develop an IPF method to generate small area detailed data for Monterrey.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Aguayo-Téllez, E. (2018). *Valoración del Trabajo Productivo Por género y Región, Chapter 1*, 1–14, edited by J. Chapa and E. Ayala. Mexico: Pearson.

Anderson, B. (2013). "Estimating Small-Area Income Deprivation: An Iterative Proportional Fitting Approach." In *Spatial Microsimulation: A Reference Guide for Users*, 49–67, edited by R. Tanton and K. Edwards. Dordrecht: Springer.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Vol 4. Dordrecht: Springer Science & Business Media.

Avram, S., and J. Dronkers. (2012). “Social Class Dimensions in the Selection of Private Schools a Cross-National Analysis Using Pisa.” In *Private Schulen in Deutschland*, 201–23. Wiesbaden: Springer.

Barthélemy, J., and P. Toint. (2015). “A Stochastic and Flexible Activity Based Model for Large Population. Application to Belgium.” *Journal of Artificial Societies and Social Simulation* 18(6), 15–35. https://doi.org/10.18564/jasss.2819

Birch, M. W. (1963). “Maximum Likelihood in Three-Way Contingency Tables.” *Journal of the Royal Statistical Society: Series B (Methodological)* 25(1), 220–33.

Birkin, M., and M. Clarke. (1988). “Synthesis – A Synthetic Spatial Information System for Urban and Regional Analysis: Methods and Examples.” *Environment and Planning A* 20(12), 1645–71.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland. (1975). *Discrete Multivariate Analysis: theorY and Practice*. Cambridge: MIT Press.

Brown, D. T. (1959). “A Note on Approximations to Discrete Probability Distributions.” *Information and Control* 2(4), 386–92.

CEPAL. (2018). *Medición de la Pobreza Por Ingresos: Actualización Metodológica y Resultados, Metodologías de la CEPAL, No 2*. Santiago: CEPAL.

Clarke, M., P. Forte, M. Spowage, and A. G. Wilson. (1984). “A Strategic Planning Simulation Model of a District Health Service System: The in-Patient Component and Results.” In *Third International Conference on System Science in Health Care*, edited by W. Eimeren, R. Engelbrecht, and C. D. Flagle, 949–54. Berlin: Springer Berlin Heidelberg.

CONEVAL. (2015). “*Medicion de la Pobreza Evolucion de Las Carencias Sociales 2015*.” https://www.coneval.org.mx/Medicion/EDP/Paginas/Datos-del-Modulo-de-Condiciones-Socioeconomicas.aspx.

Cover, T. M., and J. A. Thomas. (2006). *Elements of Information Theory*, 2nd edn. New Jersey: Wiley-Interscience.

Darroch, J. N. (1962). “Interactions in Multi-Factor Contingency Tables.” *Journal of the Royal Statistical Society: Series B (Methodological)* 24(1), 251–63.

DiCiccio, T. J., and B. Efron. (1996). “Bootstrap Confidence Intervals.” *Statistical Science* 11(3), 189–228.

Duncan, O. D., and B. Duncan. (1955). “A Methodological Analysis of Segregation Indexes.” *American Sociological Review* 20(2), 210–7.

Edwards, K. L., G. P. Clarke, J. K. Ransley, and J. E. Cade. (2011). “Serial Cross-Sectional Analysis of Prevalence of Overweight and Obese Children between 1998 and 2003 in Leeds, Uk, Using Routinely Measured Data.” *Public Health Nutrition* 14(1), 56–61.

Fabrizi, E., and C. Trivisano. (2016). “Small Area Estimation of the Gini Concentration Coefficient.” *Computational Statistics & Data Analysis* 99, 223–34.

Folch, D. C., and S. J. Rey. (2016). “The Centralization Index: A Measure of Local Spatial Segregation.” *Papers in Regional Science* 95(3), 555–76.

Fratar, T. J., A. M. Voorhees, and M. S. Raff. (1954). “Forecasting Distribution of Interzonal Vehicular Trips by Successive Approximations/with Discussion.” In *Proceedings of the Thirty-Third Annual Meeting of the Highway Research Board* Vol 33. Washington, D.C.: National Academy of Sciences.

Furness, K. P. (1965). “Time Function Iteration.” *Traffic Engineering and Control* 7(7), 458–60.

