

NBER WORKING PAPER SERIES

THE UPTICK IN INCOME SEGREGATION:
REAL TREND OR RANDOM SAMPLING VARIANCE

John R. Logan
Andrew Foster
Jun Ke
Fan Li

Working Paper 23656
<http://www.nber.org/papers/w23656>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2017, Revised January 2018

This research was supported by a research grant from National Institutes of Health (1R21HD078762-01A1), the Sociology Program of the National Science Foundation, and by the staff of the research initiative on Spatial Structures in the Social Sciences at Brown University. Cici Bauer provided helpful suggestions for calculation of segregation measures. The Population Studies and Training Center at Brown University (R24 HD041020) provided general support. The authors have full responsibility for the findings and interpretations reported here. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by John R. Logan, Andrew Foster, Jun Ke, and Fan Li. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Uptick in Income Segregation: Real Trend or Random Sampling Variance
John R. Logan, Andrew Foster, Jun Ke, and Fan Li
NBER Working Paper No. 23656
August 2017, Revised January 2018
JEL No. R23

ABSTRACT

Recent studies have reported a reversal of an earlier trend in income segregation in metropolitan regions, from a decline in the 1990s to an increase in the 2000-2010 decade. This finding reinforces concerns about the growing overall income inequality in the U.S. since the 1970s. Yet the evidence may be systematically biased to show an upward trend because the effective sample for the American Community Survey (ACS) is much smaller than it was for Census 2000, to which it is being compared. There is a possibility that the apparent changes in disparities across census tracts result partly from a higher level of sampling variation and bias due to the smaller sample. This study uses 100% microdata from the 1940 census to simulate the impact of different sampling rates on estimates of several measures of segregation and to propose and test the effectiveness of approaches to correcting the bias. It then applies those approaches to publicly available data for 2000 and 2007-2011. The reduction in sample sizes associated with the ACS results in exaggeration of evidence for increasing income segregation for all measures tested here, especially for subgroups (African Americans are studied here as an example). The methods of correction applied here will yield more conclusive and unbiased results when applied to the original sample data that is held internally by the Census Bureau.

John R. Logan
Department of Sociology
Brown University
Providence, RI 02912
john_logan@brown.edu

Jun Ke
Department of Biostatistics
Brown University
Providence, RI 02912
jun_ke@brown.edu

Andrew Foster
Department of Economics and Community Health
Brown University
64 Waterman Street
Providence, RI 02912
and NBER
afoster@brown.edu

Fan Li
Department of Biostatistics
and Bioinformatics
Duke University
2424 Erwin Rd
Durham, NC 27710
frank.li@duke.edu

The Uptick in Income Segregation: Real Trend or Random Sampling Variation?

Because neighborhoods are so consequential in people's lives and futures (Sampson 2012), urban social scientists have long been interested in neighborhood-level segregation. Although most of this literature focuses on separation by race and ethnicity, attention has also been given to segregation by social class or income. An early version of this research was devoted specifically to "underclass" neighborhoods, areas with high levels of poverty, unemployment, or other signs of distress (Ricketts and Sawhill 1988). Subsequent studies analyzed trends in income segregation across all income levels, with particular attention to income segregation within racial/ethnic groups.

The most recent research has been conducted in a period when social scientists, policymakers, and the public have become more acutely aware of issues associated with rising **income inequality** (Picketty 2013). Several recent reports have found that **income segregation**, too, is on the rise, increasing the estrangement of rich from poor and possibly leading to a decline in support for meeting the needs of less affluent Americans (Florida and Mellander 2015, Fry and Taylor 2012, Bischoff and Reardon 2014). Two patterns stand out in recent studies. First, past changes in overall income segregation have been unsteady, declining in one decade and rising in another, but segregation has been found to rise substantially after 2000. Second, income segregation is described as higher and rising more quickly within minority populations. These findings are widely enough accepted that they are referenced in public statements by political leaders: "What used to be racial segregation now mirrors itself in class segregation. This great sorting (has) taken place. It creates its own politics. There are some communities where ... I don't even know people who have trouble paying the bills at the end of the month. I just don't

know those people. And so there's less sense of investment in those children.” (President Barack Obama at a 2015 Poverty Summit, quoted by Liptak 2015).

There have been many studies of income segregation in recent decades, and these reach different conclusions depending on the time period, the measure of income segregation, and the segment of the population being studied (Abramson et al 1995, Mayer 2001, Watson 2009, Jargowsky 1996, 2003, Reardon and Bischoff 2011). Several studies distinguish between income segregation among whites and blacks (Jargowsky 1996, Massey and Fischer 2003, Watson 2009, Yang and Jargowsky 2006, Reardon and Bischoff 2011, Reardon, Fox and Townsend 2015). Representative of the most recent research is a widely cited study by Bischoff and Reardon (2014), using the ACS from 2007-2011. Some of their results are reproduced in Table 1, including an overall summary measure of family income segregation (H), a measure of the separation of families in the bottom decile of the income distribution from all others (H10), and a measure of the separation of families in the top decile (H90). Table 1 reports their measures for the total population and the black population for 2000 and 2007-2011 for several large metropolitan areas along with the mean and standard deviation of all the large metros in their sample.

Table 1 about here

The observed patterns of change vary depending on the measure and population segment that is studied. For the total population H and H90 increased in all 6 example metros and rose on average in large metros by 0.013 and .015, respectively. H10 increased in 4 of the 6 example metros and rose by 0.017 in the average large metro (these increases are all equivalent to about 0.5 standard deviations). These results support reports of increasing income segregation. Results for the black population show even higher values of segregation and larger increases

over time. H, H10 and H90 increased for all 6 example metros and the average value across metros with large black populations rose by .082, .099, and .110, respectively (more than one standard deviation).

We cast doubt on these findings. Our main insight is that all of these studies rely on sample data collected by the Census Bureau. Yet it is well known that the effective samples to estimate income distributions within census tracts were substantially downsized with the introduction of the American Community Survey (ACS) after 2000, while sample sizes within census tracts for minority populations have always been smaller than for the non-Hispanic white or total population. This recognition raises the general problem of small area estimation (Rao 2003). Estimates from random samples are known to be unbiased, but the variance of estimates can be quite large when samples are small. Social scientists in the past have treated the census's income tabulation in census tracts as though it were not based on a sample, presuming that the one-in-six long form data were sufficiently reliable for their analyses. But as Voss (2012) observes, in the ACS "standard errors of most estimates are so large that even substantial differences in numbers lack statistical significance" (see also U.S. Census Bureau 2009). For example, the most recent estimate of the median household income in relatively affluent tract 107.01 in Boston's Back Bay neighborhood in the 2013 ACS (with a typical population size of 1562 households) is \$99,234. The Census Bureau calculates a standard error of plus or minus \$13,552. So, staying within that confidence interval, the tract's median income may well have been as low as \$86,000, or it may have been as high as \$112,000. We suspect that the less reliable the income estimate is for every census tract, the larger will be the estimated variation across tracts – a value that is at the heart of all measures of income segregation.

If so, it is possible that in the full population – not in the samples enumerated by the Census Bureau – there actually was no change or even a decline in income segregation. Further, because the calculations for the black population in Table 1 were drawn from original samples that were typically less than half as large as those for the total population, and often only 10-20% as large, the apparently stronger pattern of increases in black income segregation may also be illusory. Reardon and Bischoff (2011, p.1143) were aware of the problem that “When the number of households in a unit is small (either because of a small population or because of sampling), estimates of within-unit variation will be biased downward, meaning that estimates of segregation will be biased upward.” They proposed to equalize sample sizes across metropolitan areas by a second stage of sampling, arguing that the resulting estimates would be “comparable across race groups, metropolitan areas, and years, regardless of population size” (2011, p. 1144). We show in Appendix I that procedures of this type do not compensate for differences in sample sizes in the original census samples.

Pitfalls and solutions in measuring income segregation

We are aware of two sources of error in the observed measures of income segregation based on publicly available data. The one that we focus on in this study is bias associated with limited sample size. The other is the Census Bureau’s withholding of data in public tabulations of income at the census tract level. The latter can only be overcome by gaining access to the original sample data in a Federal Statistical Research Data Center (RDC), but we mention it here to underline the fact that even the bias-corrected measures that we will present below – while an improvement over past results – are not definitive.

Data suppression in the census

The Census Bureau's concern with disclosure of information about individual persons has led to a specific pattern of data suppression: race-specific income data in tracts with few residents of a given subgroup are not published. This practice has consequences for measures of group-specific income segregation, and it may also have affected studies showing that affluent blacks and Hispanics are under-represented in higher income neighborhoods (Reardon, Fox, and Townsend 2015, Lichter, Parisi and Taquino 2012, Sharkey 2014). To illustrate the nature of the problem, we have analyzed suppression of income data for black households in the 2007-2011 ACS. We found that nationally there were over 17,000 census tracts with black residents for which no income distribution for black-headed households is reported. These tracts have a total black population over 1.2 million. The median income of the suppressed tracts (for the total population) is over \$70,000, and in most tracts members of all racial groups have similar income levels. This compares to the median income for families with a black householder in non-suppressed tracts of under \$45,000. The implication is that small, affluent minority populations in affluent (predominantly white and Asian) neighborhoods have not been taken into account in past research. The same issue arises for poor whites whose data will be suppressed in predominantly minority neighborhoods. In both cases, suppression could lead to an underestimate of income segregation.

Another disclosure practice is that in published tables at the tract level (in both Census 2000 and the ACS), the highest category of income is \$200,000 and above. This top-coding is of concern to studies of income inequality because it obstructs estimation of the income levels of households/families in the highest income category (Armour, Burkhauser and Larrimore 2016). It is also an issue in studies of income segregation where assumptions must be made about the

distribution of values in each category of income to calculate the variation of income within tracts (Reardon 2011). The top category is the most obvious problem because it has no necessary upper bound and the upper limit could vary greatly across tracts. It is especially problematic for measures calculated from the actual incomes rather than rank-ordered percentiles. Jargowsky (1996) uses such a measure and deals with the problem by assuming that the top category has a Pareto distribution.

Sampling strategies and weighting

We set these concerns aside now to focus on the issue of sample size that the approaches described below can ameliorate. Despite very large national samples, the decennial census (for what researchers refer to as sample-count variables) and the ACS have relatively small samples for individual census tracts. This is a problem shared with large-scale health surveys, which despite impressive national sample size have insufficient samples for reliable estimates of characteristics of smaller geographical areas. Statisticians define a “small area” as one where “the domain-specific sample is not large enough to support direct estimates of adequate precision” (Rao 2003, p. 1). Hence, depending on the data source a county or even a state may be “small.”

Demographers and public officials have become more aware of concerns about the nature of estimates of small area characteristics because of the substitution of the decennial long-form census (a one in six sample) by the annual American Community Survey (ACS). At the level of census tracts (for which the ACS pools data from five consecutive years) and even counties (which, depending on their size, are reported annually or with pooled three-year counts), ACS data are “noisier” than comparable data from 2000 and before (Navarro 2010). This is largely because the ACS samples are smaller. Tract estimates in the ACS are also affected by the use of

population controls from estimates made at the state and county level rather than at the tract level. Starsinic (2005) estimated that the standard errors from the five-year pooled ACS at the tract level would be about 50% higher than in Census 2000 long form data (see also National Research Council 2015, p. 24-40). The Census Bureau has attempted to deal with ACS's large confidence intervals through changes in the sampling design and through weighting techniques to account for probability of selection, nonresponse, and coverage adjustments (Asiala 2012). The National Research Council (2015) report on these efforts concludes that changes in sampling rates have tended to equalize the precision of estimates across tracts of different population sizes, but at the cost of decreasing the reliability of estimates for larger areas, resulting in minimal net improvement.

Complex weighting has another cost: while reducing bias, it increases the variance in sample weights, which in turn increase the margin of error of the final estimates. Hence the weighting procedures can be seen as “an implicit policy statement that unbiased (accurate) estimates are more important than precise (low-variance) estimates” (Spielman, Folch and Nagle 2014, p. 151). When applied to measuring the income distribution within tracts, this general problem is exacerbated by the fact that – aside from sample size – some kinds of places have less precise sample data than others. Specifically, Spielman, Folch and Nagle (2014, p. 151-154) show that in the 2007-2011 ACS the tracts with the lowest and highest median incomes have larger margins of error than tracts closer to the average income.

Correcting the bias

We now discuss in detail two standard types of measures of income segregation that are susceptible to bias related to sample size and derive potential approaches to correcting them. In a subsequent section we will use 100% microdata from the 1940 census to quantify the

relationship between sampling rate and bias and test the efficacy of our proposed corrections. We will then apply the corrections to the publicly available income data from the 2000 Census and 2007-2011 ACS to show how bias has affected conclusions about levels and changes in income segregation.

Researchers have employed several different measures of income segregation. The simplest is to divide the income distribution into a small number of categories, perhaps three, and to compute a standard segregation index (the Index of Dissimilarity) between the bottom and top categories, the rich and poor. This is the approach taken by Massey and Eggers (1993) and Massey and Fischer (2003). The simplicity is also a weakness, because such measures do not make use of the full income distribution provided by the census. We focus on two types of measures that do exploit the multiple and ordered category nature of the data. The first is the rank-order information theory index (H) used by Bischoff and Reardon (2014) and reported in Table 1.² It “compares the variation in family incomes within census tracts to the variation in family incomes in the metropolitan area” (2014, p. 212), having first recoded incomes into percentile ranks. Two other measures are based on variance. One version is the rank-order variance ratio index R, which like H relies on a rank ordering of incomes and has percentile specific variants such as R10 and R90. The rank-order variance ratio index “can be interpreted as a measure of the average variance of the neighborhood cumulative percentile density function” (Reardon 2011 p. 26).³ Another is the correlation ratio, which Jargowsky (1996) refers to as the Neighborhood Sorting Index (NSI). The NSI is simply the square root of the between-tract variance in income divided by the total variance of income, a familiar measure in analysis of variance. Like H, it “implicitly controls for the overall income level because it is based on

deviations from mean household income and also controls for income inequality because it is expressed as a percentage of total income variance” (Jargowsky 1996, p. 998).

We now present alternative approaches to reduce bias for each type of measure. First, we propose a method to estimate (and therefore to correct for) the upward bias in entropy-based measures (e.g., H , H_{10} , H_{90}) that draws solely on knowledge of the tract-level sample sizes and tract population counts. Then we propose an approach to estimate the variances within tracts using either the original interval-scale measure of family income or a rank-order measure, which can be summed to the city or metro level and used to estimate NSI or R.

Neither of these approaches solves the problem of variation in segregation estimates across samples, which is inherently greater when samples are smaller (e.g., in 2007-2011 compared to 2000 or for minority subgroups compared to the whole population). We will show, however, that they minimize the systematic bias that is found in uncorrected measures.

1. Correcting the bias in H

We have derived a convenient and feasible approximation of the bias in entropy based measures such as H and H_{90} related to sample size. The idea is to construct a quadratic Taylor expansion of the entropy function and apply this function to the sample income distributions and actual population counts by tract. This procedure builds on insights by Miller (1955). It yields an estimate of the bias that can be then subtracted from the sample estimate to get an approximate true estimate. It is most useful when sample sizes are not so large that the bias is trivial and not so small that the quadratic function is a poor fit to the entropy distribution over the relevant range.

We first note that with independent sampling without replacement the proportion of sampled households, s_j , in a tract with income above some percentile of the population has a

hypergeometric distribution with mean p_j and variance $\frac{p_j(1-p_j)}{N_j} \frac{M_j - N_j}{M_j - 1}$ where p_j is the proportion in the tract population above this percentile, M_j is the tract population, and N_j is the number in the sample. The population entropy of this tract (using the natural log form for notational convenience) is

$$\mathbb{E}(p_j) = p_j \ln(p_j) + (1-p_j) \ln(1-p_j) .$$

A second order approximation to the sample entropy is

$$\mathbb{E}(s_j) = \mathbb{E}(p_j) + (\ln(p_j) + \ln(1-p_j))(s_j - p_j) - \frac{1}{2p_j(1-p_j)}(s_j - p_j)^2 + O((s_j - p_j)^3).$$

Taking expectations with respect to s_j yields

$$\mathbb{E}(\mathbb{E}(s_j)) = \mathbb{E}(p_j) - \frac{1}{2} \frac{1}{N_j} \frac{M_j - N_j}{M_j - 1} + O_j$$

where O_j is the integral over the order statistic. Interestingly the population variable p_j does not appear in this expression except through O_j and thus the approximate bias can be calculated without knowing the true tract population income distribution. For example, the formal expression for the expected approximate bias for $H90$ adds up the tract-specific terms:

$$\mathbb{E}(H90) = H90 - \sum_j \frac{M_j(M_j - N_j)}{2M E(.1)N_j(M_j - 1)} - \sum_j \frac{M_j O_j}{M E(.1)}$$

where $M = \sum_j M_j$ and $E(.1) = .1 \ln(.1) + .9 \ln(.9)$.

In order to assess the accuracy of the approximation we need to be able to compute bounds on O_j . For any given M_j and N_j the expected (over possible samples) difference between the actual and approximate entropy as a function of the (unknown) number K of households in the population below the specified threshold is

$$O_j^*(K) = \sum_{k=0}^K \frac{\binom{K}{k} \binom{M_j - K}{n_j - k}}{\binom{M_j}{n_j}} \left(\frac{k}{n_j} \ln \left(\frac{k}{n_j} \right) + \left(1 - \frac{k}{n_j} \right) \ln \left(1 - \frac{k}{n_j} \right) - \frac{K}{M_j} \ln \left(\frac{K}{M_j} \right) + \left(\frac{K}{M_j} - 1 \right) \ln \left(-\frac{K}{M_j} + 1 \right) \right. \\ \left. - \left(\ln \left(\frac{K}{M_j} \right) - \ln \left(-\frac{K}{M_j} + 1 \right) \right) \left(\frac{k}{n_j} - \frac{K}{M_j} \right) + 1/2 \frac{N}{K} \left(\frac{k}{n_j} - \frac{K}{M_j} \right)^2 \left(\frac{K}{M_j} - 1 \right)^{-1} \right)$$

Note that while K is not known in practice it can take on only one of $M_j + 1$ possible values, each of which may be checked given M_j and N_j . The bounds of the set of $O_j^*(K)$ for $K \in [0, 1, \dots, M_j + 1]$ provide bounds for O_j .

For $M_j = 1000$ and $N_j = 50$, for example, the maximum of the absolute value of O_j is .0065, which applies when there is only one household in the tract population above (or below) the percentile cutoff. This is a rare scenario, and the average bias across tracts (which is relevant for the calculation of entropy statistics) is likely to be much smaller. Thus our approximation provides a useful basis for estimating bias. It is also worth noting that the adjustment for sampling without replacement will be small if the sample is small relative to the population. For example, for $M_j = 1000$ and $N_j = 50$ the term $\frac{M_j - N_j}{M_j - 1} = .95$. If we drop this term (thus assuming sampling with replacement) our approximate bias expression depends only on tract sample size:

$$\mathbb{E}(E(s_j)) = E(p_j) - \frac{1}{2} \frac{1}{N_j} + O_j^*.$$

Computation of bias for H is a straightforward extension of the above because the bias term does not depend on the percentile under consideration. In particular,

$$\mathbb{E}(H) - H \approx \sum_j \frac{M_j}{M} \frac{M_j - N_j}{N_j(M_j - 1)}$$

As before, if the sample is small relative to the population this expression reduced to

$$\mathbb{E}(H) - H \approx \sum_j \frac{M_j}{M} \frac{1}{N_j}$$

So the approximate bias in H is just the average inverted tract population size.

2. *Sparse-Sampling Variance Decomposition (SSVD)*

In contrast to the entropy-based measures the level of bias for variance-based measures of income segregation depends not just on sample and population sizes, but also on the distribution of incomes within and across tracts. Unfortunately these distributions are only imperfectly observed through the sample, and the bias correction must account for this fact as well. It was therefore necessary to develop a different approach to bias-correction for these measures than was used for the entropy measures, which we refer to as Sparse-Sampling Variance Decomposition (SSVD).

The idea of this approach is as follows. The variance-based measures decompose total variation into variation within and across census tracts. We presume that the total variance in income can be reliably estimated from the sample data in large cities and metropolitan areas. We then estimate the total within variance, from which the between variance can be deduced. We take advantage of being able to leverage results from a large number of census tracts to average-out errors in the estimation of the distribution within each tract. Although to our knowledge this has not been done before in this context, it is parallel to the use of “small t, large n” panel data methods (Mundlak 1978).

We begin by applying SSVD to estimating NSI. The expected total variation of income within tracts for the city is the average of the tract-specific variances weighted by the number of households in the tract. Tract populations are of course known and the variance based on the sample in each tract (using the standard $N-1$ bias correction) is an unbiased but noisy estimate of

the underlying population variance, even with samples as small as two. But the population weights are uncorrelated with the noise (which just arises from sampling). Hence the population-weighted average of the variance estimates for each tract from the sample converges to the within variation for the population as the number of tracts gets large. In addition, because the total variance for the population only involves the calculation of a single mean for the city (rather than a mean for each tract), the per-household population-weighted total variance estimated from the sample is consistent for the corresponding population measure as the number of tracts gets large. The population across-tract variation is just the total minus the within-tract variation in the population. Thus NSI can be estimated as the square root of 1 minus the ratio of the within to the total variation calculated from the sample using population weights.

The bias in the SSVD estimate of the NSI can be shown to approach 0 as the number of tracts increases. Assume there are J tracts and that the income y_{ij} of household i in tract j is a random variable drawn from a distribution F_j with mean μ_j and variance σ_j^2 . Further let the means be drawn from a distribution with mean μ and let the variance V_μ be drawn from a distribution with mean σ^2 and variance V_σ . Let $\mu_M = \sum_j M_j \mu_j / \sum_j M_j$, P_j denote the set of households in tract j , and S_j the set of households in tract j that are in the sample.

$$\text{The NSI for the full population is } NSI = \sqrt{A/T} = \sqrt{(T - W)/T} = \sqrt{1 - W/T}$$

where A is the across, W the within and T the total variation. Then

$$\mathbb{E}W = \mathbb{E} \frac{\sum_j \sum_{i \in P_j} (y_{ij} - \frac{1}{M_j} \sum_{k \in P_j} y_{kj})^2}{M} = \frac{\sum_j M_j \sigma_j^2}{M} \quad (1)$$

$$\mathbb{E}T = \mathbb{E} \frac{\sum_j \sum_{k \in P_j} (y_{ij} - \frac{1}{\sum_l M_l} \sum_{k \in P_j} y_{kl})^2}{M} = \frac{\sum_j M_j ((\mu_j - \mu_M)^2 + \sigma_j^2)}{M} \quad (2)$$

Denote the sample analogs of W and T as

$$\tilde{W} = \mathbb{E} \frac{\sum_j \frac{M_j}{N_j - 1} \sum_{i \in S_j} (y_{ij} - \frac{1}{N_j} \sum_{k \in S_j} y_{kj})^2}{\sum_j M} \quad (3)$$

$$\tilde{T} = \frac{\sum_j \frac{M_j}{N_j} \sum_{i \in S_j} (y_{ij} - \frac{1}{M} \sum_j \frac{M_j}{N_j} \sum_{k \in S_j} y_{kj})^2}{(\sum_j \sum_{i \in S_j} \frac{M_j}{N_j})} \quad (4)$$

The expected values for the mean in tract j , the overall mean, and the overall variance are provided from sample values corrected for degrees of freedom:

$$\mathbb{E} \frac{1}{N_j} \sum_{i \in S_j} y_{ij} = \mu_j, \quad \mathbb{E} \frac{1}{M} \sum_j \frac{M_j}{N_j} \sum_{k \in S_j} y_{kj} = \mu, \quad \text{and} \quad \mathbb{E} \frac{1}{N_j - 1} \sum_{i \in S_j} (y_{ij} - \frac{1}{N_j} \sum_{k \in S_j} y_{kj})^2 = \sigma_j^2.$$

Consequently $plim \frac{\tilde{W}}{\tilde{T}} - \frac{W}{T} = 0$ and $plim \widetilde{NSI} - NSI = 0$.