Gaytán, E. A. A., and J. C. C. Cantú. (2014). “Dónde Quedó el Trabajo? Una Propuesta Para Medir la Participación del Ingreso Laboral en méxico.” *El trimestre Económico* 81(324), 989–1015.

Jargowsky, P. A., and J. Kim. (2009). “The Information Theory of Segregation: Uniting Segregation and Inequality in a Common Framework.” In *Occupational and Residential Segregation*, edited by Y. Flückiger, S.F. Reardon, and J. Silber, 3–31. Bingley: Emerald Group Publishing Limited.

Johnston, R. J., and A. M. Hay. (1982). “On the Parameters of Uniform Swing in Single-Member Constituency Electoral Systems.” *Environment and Planning A* 14(1), 61–74.

Johnston, R. J., and A. M. Hay. (1983). “Voter Transition Probability Estimates: An Entropy-Maximizing Approach.” *European Journal of Political Research* 11(1), 93–8.

Johnston, R. J., and C. J. Pattie. (1993). “Entropy-Maximizing and the Iterative Proportional Fitting Procedure.” *The Professional Geographer* 45(3), 317–22.

Kavroudakis, D., D. Ballas, and M. Birkin. (2013). "Using Spatial Microsimulation to Model Social and Spatial Inequalities in Educational Attainment." *Applied Spatial Analysis and Policy* 6(1), 1–23.

Kruithof, J. (1937). "Telefoonverkeersrekening." *De Ingenieur* 52, 15–25.

Li, J., K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu. (2017). "Feature Selection: A Data Perspective." *ACM Computing Surveys (CSUR)* 50(6), 1–45.

Lin, Y., and N. Xiao. (2023). "Generating Small Areal Synthetic Microdata from Public Aggregated Data Using an Optimization Method." *The Professional Geographer*, 1–11. https://doi.org/10.1080/00330124.2023.2207640

Logan, J. R., A. Foster, J. Ke, and F. Li. (2018). "The Uptick in Income Segregation: Real Trend or Random Sampling Variation?" *American Journal of Sociology* 124(1), 185–222.

Logan, J. R., A. Foster, X. Hongwei, and W. Zhang. (2020). "Income Segregation: Up or Down, and for Whom?" *Demography* 57(5), 1951–74.

Lovelace, R., and D. Ballas. (2013). "'Truncate, Replicate, Sample': A Method for Creating Integer Weights for Spatial Microsimulation." *Computers, Environment and Urban Systems* 41, 1–11.

Lovelace, R., M. Birkin, D. Ballas, and E. Van Leeuwen. (2015). "Evaluating the Performance of Iterative Proportional Fitting for Spatial Microsimulation: New Tests for an Established Technique." *Journal of Artificial Societies and Social Simulation* 18(2), 1–21.

Lu, Y., I. Cohen, X. S. Zhou, and Q. Tian. (2007). "*Feature Selection Using Principal Feature Analysis*". In Proceedings of the 15th ACM International Conference on Multimedia, 301–304.

Massey, D. S., and N. A. Denton. (1988). "The Dimensions of Residential Segregation." *Social Forces* 67(2), 281–315.

Monkkonen, P. (2011). "Housing Finance Reform and Increasing Socioeconomic Segregation in Mexico." *International Journal of Urban and Regional Research*, 36(4), 757–772. https://doi.org/10.1111/j.1468-2427.2011.01085.x

Mosteller, F. (1968). "Association and Estimation in Contingency Tables." *Journal of the American Statistical Association* 63(321), 1–28.

Musterd, S. (2020). "Urban Segregation: Contexts, Domains, Dimensions and Approaches." In *Handbook of Urban Segregation*, 1–16, edited by S. Musterd, Northampton: Edward Elgar Publishing.

Panori, A., D. Ballas, and Y. Psycharis. (2017). "Simathens: A Spatial Microsimulation Approach to the Estimation and Analysis of Small Area Income Distributions and Poverty Rates in the City of Athens, Greece." *Computers, Environment and Urban Systems* 63, 15–25.

Park, R. E., E. W. Burgess, R. D. McKenzie, and L. Wirth, (1925). *"The City (The University of Chicago Studies in Urban Sociology)."* Chicago, Ill.