A critique of the NSI is that it is affected by changes in the distribution of income even when the ranking is preserved. Fortunately the SSVD can be applied to variance-based measures calculated from rank-order data such as R – simply transform the income data for the sample into cumulative percentiles and then do a variance decomposition of the resulting percentiles. In particular, for each percentile p one can calculate the fraction of households in each tract below that p and compute the tract-population weighted variance of this measure across tracts relative to the total variance in the sample. This measure for $p=90$, say R_{90}^R then, like H90 indicates the extent to which the top 10 percent of the population is segregated from other 90 percent. To compute R^R we average the across and total variation across all percentiles and then divide.

Because both R^R and R_{90}^R are based on across-variance estimates the SSVD decomposition follows exactly. Moreover, the approach is computationally efficient because the integration over p needed for R^R has an analytic solution and thus no numerical integration is needed.¹⁰

The integration in the SSVD estimate of the rank order estimate conveniently has a closed form solution:

$$\begin{aligned}\tilde{R}^R &= 1 - 6 \int_{p=0}^1 \left(\sum_j \frac{M_j}{M} \frac{1}{N_j - 1} \sum_{i \in S_j^*} (\mathbb{I}(f_{ij}^s < p)) - \frac{1}{N_j} \sum_{k \in S_j^*} \mathbb{I}(f_{ik} < p) \right)^2 dp \\ &= 1 - 6 \sum_j \frac{M_j}{M} \frac{N_j}{N_j - 1} \sum_{i \in S_j^*} f_{ij}^s \left(\frac{2i - 1}{N_j} - 1 \right)\end{aligned}$$

where f_{ij}^s is the cumulative percentile of a household in the sampled city population, accounting for any differential sampling weights by tract, and S_j^* indexes the sample S_j from tract j ordered such that if $i \leq i^*$ then $f_{ij}^s \leq f_{i^*j}^s \forall i, i^* \in S_j^*$. The number 6 comes from fact that

$$\int_0^1 p(1-p)dp = 1/6$$

The second expression, which is obtained by bringing the integral inside of

the summations, indicates that the SSVD estimate of the rank order variance ratio is simply a weighted average of the cumulative percentiles. Note that since the sample households are ordered within a tract, the track-specific weights are antisymmetric $w(i) = -w(N - i + 1)$ around the median ranked household in the tract.

Validating the bias correction procedures

Because the Census and ACS have collected income information only for samples of the population, there is no “true value” for income segregation using contemporary data against which we can compare bias-corrected measures. But using full-count historical microdata from 1940 we can carry out an exercise that is impossible with contemporary data: to draw samples of

varying proportions, then to calculate measures of income segregation across census tracts from those samples. Because the “real” level of segregation is known from the 100% data, we can determine how the “observed” level is affected by sampling proportion.

In carrying out this test we are mindful of the difference in sampling rates for income between the 2000 Census and the ACS. The 2000 Census long-form data were from a one-in-six sample of the population. The NRC (2015, p. 9) calculates a generalized design effect for the 2000 Census of 1.12, representing the degree to which the effective sample size from the Census’s design differs from a simple random sample. This reflects, for example, how the Census dealt with overall non-response and the use of population controls in developing weights. Hence we treat Census 2000 as approximately a 15% sample. The actual sampling rate for income may be lower than this, because income is among the variables for which non-response is especially high.

What sample proportion does the ACS represent for this purpose? The ACS is not conducted as part of a full census enumeration as the long form surveys in decennial censuses used to be. It utilizes a complex system of sampling and weighting, and it has changed in important ways over time. The sample size increased in 2011. At that time also the Census Bureau increased the differential in sampling proportions between smaller and larger census tracts in order to improve estimates for smaller tracts. One estimate for the 2007-2011 sample (National Research Council 2015, p. 24) is that the median tract sample size was 296 households (compared to 605 households in Census 2000). After taking into account the generalized design effect of 1.41, the effective sample size for the median tract was only 209 households (compared to 533 in Census 2000). These calculations convey the order of magnitude in the ACS’s reduction of sample size compared to Census 2000.

A more precise calculation can be made from the 2007-2011 ACS summary files for census tracts, which report the sample size in every tract. Across all tracts in the United States, the average final sample proportion was 8.2%. This is reduced to an effective sampling rate of 5.8% after taking into account the design effect but again not considering special concerns of non-response on income data. In the following analyses, we will treat sample proportions of around 5% as representative of the ACS. Spielman, Folch and Nagle (2014, p. 152) cite Census Bureau estimates that imputation rates for income variables approach 20%, suggesting that the actual ACS sampling rate for income data may be less than 5%.

Results from sample draws from the 1940 census

The 1940 census was the first to collect data on income. For each employed household member the enumerator listed the person's wage and salary income. Our analysis is for total household wage and salary income, combining the figures for the household head and all other household members. We did not measure family income because family relationships are not clearly defined in the 1940 data. In addition to income segregation for the total population we compute indices for a subgroup whose population share in 1940 was comparable to the share of black residents in the contemporary data: foreign-born whites (based on the country of birth of the household head). Results for foreign-born white households reveal how observed trends in segregation for specific subgroups may be affected by their smaller numbers.

We use data for all households enumerated in the 1940 census in Chicago (not presented here, we also compared Chicago results with five other large Northern and Midwestern cities, showing the same pattern of results). We drew samples of 5%, 10%, 15%, and 20%, repeating the procedure 100 times for every level of the sampling proportion and calculating every segregation measure for every draw.⁴ Non-residential tracts (10 households or less) are omitted;

a minimum of one household was sampled in every tract. This procedure yields a sampling distribution of estimated income segregation at each sampling rate. Because all of these values are from the same population in a single year, any differences between sampling distributions for lower or higher sampling proportions are due solely to varying sample size.

Figures 1 and 2 display the results of these sample draws for all Chicago households in 1940. Figure 1 reports values for the entropy-based measures H, H10, and H90; Figure 2 reports values for the variance-based measures R and NSI. On each plot one set of values represents the bias-corrected measures, and the other represents the uncorrected measures. At each of four selected sampling rates the plot displays the distribution of values from the 100 samples that were drawn: the maximum and minimum values, the median value, and the values at the 25th percentile and the 75th percentile (in addition, the median estimates for all the analyses of Chicago data are listed in Appendix Table 1). A straight line across the graph shows the true value of the measure as calculated from the original 100% microdata. The figures are useful as a visualization of the differences across measures, and they are especially helpful in displaying the bias in the uncorrected estimates.

Figures 1-2 about here

An important feature of the values in both figures is the sampling variation. For example, even the best performing estimate for H in Figure 1 (the corrected estimate with a 20% sampling rate) has a risk of yielding either too high or too low a value, even though on average – with many samples – it is within .01 of the correct figure of .062. In this exercise we are able to draw many samples and compare them, but in actual research there is typically only one sample. Consequently, researchers should be aware of not only the potential bias in estimates but also their sampling variation, which (as shown in the figures) diminishes with larger samples. We

draw attention to this statistical principle because although we show here that it is possible to correct for bias, sampling variation remains a concern with the lower sampling rate of the ACS compared to the decennial census long form that it replaced. Correcting for bias does not reduce the risk of drawing wrong conclusions due to sampling variation.

Figure 1 applies the correction calculated from average tract sample size to H, H10, and H90. Note that the median values of the corrected estimates fall within .03 of the actual values even at a 5% sampling rate for each of the three entropy measures. Bias in the uncorrected values, in contrast, is very high (.19) with a 5% sampling rate, falling substantially (.09) at 10%, and then continuing to improve but still present even at a sampling rate of 20%. The bias correction thus works well, but not perfectly, reflecting the fact that a second order approximation was employed in the derivation of that correction.

Figure 2 applies the SSVD to NSI and R. Again there is a spread in all sample estimates that is reduced at higher sampling rates. The uncorrected estimates for R and NSI are high at a 5% or 10% sampling rate, and some bias remains even at 20%. In contrast, there is no bias in the corrected estimates of R at any sampling rate. The correction of NSI greatly reduces the bias in the uncorrected estimate, though at a 5% sampling rate the median estimate of .246 is somewhat above the true value of .237.

These results support our initial concern about the upward bias in measures of income segregation. A clear implication is that if income segregation in the population has not changed between time 1 and time 2, but the data by which we measure segregation changes from a higher to a lower sampling rate, we will observe an apparent – but false – increase in segregation. We now turn to estimates for population subgroups where we expect greater impacts of variation in the sampling rate. For this purpose we select households with a foreign-born white head of

household (about 15% of households, similar to the share of a minority group in many contemporary metropolitan areas). Results are displayed in Figures 3-4.

Figures 3-4 about here

There is a similar pattern with all five segregation measures. Here we describe results for H in detail to point out the key findings. The actual population value of H for foreign-born white households in Chicago in 1940 was .045, lower than the level of income segregation of the whole population (.062). Yet the uncorrected estimate of H at a 5% sampling rate (comparable to today's ACS) was .093 for foreign born, higher than the estimated .081 for the total population at the same sampling rate. The greater bias for the foreign-born population at this sampling rate might have led researchers to conclude incorrectly that this minority population was especially highly segregated. Based on the corrected estimates, however, at every sampling rate the foreign-born population is shown to be less segregated.

Our proposed sample-size correction methods for entropy measures and SSVD for variance-based measures are shown here to be successful in reducing the bias from varying sampling rates. If we compare the corrected estimates to the true values (from the 100% sample), we find that it is more effective in some instances than in others. For the total population, the corrections yield median estimates that are close to the true values even at a 5% sampling rate for H, H10, H90, and R, but there is a slight upward bias on the median value of the NSI. For foreign born whites, who represent a smaller population segment and therefore smaller sample sizes, the corrected estimates (in relation to the true values) are excellent for H90 and R, slightly high for H10, and somewhat higher for NSI. Still, the corrected 5% median estimate of NSI is closer to the true value than the uncorrected 20% estimate.

Applying corrections to contemporary data

How does bias from the reduced sampling rate of the ACS affect estimates of change in income segregation? Related to this, how does the smaller sample size for subgroups of the population affect estimates of their segregation in comparison to the whole population (or to larger groups such as whites)? The techniques demonstrated here can be applied to contemporary data, but with several provisos. Most important, without access to internal census files it is impossible to take into account subgroups living in census tracts where their income distribution is not reported in the public data. As noted above, possibly many affluent blacks or Hispanics live in high-income, predominantly white census tracts but in too small numbers in any given tract to be reported. Possibly also many poor white or Asians live in low-income, predominantly minority census tracts but do not reach the threshold to be reported. To the extent that this occurs – and this cannot be estimated with public data – the observed income segregation of blacks and Hispanics may be skewed downward and that of whites and Asians upward. We cannot deal with this issue here.

We can, nevertheless, shed light on the potential of bias due to lower sampling rates in the ACS and smaller samples for population subgroups such as African Americans. For this purpose, the key step is to simulate samples of individual families from the grouped data published by the Census Bureau. The grouped data (counts of the number of families within each of many categories of income) conceal the exact distribution of incomes of families within each category, and (as other researchers have done) we estimate those distributions by making reasonable assumptions. We describe that procedure and examine its efficacy by reference to the 1940 results, concluding that estimates are robust for H, H10, H90, and R, but not for NSI. We then apply the same procedures to contemporary data for each of the rank-order measures to

determine how the bias that we have investigated here affects estimates of trends in income segregation.

Estimating income segregation from grouped data

The challenge in estimating measures of segregation using grouped data is that segregation measures may be sensitive to the distribution of incomes within categories and within tracts. Arbitrarily assigning the mean of the interval to each category may be reasonable for some purposes; however, initial explorations using the 1940 data suggested this approach did not work well for the measurement of income segregation. The problem is especially acute for the top category for which there is in principle no upper bound.

The limitations of group-level data have been addressed in previous work using assumptions about the smoothness of income distributions within tracts. Reardon (2011) proposes calculating a percentile-specific H for each category's lower or upper bound using the fraction of households in each tract above and below the bound in each tract. He then uses a fourth-order polynomial to interpolate values of H_{10} and H_{90} and to integrate over the full distribution of percentiles to obtain H . In constructing estimates of the NSI, Jargowsky (1996) makes use of published tract means, which the Census Bureau constructs from the unit-level data, but has to interpolate to estimate the metropolitan area variance. In particular, he assumes a linear distribution of households in lower income categories and a Pareto distribution for income categories above the metropolitan area mean.⁵ Integration over these category-wise distributions within tracts can then yield the desired measures.

Unfortunately, these methods do not apply in a straightforward way to the construction of bias-corrected methods. There are two issues. First, for the variance-based measures the SSVD requires information on the within-tract variation. While measures of variation by tract are

reported in some published tables, and indeed could be provided in most cases without violating confidentiality, we were unable to find consistent measures across the two censuses by race. Second, preliminary explorations suggested that simply subtracting a bias-correction from a standard entropy-based measure that integrates over the estimated distributions or interpolates quantiles may over-correct for sampling bias for relatively small samples. The reason is that the bias corrections assume that the data are sampled but smoothing itself removes some of the sampling variability.

To address these problems, we developed an approach based on Jargowsky's procedure. Instead of estimating the distribution only for the overall metropolitan area, we applied the procedure to each tract. Then, to preserve the sampling variability inherent in the original individual data, we sampled incomes from each distribution, with the number of sampled cases per category and tract being equal to the actual number of households. This procedure in effect recreates a unit-level data set from a grouped data set.⁶

We expect that estimating tract-specific distributions from grouped data will work if the noise in the estimation process averages out across tracts. Because we cannot demonstrate theoretically that it will, we turned once again to experimenting with the 1940 Chicago case. Starting with the 100% household level income data, we sampled from these data by tract using a specified sampling rate without replacement. We grouped the data into 10 pre-specified categories, keeping only the numbers of households by tract and category and the category boundaries. We then used the Jargowsky procedure to estimate the full tract and category-specific distributions based on the grouped data. Finally, we sampled from these estimated distributions with the number of households in each category and tract equal to the number retained in the simulated grouped data. The relevant algorithms were then applied to construct

biased and bias-corrected estimates of the relevant measures. The second-stage sampling was repeated 100 times and the results were averaged. The whole procedure, starting with the 100% household level income, was then repeated 100 times for each specified sampling rate.

Figures 5 and 6 plot the estimates of income segregation parallel to results in Figures 1-2 but now calculated from the grouped data (median estimates are reported in Appendix Table 1). True values (the horizontal line in the figures and the values in Appendix Table 1) are from the 100% individual data. Results of the correction are excellent for the entropy measures H, H10 (with some remaining upward bias at the lower sampling rates) and for R. H90 is a bit off but the corrected measures are still a substantial improvement over the uncorrected estimates even at a 20% sample.

It is noteworthy that in Figures 5 and 6, even when (for H90 and H-R) the intervals of the estimates do not overlap the true value they tend to be equal across sampling values. This relationship is a useful feature as it suggests that our correction procedure does succeed in removing differences that are attributed to changing sample sizes even when the grouping procedure results in some loss of information. Indeed, the process of grouping the data can yield bias even with a 100% sample as a consequence of the interpolation process. This pattern is especially evident with respect to H-R, for which the corrected estimates are identical at .060 across all sampling rates (even 100%, not shown in the table) for the grouped data. The comparable figure for the individual-level data, however, is .062. Analogously the gap between the group-based estimates of H90 and the true value in Figure 6 is attributable to the fact that even with a 100% sample, the grouped data yield an estimate of 0.063 as opposed to 0.059 for the individual data.

Further, the procedures using grouped data yield uncorrected and corrected estimates of NSI that are very different from and further from the true value than those calculated directly from the individual household data. At a 20% sampling rate, the corrected NSI virtually matched the true value when estimated from individual data for both the total population and the foreign born whites. But at this sampling rate, the corrected NSI from grouped data is too low for the total population and too high for foreign born whites. While even with grouped data the estimates are relatively invariant with respect to the sample size when SSVD is applied, we are not satisfied with the level of precision. The difficulty with the NSI may not be correctable, because it is due to the combination of the great sensitivity of NSI to the highest income values and our inability to reproduce the original distribution of these values from the grouped data. Consequently, we will not offer estimates of contemporary changes in NSI here, deferring to a future time when the internal census files may be available.

Figures 5-6 about here

Correcting measures of contemporary income segregation

We now turn to an effort to estimate income segregation in 2000 and 2007-2011. As already noted, these estimates are not “true” because they are distorted by working with grouped data and they do not compensate for suppression of income data for blacks in census tracts with small black populations. However they do remove the bias due to smaller sample sizes in 2007-2011 and for blacks compared to the total population. To maximize comparability with results previously reported by Bischoff and Reardon (2014) we use family income data at the tract level from Census 2000 (SF3) and ACS 2007-2011. Income segregation for the total population is for 118 metropolitan divisions or statistical areas with population above 500,000 in the 2007-2011 data; segregation for black families is measured for a subset of 66 metros with more than 30,000

black families. Metropolitan area boundaries are fixed by the definitions used in Census 2010. We applied the method described above with the 1940 data and summarize results in two tables, Table 2 for the total population and Table 3 for families with black household heads.

Table 2 lists values for the whole population for the same six metropolitan regions that were presented in Table 1 as well as the mean and standard deviation for large metropolitan regions. Our uncorrected figures differ from those reported by Bischoff and Reardon (2014); theirs are generally larger because they imposed a cap on sample size whose effect was greater for larger regions (see Appendix 1). For every measure and at both time points the mean is lower for uncorrected measures than for those that have been adjusted. The adjustments also result in changes in the trend over time. Averaging across cities, an increase of H of .017 (more than half a standard deviation) becomes an increase of .011; the corrections reduce the apparent gain in H in all six metros listed in the table. The increase in the mean H10 of .02 reduces to increase only .011 (in the case of Chicago it converts a decline of .005 to a decline of .015). The average increase of H90 for all cities was .026 without the correction but only .017 after correction. Similarly, R rose by .019 before correction, and by .014 after correction.

Table 2 about here

Did income segregation increase for the total population? Pending a more comprehensive analysis that may be completed in the future with internal census files and acknowledging variability in patterns across metropolitan regions, we conclude that it did increase, but that about a third of the previously reported increase is attributable to the lower sampling rate in the 2007-2011 ACS.

We turn now to income segregation among black families. We have emphasized that the upward bias from sample data is more pronounced for estimates of segregation within subgroups,

so we expect corrections to be greater for black families. This bias predates the ACS, because the tract-level sample sizes for the black income distribution have always necessarily been smaller than for the total population. Hence several studies over the years have reported that segregation among blacks is greater than in the general population. Bischoff and Reardon (2014, p. 215) draw particularly strong conclusions: “The trends in income segregation among black and Hispanic families are much more striking than those among white families... By 2009, income segregation among black families was 65 percent greater than among white families. Although income segregation among blacks grew substantially in the 1970s and 1980s, it grew at an even faster rate from 2000 to 2009, after declining slightly in the 1990s. Indeed, in the nine years from 2000 to 2009, income segregation among black families grew by almost two standard deviations.”

These conclusions are clearly contradicted by the corrected estimates for black families in Table 3. First, is segregation higher among black families? Comparing the mean values for 2007-2011 in Table 2 (total families) and Table 3 (black families) the differences are smaller after correction, and while two show blacks to be more segregated by income, two show them to be less segregated. In 2007-2011 H was higher for blacks by .011 (before correction it was higher by .033), and H10 was higher by .005 (higher by .036 before correction). On the other hand, H90 was lower by .011 (higher by .023 before correction), and R was lower by .009 (but higher by .013 before correction). These comparisons illustrate how attention to sampling can lead to very different conclusions.

Table 3 about here

Did black income segregation increase and did it increase more than in the total population? Uncorrected measures show that H increased by .017 in the total population and by

much more – .056 – for black families. The difference is substantially reduced after correction: H for blacks increased by .027 as compared to .011 for total population). After correction, H90 for black families increased by .034 (compared to .017 for total), and R by .032 (compared to .014 for total). H10, however, declined by the much larger margin of .078 (while increasing by .011 for total). Bischoff and Reardon (2014, p. 226) had drawn special attention to changes in H10, concluding that “[L]ow-income black and Hispanic families are much more isolated from middle-class black and Hispanic families than are low-income white families from middle-and high-income white families. The rapid growth of income segregation among black families has exacerbated the clustering of poor black families in neighborhoods with very high poverty rates.” Results from uncorrected measures fit readily into a common narrative about race and concentrated poverty. Corrected estimates for black families point in the opposite direction. There are also important differences in the comparisons across the six specific large cities reported in Table 3. For example, the uncorrected estimates for R show increases in income segregation among blacks in every city but Detroit. However, the corrected estimates show declines in three of the cities (Chicago, Detroit, and Pittsburgh) but increases in the other three (Cleveland, Los Angeles and Philadelphia). What seemed a simple set of results from the uncorrected measures is actually more nuanced and demands more explication.

Conclusion

It is plausible that income segregation has increased since 2000 along with the rising level of overall income inequality in the United States. Income segregation may also be higher and increasing faster for minorities than for the general population. We will not know for sure until corrected measures of income segregation are estimated from the original individual-level

sample data without suppression, but certainly past research based on the 2000 Census and 2007-2011 ACS was systematically biased to show an upward trend. Estimates using methods of correction for this bias suggest that income segregation did increase for the total population but not by as much as has been previously reported. Estimates for black families suggest that income segregation may not be higher and may actually have fallen for the lowest tenth of black families.