Peng, H., F. Long, and C. Ding. (2005). "Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy." *IEEE Transactions on pattern Analysis and Machine Intelligence* 27(8), 1226–38.

Philips, I., G. Clarke, and D. Watling. (2017). "A Fine Grained Hybrid Spatial Microsimulation Technique for Generating Detailed Synthetic Individuals from Multiple Data Sources: An Application to Walking and Cycling." *The International Journal of Microsimulation* 10(1), 167–200.

Reardon, S. F. (2009). "Measures of Ordinal Segregation." In *Occupational and Residential Segregation*, edited by J. Silber, Y. Fluckiger and S. F. Reardon. Bingley: Emerald Group Publishing Limited.

Reardon, S. F.. (2011). "Measures of Income Segregation." Unpublished Working Paper, Stanford Center for Education Policy Analysis.

Reardon, S. F., and K. Bischoff. (2011). "Income Inequality and Income Segregation." *American Journal of Sociology* 116(4), 1092–153.

Reardon, S. F., and D. O'Sullivan. (2004). "Measures of Spatial Segregation." *Sociological Methodology* 34(1), 121–62.

Rey, S. J., A. T. Murray, and L. Anselin. (2011). "Visualizing Regional Income Distribution Dynamics." *Letters in Spatial and Resource Sciences* 4(1), 81–90.

Rose, A. N., and N. N. Nagle. (2017). "Validation of Spatiodemographic Estimates Produced through Data Fusion of Small Area Census Records and Household Microdata." *Computers, Environment and Urban Systems* 63, 38–49.

Smith, D. M., J. R. Pearce, and K. Harland. (2011). “Can a Deterministic Spatial Microsimulation Model Provide Reliable Small-Area Estimates of Health Behaviours? An Example of Smoking Prevalence in New Zealand.” *Health & Place* 17(2), 618–24.

Stephan, F. F., W. Edwards Deming, and M. H. Hansen. (1940). “The Sampling Procedure of the 1940 Population Census.” *Journal of the American Statistical Association* 35(212), 615–30.

Stewart Fotheringham, A., and D. W. S. Wong. (1991). “The Modifiable Areal Unit Problem in Multivariate Statistical Analysis.” *Environment and Planning A* 23(7), 1025–44.

Stone, R., and J. A. C. Brown. (1962). *A Computable Model of Economic Growth*. Cambridge: Chapman & Hall.

Terrance Ireland, C., and S. Kullback. (1968). “Contingency Tables with Given Marginals.” *Biometrika* 55(1), 179–88.

Van Ham, M., D. Manley, N. Bailey, L. Simpson, and D. Maclennan. (2012). “Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research.” In *Understanding Neighbourhood Dynamics*, 1–21. Dordrecht: Springer.

Wong, D. W. S. (1992). “The Reliability of Using the Iterative Proportional Fitting Procedure.” *The Professional Geographer* 44(3), 340–8.

Wong, D. W. S. (2002a). “Spatial Measures of Segregation and Gis.” *Urban Geography* 23(1), 85–92.

Wong, D. W. S. (2002b). “Modeling Local Segregation: A Spatial Interaction Approach.” *Geographical and Environmental Modelling* 6(1), 81–97.

Wong, D. W. S. (2003). “Spatial Decomposition of Segregation Indices: A Framework toward Measuring Segregation at Multiple Levels.” *Geographical Analysis* 35(3), 179–94.

Wong, D. W. S. (2004a). “Comparing Traditional and Spatial Segregation Measures: A Spatial Scale Perspective.” *Urban Geography* 25(1), 66–82.

Wong, D. W. S. (2004b). “The Modifiable Areal Unit Problem (Maup).” In *WorldMinds: Geographical Perspectives on 100 Problems*, edited by D. G. Janelle, B. Warf, and K. Hansen. 571–572. Dordrecht: Springer.

Yao, J., D. W. S. Wong, N. Bailey, and J. Minton. (2019). “Spatial Segregation Measures: A Methodological Review.” *Tijdschrift Voor Economische en Sociale Geografie* 110(3), 235–50.

Založnik, M. (2011). “*Iterative Proportional Fitting: Theoretical Synthesis and Practical Limitations*.” PhD thesis, University of Liverpool.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher’s web site.