This research is immediately consequential our understanding of patterns and trends in income segregation. More generally these findings add to a growing awareness of the relevance of sampling issues to research based on the population census and ACS. Statisticians have always been aware of these issues, but in the past most researchers took the reliability and precision of census data for granted. The one-in-six long form data from 2000 and earlier were routinely treated as though they fully captured population characteristics. We are now in a better position to evaluate how this approach can lead researchers to wrong conclusions, thanks to the 100% census samples that are now being made available by the Minnesota Population Center. The 1940 data are especially useful because they include a more complete set of socioeconomic indicators (not only income, but also education, home value, rent, and recent unemployment experience). By drawing repeated samples from these full population data, researchers can now readily study the sampling variability of any parameter of interest. We have exploited this opportunity to examine the behavior of measures of income segregation. We have shown that there can be considerable bias even in a 15% similar to the long-form data collected in decennial censuses up to 2000. Now the Bureau of the Census has shifted to smaller annual samples in the American Community Survey that are no larger than 5% even when data are pooled over five years, and measures of income segregation have greater upward bias. The problem is necessarily

accentuated when dealing with subgroups of the population. We have given attention here to the foreign-born white population in 1940 and black families in the last decade. The same issues may be relevant for other subgroups, such as families with children (as studied by Owens 2016).

Fortunately sampling bias in these measures can be corrected. Results from repeated sample draws from the 100% data in 1940 give a high level of confidence in the efficacy of a simple population- and sample size-correction for entropy-based measures, and the SSVD estimation of variance-based measures. The remaining obstacles can be overcome by access to confidential census data centers, where researchers can study the original individual-level sample data without suppression. Despite the reduced sample size of the ACS, it will be possible in this way to draw clear conclusions about changes in income segregation in the country as a whole, in different regions, different parts of the income distribution, and different categories of people by race, family composition, or age.

A closely related question for further research is whether other measures of spatial inequality, such as traditional measures of racial/ethnic segregation, are affected by sampling in the same way. Fortunately some data are collected from full population samples, including even the counts of Asians and Hispanics of different national origins. Studies of segregation of some other potentially important categories of people, though, rely on sample data. These include distinctions within the black population between people of recent African origin, Afro-Caribbeans, and African Americans with a longer history in the U.S. Immigration researchers typically wish to distinguish people not only by race/ethnicity but also by nativity, and the Census Bureau provides the necessary tabulations by census tract from sample data. As noted above, some researchers wish to distinguish between family households with and without children. Studies of this type are being conducted, and it is important to know whether the usual

indices of segregation are subject to the same kind of bias as the income segregation measures evaluated here.

Footnotes

1. The 2000 measures are calculated from tract data from SF3 (the sample count data) of Census 2000, and 2007-2011 measures are calculated from the 2007-2011 five-year pooled American Community Survey tract data. Reardon and Bischoff (2014) included metropolitan areas with population greater than 500,000 in 2007 ($n=117$); for black income segregation they included a subset of these areas where the number of black families was greater than 10,000 in every census year since 1970 and in 2009 ($n=65$). Measures of H , H_{10} , and H_{90} in Table 1 are based on data provided by them. The means and standard deviations in Table 1 are unweighted.

2. Bischoff and Reardon (2014, pp. 227-228) describe H^R as follows (citations omitted). “For any given value of p , we can dichotomize the income distribution at p and compute the residential (pairwise) segregation between those with income ranks less than p and those with income ranks greater than or equal to p . Let $H(p)$ denote the value of the traditional information theory index of segregation computed between the two groups so defined. Likewise, let $E(p)$ denote the entropy of the population when divided into these two groups. That is,

$$E(p) = p \log_2 \frac{1}{p} + (1 - p) \log_2 \frac{1}{(1 - p)}$$

and

$$H(p) = 1 - \sum_j \frac{t_j E_j(p)}{TE(p)},$$

where T is the population of the metropolitan area and t_j is the population of neighborhood j .

Then the rank-order information theory index (H^R) can be written as

$$H^R = 2 \ln(2) \int_0^1 E(p) H(p) dp$$

Thus, if we computed the segregation between those families above and below each point in the income distribution and averaged these segregation values, weighting the segregation between families with above-median income and below-median income the most, we get the rank-order information theory index.”

3. The percentile- p specific variance-ratio index may be defined as

$$R_p^R = \sum_j \frac{M_j}{\sum_k M_k} \left(\frac{1}{M_j} \sum_{i \in P_j} \mathbb{I}(f_{ij} < p) - p \right)^2 / (p(1-p))$$

where $\mathbb{I}()$ is the indicator function, f_{ij} is the cumulative percentile of income of household i in tract j relative to the city, P_j indexes the population of households in tract j , $M_j = |P_j|$ is the number of households in tract j and J is the number of tracts. Note that the total variation, and thus the denominator in this expression is $p(1-p)$ because the fraction of households with income less than p in the population is exactly p .

Similarly, the rank-order variance ratio index is defined as

$$R^R = 6 \int_{p=0}^1 \frac{M_j}{\sum_j M_j} \sum_j \left(\frac{1}{M_j} \sum_{i \in P_j} \mathbb{I}(f_{ij} < p) - p \right)^2 dp.$$

4. As reviewers pointed out, the reliability of samples depends directly not on the sampling proportion but on the number of sampled cases, particularly the number of cases in each census tract from which the tract’s income distribution is estimated. We repeated our analyses for average sample sizes per tract of 50, 100, 150, 200, and 250. Tracts averaged about 1300 households, so a sample of 50 would be just under 4%, while a sample of 200 be around 15%. Results of these analyses showed the same patterns as did the analyses where we varied sampling rate.

5. The Pareto Distribution $\{F(Y) = 1 - \alpha Y^{-\beta}\}$ describes the distribution of a population with incomes of Y or greater. One way to estimate α and β is by using Quantile Method proposed by Quandt (1964). Choose two probability levels P_1 and P_2 and determine the corresponding quantiles Y_1 and Y_2 from the income category:

$$P_1 = 1 - \alpha Y_1^{-\beta}, \quad P_2 = 1 - \alpha Y_2^{-\beta},$$

$$\log(1 - P_1) = \log \alpha + \beta \log Y_1,$$

$$\log(1 - P_2) = \log \alpha + \beta \log Y_2,$$

Then:

$$\beta = \frac{\log N_1 - \log N_2}{\log Y_1 - \log Y_2},$$

$$\alpha = \frac{1 - P_1}{Y_1^{-\beta}} = \frac{N_1 / N}{Y_1^{-\beta}}.$$

Where N_1 and N_2 are the number of household whose income is at least greater than Y_1 , Y_2 , respectively. N is the total number of households among all the tracts. Following Jargowsky (1996) we use the Pareto Distribution to estimate the variance in each category:

$$\sum_{i=1}^{n_b} \delta^2 = \int_{y_1}^{y_2} (y - \bar{y})^2 f(y) dy$$

6. We also implemented Reardon's approach of estimating H using a fourth order polynomial to the metro-level entropy function, which was explicitly designed for group level data and thus does not require second-stage sampling. In Figures 5-6, Tables 2-3, and Appendix Table 1 this

estimate of H from grouped data is identified as $H-R$. We include these results to show that our correction procedure improves the estimates for either approach. The uncorrected $H-R$ estimates are better than the uncorrected H estimates, and we attribute this to its having smoothed out some sampling fluctuations of values within categories. Because we cannot use Reardon's approach for the variance-based measures, we focus here on results using our modification of Jargowsky's method.

References

- Abramson, Alan J., Mitchell S. Tobin & Matthew R. VanderGoot, 1995. The changing geography of metropolitan opportunity: The segregation of the poor in U.S. metropolitan areas, 1970 to 1990. *Housing Policy Debate*, 6:1, 45-72, DOI: 10.1080/10511482.1995.9521181
- Asiala, Mark E. 2012. “American Community Survey Sampling and Weighting” Conference on Measuring People in Place, Boulder CO, October 5. Accessed February 8, 2013 at http://www.colorado.edu/ibs/cupc/workshops/measuring_people_in_place/
- Bischoff, Kendra and Sean F. Reardon. 2014. Residential Segregation by Income, 1970–2009. Pp. 208-233 in John R. Logan (editor), *Diversity and Disparities: America Enters a New Century*. New York: Russell Sage Foundation.
- Bureau of the Census. 2006. “Design and Methodology: Current Population Survey. Technical Paper 66.” <http://www.census.gov/prod/2006pubs/tp-66.pdf> [accessed 3/22/15].
- Card, David, Jörg Heining, and Patrick Kline. 2012. “Workplace heterogeneity and the rise of West German wage inequality” . No. w18522. National Bureau of Economic Research.
- Florida, Richard and Charlotta Mellander. 2015. “Segregated City: The Geography of Economic Segregation in America’s Metros” Report by the Martin Prosperity Institute, University of Toronto. Accessed 11/24/15 at <http://martinprosperity.org/media/Segregated%20City.pdf>.
- Firebaugh, Glenn. 1999. “Empirics of World Income Inequality” *American Journal of Sociology* 104: 1597-1630.
- Fry, Richard and Paul Taylor. 2012. “The Rise of Residential Segregation by Income” Report by the Pew Research Center. Accessed 11/24/15 at

<http://www.pewsocialtrends.org/files/2012/08/Rise-of-Residential-Income-Segregation-2012.2.pdf>.

- Harville, D. A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems" *Journal of the American Statistical Association* 72 (358): 320–338. doi:10.2307/2286796.
- Jargowsky, Paul A. 1996. "Take the Money and Run: Economic Segregation in U.S. Metropolitan Areas." *American Sociological Review* 61(6): 984–98.
- Jargowsky, Paul A. 2003. "Stunning Progress, Hidden Problems: The Dramatic Decline of Concentrated Poverty in the 1990s." In *Redefining Urban and Suburban America: Evidence from Census 2000*, vol. 2, ed. Bruce Katz, Alan Berube, and Robert Lang. Washington, D.C.: Brookings Institution Press.
- Liptak, Kevin. 2015. "Obama draws line between racial segregation of the past and class segregation today" CNN Politics, 5/12/15. Accessed 11/24/15 at <http://www.cnn.com/2015/05/12/politics/obama-combat-poverty-baltimore/>
- Massey, Douglas S and Mitchell L. Eggers . 1993. "The Spatial Concentration of Affluence and Poverty During the 1970s." *Urban Affairs Quarterly* 29:299-315.
- Massey, Douglas S. and Mitchell L. Eggers. 1990. "The Ecology of Inequality: Minorities and the Concentration of Poverty, 1970-1980." *American Journal of Sociology* 95:1153-88.
- Massey, Douglas S., Andrew B. Gross, and Kumiko Shibuya. 1994. "Migration, Segregation, and the Geographic Concentration of Poverty." *American Sociological Review* 59:425-45.
- Massey, Douglas S., and Mary J. Fischer. 2003. "The Geography of Inequality in the United States, 1950–2000." *Brookings-Wharton Papers on Urban Affairs*, 1–40.

- Mayer, Susan E. 2001. "How Growth in Income Inequality Increased Economic Segregation," Joint Center For Poverty Research Working Paper 230.
- Miller, George. 1955. "Note on the Bias of Information Estimates" *Information Theory in Psychology: Problems and Methods*. II-B: 95-100.
- Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data" *Econometrica* 46: 69-85.
- National Research Council. 2015. *Realizing the Potential of the American Community Survey: Challenges, Tradeoffs, and Opportunities*. Washington, D.C.: National Academies Press.
- Owens, Ann. 2016. "Inequality in Children's Contexts: Income Segregation of Households with and without Children" *American Sociological Review* 81: 549-574.
- Petersen, Trond, Andrew M. Penner, and Geir Høgsnes. 2014. "From Motherhood Penalties to Husband Premia: The New Challenge for Gender Equality and Family Policy, Lessons from Norway¹" *American Journal of Sociology* 119: 1434-1472.
- Picketty, Thomas. 2013. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.
- Quandt, Richard E. 1964. "Old and New Methods of Estimation and the Pareto Distribution" *Econometric Research Program, Research Memorandum No. 70*. Princeton NJ: Princeton University. Accessed 3/29/16 at <https://www.princeton.edu/~erp/ERParchives/archivepdfs/M70.pdf>.
- Rao, J.N.K. 2003. *Small Area Estimation*. New York: Wiley-Interscience.
- Reardon, Sean F. 2011. "Measures of Income Segregation." Working paper. Stanford, Calif.: Stanford University, Center for Education Policy Analysis. Available at:

- http://inequality.stanford.edu/sites/default/files/reardon_measures-income-seg.pdf
(accessed September 1, 2017).
- Reardon, Sean F. and Kendra Bischoff. 2011. "Income Inequality and Income Segregation Source" *American Journal of Sociology*, 116:1092-1153.
- Reardon, Sean F., Lindsay Fox and Joseph Townsend. 2015. "Neighborhood Income Composition by Household Race and Income, 1990-2009" *Annals of the American Academy of Political and Social Science* 660: 78-97.
- Ricketts, Erol R. and Isabel V. Sawhill. 1988. "Defining and Measuring the Underclass." *Journal of Policy Analysis and Management* 7:316-25.
- Sampson, Robert. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago: University of Chicago Press.
- Schnabel, R. B., Koontz, J. E. and Weiss, B. E. 1985. "A modular system of algorithms for unconstrained minimization" *ACM Trans. Math. Software*, 11, 419-440.
- Spielman, Seth E. , David Folch, and Nicholas Nagle. 2014. "Patterns and causes of uncertainty in the American Community Survey" *Applied Geography* 46: 147-157.
- Starsinic, Michael. 2005. "American Community Survey: Improving Reliability for Small Area Estimates" *Proceedings of the 2005 Joint Statistical Meetings*, Pp. 3592-3599.
- Voss, Paul. 2012. "Usability of the ACS" Conference on Measuring People in Place, Boulder CO, October 5.
http://www.colorado.edu/ibs/cupc/workshops/measuring_people_in_place/
- Watson, Tara. 2009. "Inequality and the Measurement of Residential Segregation by Income." *Review of Income and Wealth* 55(3): 820-44.

Yang, Rebecca, and Paul A. Jargowsky. 2006. "Suburban Development and Economic Segregation in the 1990s." *Journal of Urban Affairs* 28:253–73.

Appendix I: Two-stage resampling as a possible solution to bias

We use 1940 census data here to test a resampling procedure developed by Bischoff and Reardon to correct for differences in sampling rates of H in the 2000 and 2010 census using aggregate tract level data. The idea of the resampling approach is that if similarly sized small samples are drawn based on observed distributions for the two censuses then small-sample biases, while present, will be comparable across time and thus trends will be correctly estimated. The procedure, as explained in a private communication from its authors and similar in concept to their handling of data from 2000 in a previous study (Reardon and Bischoff 2011), is that a simulated micro data set is created with the number of households in each category of income by tract equal to the number implied by the published distributions. Then 100 samples of $50 \cdot N$, where N is the number of tracts, are drawn from the simulated micro data for each year. H is calculated for each sample using grouped data procedure for which Bischoff and Reardon provided Stata code. The procedure involves fitting 4-th order polynomials to the cumulative distributions and then integrating over this distribution to H. The 100 sample Hs are then averaged to get a final H estimate for each year.

We carried out a two stage procedure to mimic the data generating process underlying this procedure, and we applied it to H, H90, and NSI with similar results for each measure. Here we display the results for H. In the first stage we constructed 100 5% and 15% samples and one 100% sample of the 1940 micro data and then categorized the data by tract to create the equivalent of the published tables for each sample. In a second stage we carried out the proposed resampling procedure for each first-stage sample. We then plotted the difference between the estimate and a “true” estimate based on the 100% sample without any second stage sampling. A

box plot of the resulting differences by original sampling rate is presented in Appendix Figure 1 and labeled “N*50 Cen”. We also carried out a variant in which the microdata consisted of the number of households implied by the census tract sample sizes rather than the census tract population counts (“N*50 Samp”). A third variant (All) was constructed in which there was no second-stage sampling.

It is evident from Appendix Figure 1 that despite the fact that the second stage sample is of the same size for the 5 and 15 percent samples, the Bischoff and Reardon procedure (“N*50 Cen”) does not yield similar biases for them. The average difference is .009 or 16% of the true value of .056. This difference is roughly equal to the difference of .01 when no second stage is sampled at all (“All”). In short, the “N*50 Cen” approach increases the bias by roughly the same amount for the 5 and 15 percent samples and thus does not undo the difference in bias that is created from the different sampling rates used in the first stage.

The bias estimates for “N*50 Samp” are more comparable across samples. This suggests that the Bischoff and Reardon procedure would work better if it were based on the counts of the sample rather than the counts of the population that are estimated from the sample. In effect this works because it approximates what you would get by randomly discarding 2/3 of the 15% sample so that sampling rates are in fact the same. There are two drawbacks with this approach. First, in contrast to the “N*50 Cen” procedure, the “N*50 Samp” procedure will not work to compute segregation among subgroup populations for which 5% of the population per tract on average is less than 50. Second, bias depends on the within tract distributions (and thus segregation) as well as on sampling size, so there is no guarantee that bias will be the same in different cities with different levels of segregation. Simulations available from the authors on

request suggest that this latter source of bias (due to differing within tract variation) is not large for H, but it can be important for other measures such as H10 and H90.

Appendix Figure 1: Bias from resampling procedures based on 1940 Chicago Census Data

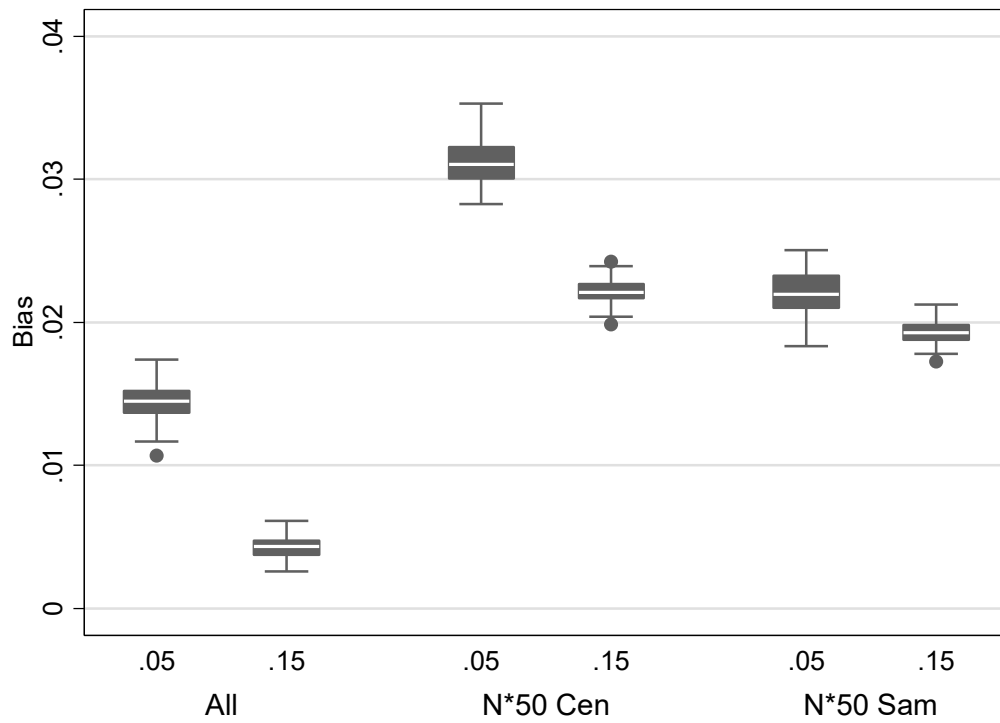


Table 1. Observed income segregation in metropolitan regions, 2000 and 2007-2011				
	Total		Black	
	2000	2007-2011	2000	2007-2011
H				
Chicago	0.164	0.168	0.147	0.186
Cleveland	0.158	0.172	0.160	0.202
Detroit	0.162	0.194	0.094	0.141
Los Angeles	0.174	0.179	0.177	0.268
Philadelphia	0.189	0.207	0.127	0.175
Pittsburgh	0.114	0.130	0.184	0.254
<i>Mean large metros¹</i>	<i>0.134</i>	<i>0.147</i>	<i>0.169</i>	<i>0.251</i>
<i>SD</i>	<i>0.027</i>	<i>0.027</i>	<i>0.055</i>	<i>0.082</i>
H(10)				
Chicago	0.200	0.186	0.153	0.197
Cleveland	0.221	0.219	0.156	0.221
Detroit	0.178	0.195	0.087	0.166
Los Angeles	0.132	0.149	0.155	0.269
Philadelphia	0.217	0.230	0.128	0.174
Pittsburgh	0.130	0.154	0.175	0.303
<i>Mean large metros¹</i>	<i>0.146</i>	<i>0.163</i>	<i>0.171</i>	<i>0.270</i>
<i>Standard deviation</i>	<i>0.031</i>	<i>0.029</i>	<i>0.054</i>	<i>0.093</i>
H(90)				
Chicago	0.211	0.233	0.168	0.253
Cleveland	0.203	0.224	0.194	0.272
Detroit	0.203	0.261	0.135	0.212
Los Angeles	0.257	0.274	0.253	0.358
Philadelphia	0.240	0.245	0.188	0.284
Pittsburgh	0.189	0.203	0.253	0.361
<i>Mean large metros¹</i>	<i>0.185</i>	<i>0.200</i>	<i>0.231</i>	<i>0.341</i>
<i>SD</i>	<i>0.036</i>	<i>0.036</i>	<i>0.079</i>	<i>0.096</i>

¹ See footnote 1 for selection of large metros and metros with large black populations. Source: Downloadable from <https://s4.ad.brown.edu/Projects/Diversity/Data/Data.htm>.

Table 2. Estimates of income segregation before and after correction for selected metropolitan regions and for the average of 118 large metropolitan regions, 2000 and 2007-2011

			Chicago	Cleveland	Detroit	Los Angeles	Philadelphia	Pittsburgh	<i>Mean large metros</i>	<i>Change in mean</i>	<i>Standard deviation</i>
H	Uncorrected	2000	0.140	0.140	0.143	0.155	0.163	0.088	0.112		0.028
		2007-11	0.153	0.153	0.185	0.167	0.178	0.106	0.129	0.018	0.029
	Corrected	2000	0.136	0.135	0.137	0.151	0.158	0.085	0.108		0.028
		2007-11	0.143	0.143	0.172	0.155	0.168	0.098	0.119	0.012	0.029
H-R	Uncorrected	2000	0.138	0.137	0.141	0.153	0.160	0.087	0.110		0.027
		2007-11	0.148	0.148	0.179	0.162	0.172	0.102	0.125	0.015	0.028
	Corrected	2000	0.133	0.132	0.135	0.149	0.155	0.083	0.106		0.027
		2007-11	0.138	0.138	0.166	0.150	0.162	0.095	0.115	0.009	0.028
H10	Uncorrected	2000	0.158	0.181	0.147	0.097	0.181	0.081	0.106		0.031
		2007-11	0.153	0.186	0.170	0.117	0.193	0.107	0.126	0.020	0.030
	Corrected	2000	0.152	0.174	0.138	0.091	0.173	0.076	0.100		0.030
		2007-11	0.137	0.170	0.150	0.098	0.177	0.095	0.111	0.011	0.029
H90	Uncorrected	2000	0.178	0.173	0.164	0.216	0.201	0.152	0.152		0.049
		2007-11	0.204	0.195	0.231	0.246	0.212	0.170	0.178	0.026	0.056
	Corrected	2000	0.171	0.166	0.155	0.210	0.193	0.146	0.146		0.049
		2007-11	0.188	0.179	0.211	0.227	0.197	0.159	0.163	0.017	0.055
R	Uncorrected	2000	0.240	0.239	0.246	0.546	0.280	0.154	0.194		0.047
		2007-11	0.249	0.250	0.301	0.269	0.292	0.176	0.213	0.019	0.048
	Corrected	2000	0.235	0.233	0.240	0.275	0.275	0.149	0.189		0.047
		2007-11	0.237	0.239	0.288	0.265	0.283	0.166	0.203	0.013	0.048

**Table 3. Estimates of income segregation for black families before and after correction,
for selected metropolitan regions and for the average of 116 large metropolitan regions, 2000 and 2007-2011**

			Chicago	Cleveland	Detroit	Los Angeles	Philadelphia	Pittsburgh	<i>Mean large metros</i>	<i>Change in mean</i>	<i>Standard deviation</i>
H	Uncorrected	2000	0.120	0.134	0.076	0.130	0.094	0.114	0.116		0.032
		2007-11	0.134	0.171	0.113	0.165	0.127	0.142	0.162	0.046	0.041
	Corrected	2000	0.110	0.123	0.067	0.119	0.084	0.098	0.103		0.029
		2007-11	0.112	0.142	0.087	0.132	0.100	0.103	0.130	0.027	0.036
H-R	Uncorrected	2000	0.117	0.132	0.074	0.127	0.092	0.112	0.113		0.031
		2007-11	0.130	0.164	0.108	0.155	0.122	0.134	0.154	0.041	0.039
	Corrected	2000	0.107	0.121	0.065	0.115	0.083	0.095	0.100		0.028
		2007-11	0.107	0.136	0.082	0.122	0.095	0.094	0.180	0.080	0.047
H10	Uncorrected	2000	0.108	0.107	0.055	0.094	0.079	0.091	0.226		0.059
		2007-11	0.128	0.166	0.105	0.154	0.123	0.162	0.162	-0.065	0.044
	Corrected	2000	0.093	0.090	0.042	0.077	0.063	0.065	0.194		0.058
		2007-11	0.093	0.122	0.065	0.104	0.082	0.101	0.116	-0.078	0.034
H90	Uncorrected	2000	0.124	0.144	0.098	0.176	0.110	0.144	0.138		0.047
		2007-11	0.176	0.189	0.153	0.208	0.176	0.181	0.201	0.063	0.048
	Corrected	2000	0.109	0.127	0.085	0.159	0.094	0.119	0.118		0.042
		2007-11	0.142	0.145	0.113	0.157	0.135	0.121	0.152	0.034	0.041
R	Uncorrected	2000	0.195	0.218	0.290	0.200	0.153	0.179	0.180		0.047
		2007-11	0.202	0.248	0.124	0.230	0.180	0.181	0.226	0.046	0.059
	Corrected	2000	0.181	0.203	0.161	0.181	0.139	0.155	0.162		0.044
		2007-11	0.178	0.219	0.113	0.195	0.152	0.140	0.194	0.032	0.058

**Appendix Table 1. Estimates of various measures of income segregation,
with samples of different sizes, and true 100% population value, Chicago 1940**

	Uncorrected estimates				Corrected estimates				True Value
	5%	10%	15%	20%	5%	10%	15%	20%	
Estimates from individual data									
Total population									
H	0.081	0.071	0.067	0.066	0.065	0.063	0.063	0.063	0.062
H10	0.053	0.038	0.033	0.031	0.028	0.026	0.026	0.026	0.026
H90	0.085	0.071	0.066	0.064	0.059	0.059	0.059	0.059	0.059
NSI	0.277	0.26	0.253	0.244	0.246	0.241	0.24	0.236	0.237
R	0.114	0.107	0.104	0.103	0.099	0.099	0.099	0.099	0.099
Foreign-born whites									
H	0.093	0.067	0.059	0.054	0.053	0.048	0.047	0.046	0.045
H10	0.087	0.054	0.042	0.036	0.026	0.024	0.023	0.023	0.023
H90	0.107	0.076	0.065	0.059	0.046	0.046	0.046	0.046	0.046
NSI	0.288	0.242	0.22	0.209	0.199	0.183	0.174	0.177	0.177
R	0.107	0.089	0.082	0.079	0.07	0.069	0.069	0.069	0.069
Estimates from grouped data									
Total population									
H	0.082	0.071	0.068	0.066	0.065	0.064	0.063	0.063	0.062
H-R	0.076	0.068	0.065	0.063	0.06	0.06	0.06	0.06	0.062
H10	0.053	0.038	0.034	0.031	0.027	0.026	0.026	0.026	0.026
H90	0.089	0.075	0.071	0.068	0.063	0.063	0.063	0.063	0.059
NSI	0.33	0.278	0.233	0.217	0.129	0.142	0.145	0.158	0.237
R	0.116	0.108	0.105	0.104	0.101	0.101	0.1	0.1	0.099
Foreign-born whites									
H	0.094	0.068	0.059	0.055	0.054	0.048	0.047	0.046	0.045
H-R	0.084	0.063	0.056	0.052	0.044	0.044	0.043	0.043	0.045
H10	0.088	0.054	0.042	0.037	0.028	0.025	0.024	0.023	0.023
H90	0.108	0.077	0.067	0.062	0.047	0.048	0.048	0.048	0.046
NSI	0.481	0.429	0.364	0.347	0.239	0.24	0.205	0.189	0.177
R	0.108	0.09	0.083	0.08	0.072	0.07	0.07	0.07	0.069

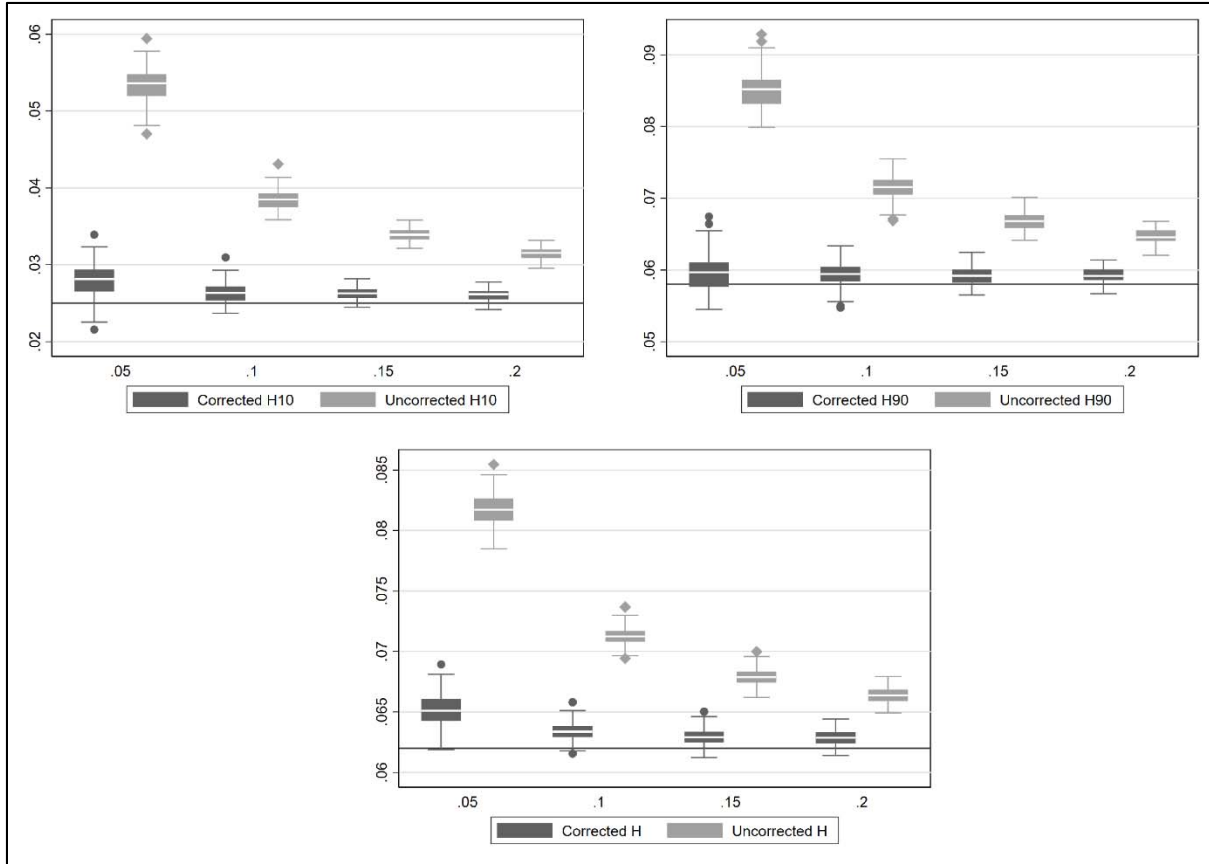


Figure 1. Estimated income segregation in Chicago 1940: true value (horizontal line), and distribution of estimates from samples of varying size, showing uncorrected and corrected values of H, H10, and H90 based on data for individual households, total population.

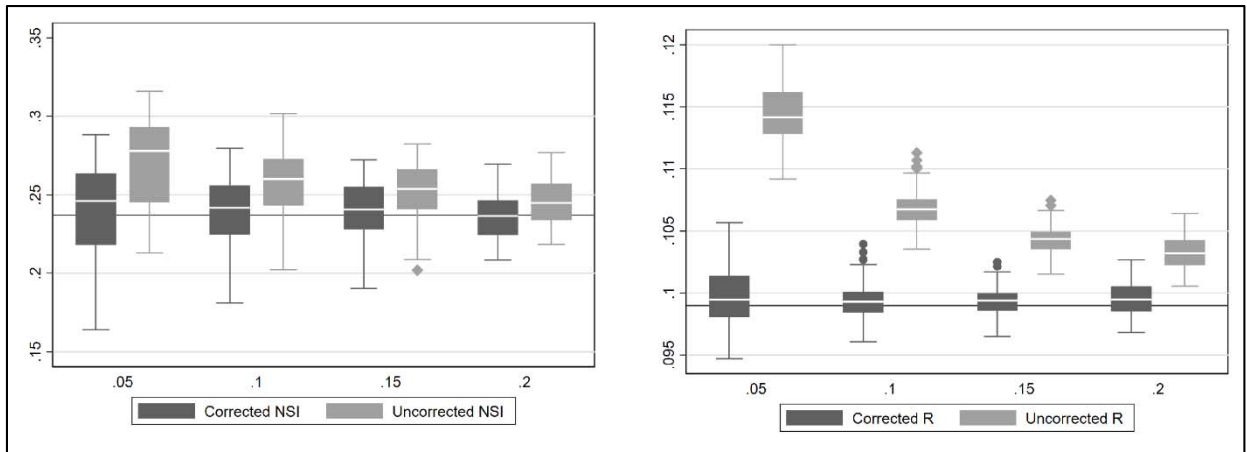


Figure 2. Estimated income segregation in Chicago 1940: true value (horizontal line), and distribution of estimates from samples of varying size, showing uncorrected values of R and NSI and adjusted values using SSVD, based on data for individual households, total population.

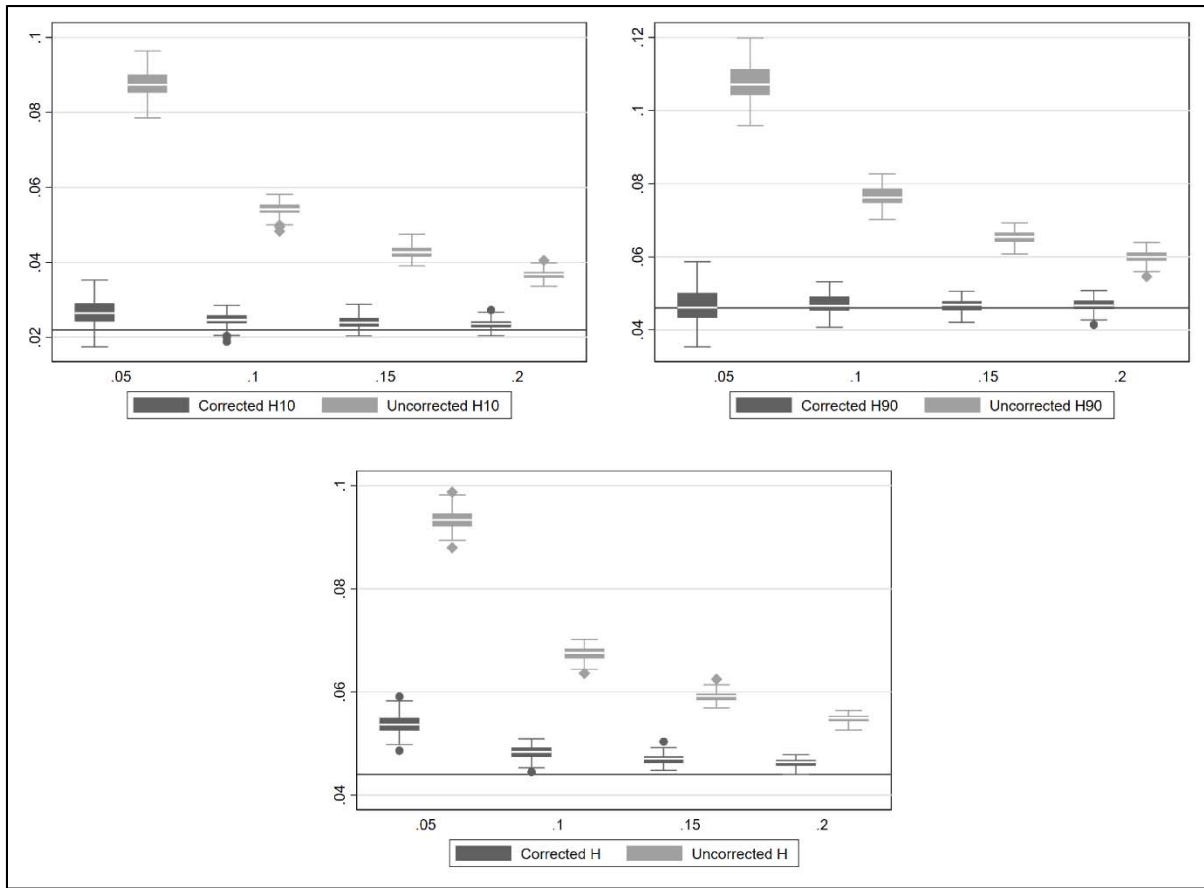


Figure 3. Estimated income segregation in Chicago 1940: true value (horizontal line), and distribution of estimates from samples of varying size, showing uncorrected and corrected values of H, H10, and H90 based on individual data, foreign-born whites.

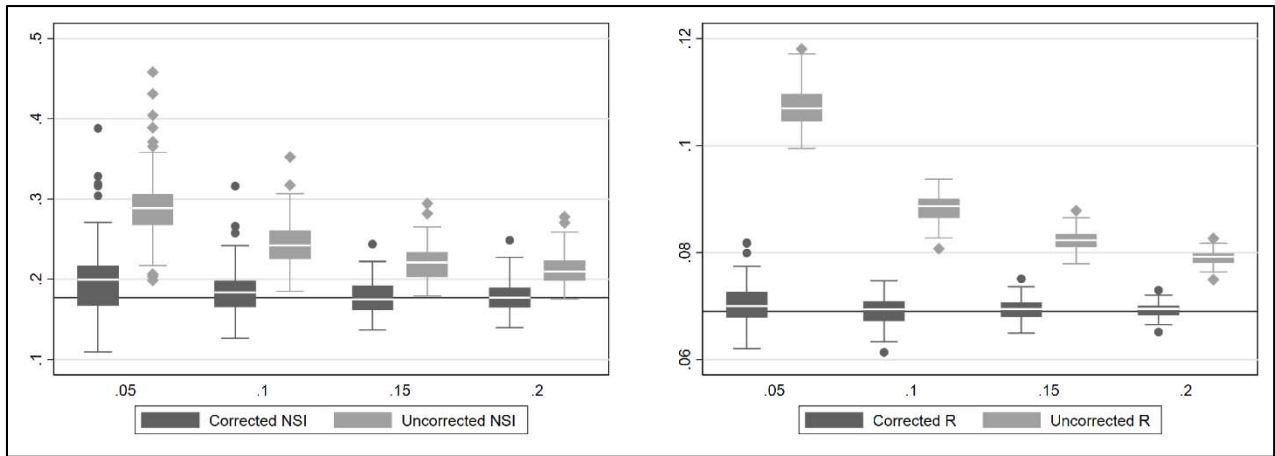


Figure 4. Estimated income segregation in Chicago 1940: true value (horizontal line), and distribution of estimates from samples of varying size, showing uncorrected values of R and NSI and adjusted values using SSVD, based on data for individual households, foreign-born whites.

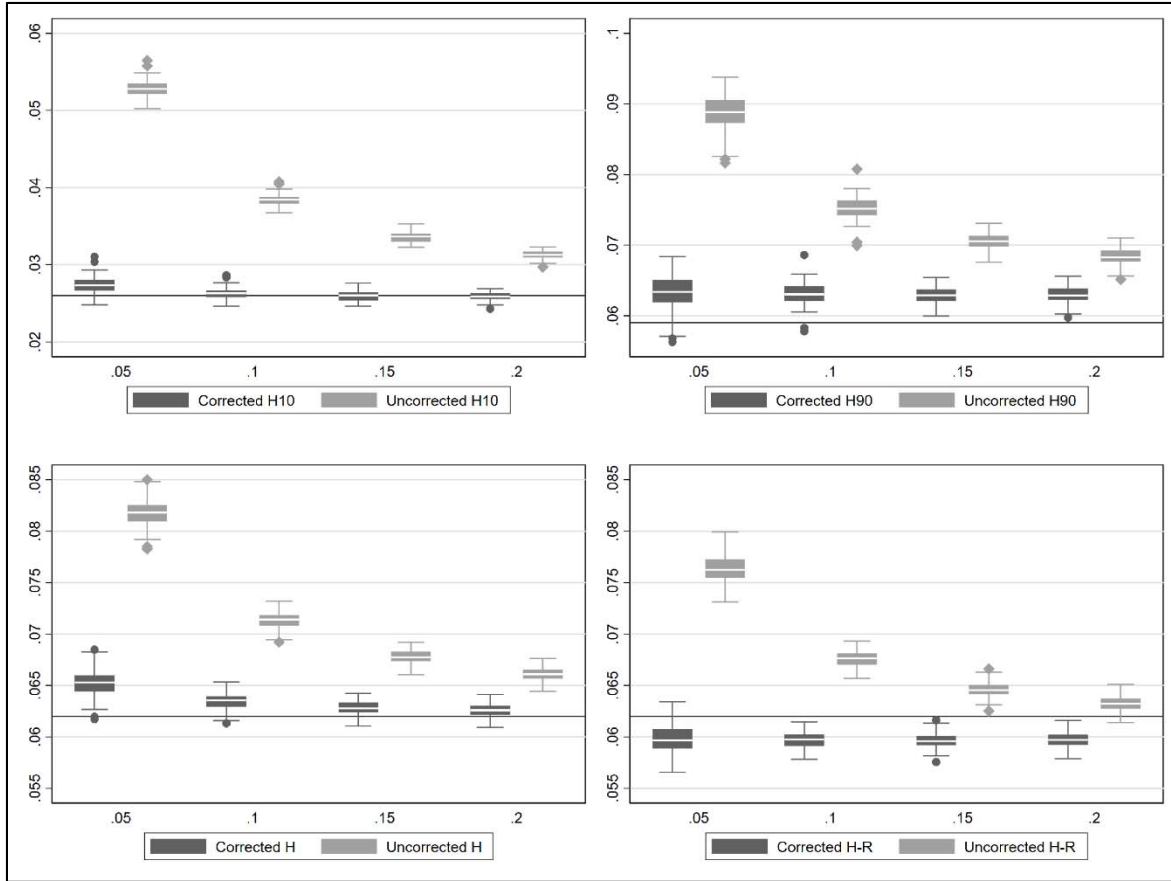


Figure 5. Estimated income segregation in Chicago 1940: true value from 100% individual data (horizontal line), and distribution of estimates from samples of varying size, showing uncorrected and corrected values of H, H-R, H10, and H90 based on grouped data, total population.

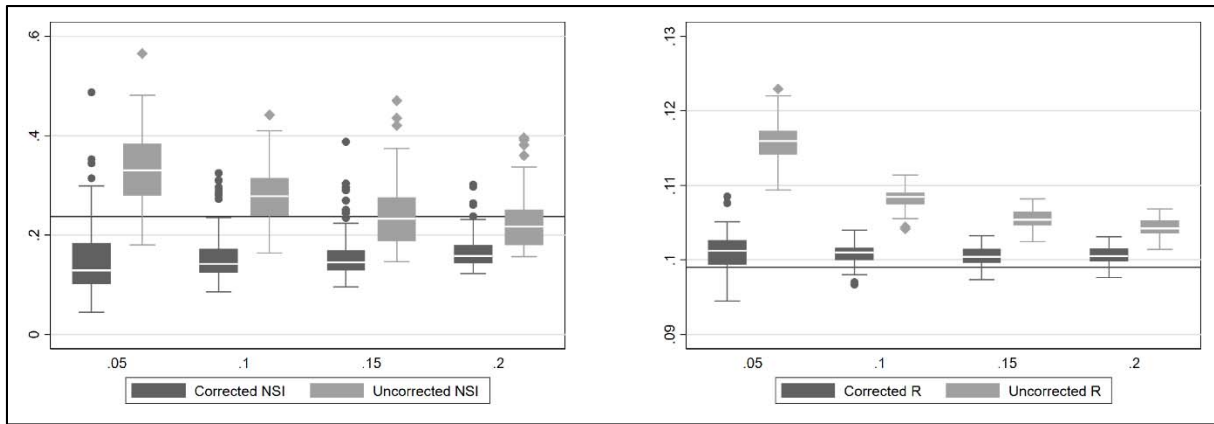


Figure 6. Estimated income segregation in Chicago 1940: true value from 100% individual data (horizontal line), and distribution of estimates from samples of varying size, showing uncorrected values of R and NSI and adjusted values using SSVD, based on grouped data, total population.